UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Molecular characterization of alpha-thalassemia and the long-range regulation of alpha-globin gene expression

Rita Sofia Lopes Pena

Mestrado em Biologia Humana e Ambiente

Dissertação orientada por:
Doutora Paula Faustino – Instituto Nacional de Saúde Doutor Ricardo Jorge
Professora Doutora Teresa Rebelo – Faculdade de Ciências da Universidade de Lisboa

2022

To my mom, for always encouraging
me to follow my dreams. I love you!

Acknowledgments

I would like to start by thanking Dr. Fernando de Almeida, president of the Directive Board of “Instituto Nacional de Saúde Doutor Ricardo Jorge” (INSA), and to Dra. Glória Isidro, coordinator of “Departamento de Genética Humana” (DGH) of INSA for allowing the realization of this study.

To my supervisor, Doctor Paula Faustino for welcoming me with open arms in the laboratory "Hemoglobinopatias, metabolismo do ferro e doenças associadas" of DGH/INSA and for the continuous support during this year. To my internal supervisor, Prof. Doctor Teresa Rebelo for agreeing to review my work.

To Dr. Pedro Lopes, who taught me everything I needed to know about the techniques used in this study and was always willing to offer advice and words of encouragement when things did not go as planned. To my lab partners, Djanira and Eduarda, I could not have completed this journey without you. To all the members of DGH for making me feel so welcome and for always being willing to help with everything.

This dissertation would not have been possible without the help of multiple people. To Dr. Gisela Gaspar and Dr. Armandina Miranda from “Unidade de Diagnóstico Laboratorial e Referência” of the “Departamento de Promoção da Saúde e Prevenção de Doenças Não Transmissíveis” (DPS) of INSA for providing the DNA samples that made this study possible and for all the hematological and biochemical analysis of the hemoglobinopathies cases. To our colleagues from "Unidade de Tecnologia e Inovação" of DGH who carried out all of the Sanger sequencing and MLPA reactions, a special thank you to Daniela Santos and Catarina Silva for all of their support with the MLPA technique. To our colleagues from "Unidade de Genética Molecular" of DGH for the DNA extractions. I would also like to give a special thank you to Prof. Doctor Francisco Pina Martins for his ongoing help with all my statistical-related questions.

I know in my heart that I could not have made it through this year without four incredible people who were there for me every step of the way, and for that, I would like to thank them individually:

- Maria, I cannot tell you how happy I am that we made it through these five years of college together. This is the second major accomplishment we achieve together and I know that we will have many more to celebrate in the future. Thank you for everything you do for me and for always being there no matter what, you are the best bestie girl I could have ever asked for. You are and always will be my person. Love you forever.
- Diogo, I lost count of how many mental breakdowns we had together this year, how many hours we spent on the phone stressing, how many times I almost ran to your lab just so we could stress together. I am so proud of us for getting to the end of this year even through all those hard parts. Thank you for somehow always knowing what I need, for always being there for me, for all the train rides and lunch breaks where we could not stop laughing, but above all, for being the best best friend in the whole world. I love you.
- Pedro, you have been there for me through it all. I cannot thank you enough for listening to all of my mental breakdowns while writing this thesis and always dropping everything to give me that hug you knew would make everything better. You are the best boyfriend I could have ever asked for. I love you to infinity and beyond.
- Djanira, I honestly do not know how I would have gotten through this year without you. Thank you for all of the memories we made, for stressing with me through all of the optimizations and

dancing like crazy people when they worked, for all of the gossip we shared, for running with me in hallways, and most importantly for always making me laugh. You have a special place in my heart.

I could not end this big list without mentioning a few other people that one way or another had an impact during these last few years and made them a lot better. So a big thank you to Samanta, Bruna, Carolina, Beatriz, Rodrigo, Francisco, and Adriana.

Last but not least, and most importantly, my wholehearted thanks goes to the three most important people in my life. To my parents, thank you for always being there for me and supporting my dreams despite any reservations you may have had. To my brother, since the day you were born that I have a best friend for life, and no matter what happens we will always have each other’s back, I could not have asked for a better brother. I love you three with my whole heart.

Resumo

As hemoglobinopatias são doenças genéticas relacionadas com a hemoglobina ($\alpha_2\beta_2$) que se classificam em qualitativas ou quantitativas. As primeiras são denominadas por variantes de hemoglobina e resultam de defeitos na estrutura da mesma, enquanto as últimas são conhecidas por talassémias, onde se inclui a α-talassémia.

A α-talassémia é, geralmente, causada pela deleção de um ou mais genes α-globínicos, o que leva à redução de síntese de cadeias α-globínicas e do tetrâmero da hemoglobina. Daí advêm glóbulos vermelhos microcíticos e/ou hipocrómicos e, nalguns casos, anemia. Existem quatro condições clínicas reconhecidas para a α-talassémia, sendo que a gravidade do fenótipo varia consoante a razão de síntese entre as cadeias α- e β-globínicas. A ausência completa de cadeias de α-globina é fatal durante a gravidez ou logo após o nascimento e, como tal, corresponde à forma mais grave de α-talassémia. Por outro lado, as variantes de hemoglobina são geralmente causadas por mutações *missense* nos genes da α-globina e podem ser clinicamente benignas ou originar diversos graus de patogenicidade.

Os genes α-globínicos encontram-se no agrupamento génico da α-globina, localizado no braço curto do cromossoma 16, sendo que a sua expressão é altamente e diferencialmente regulada durante os diferentes períodos do desenvolvimento embrionário, fetal e adulto. Esta regulação depende dos promotores dos genes α e de elementos localizados a montante dos mesmos, aos quais se dá o nome *Multispecies Conserved Sequences* (MCS-R1 a MCS-R4). O MCS-R2, também conhecido como HS-40 ou α-MRE (*Major Regulatory Element*), localiza-se 40 kb a montante dos genes α-globínicos e é essencial para a regulação da expressão dos mesmos. A sequência deste *enhancer* apresenta variabilidade polimórfica, que é característica da ancestralidade de cada população, tendo sido descritos até ao momento seis haplótipos, denominados de A a F.

Este estudo teve como objetivo contribuir para o conhecimento do espetro de deleções e mutações pontuais que causam α-talassémia ou levam ao aparecimento de variantes de hemoglobina de cadeia α, na população residente em Portugal, bem como estimar a relação existente entre os genótipos e os fenótipos observados em cada indivíduo. Para além disso, pretendeu-se identificar os diferentes haplótipos do HS-40 que estão associados à deleção de 3,7kb, a deleção α-talassémica mais comum, e determinar a sua ancestralidade na população Portuguesa.

Assim, foram analisadas amostras de sangue de 17 indivíduos com suspeita de terem α-talassémia ou serem portadores de variantes de hemoglobina de cadeia α, previamente caracterizados a nível hematológico e bioquímico. O diagnóstico de α-talassémia foi realizado através de técnicas laboratoriais como *Gap-PCR*, *Multiplex Gap-PCR Assay* e *Multiplex Ligation-dependent Probe Amplification* (MLPA). No que diz respeito ao diagnóstico das variantes, utilizou-se uma técnica de PCR convencional que permite amplificar os genes *HBA2* e *HBA1* separadamente, seguida de sequenciação de Sanger.

Nas 11 amostras estudadas para α-talassémia através das técnicas *Gap-PCR* e *Multiplex Gap-PCR* foi possível diagnosticar quatro indivíduos heterozigóticos (genótipo $-\alpha^{3,7}/\alpha\alpha$) e dois homozigóticos (genótipo $-\alpha^{3,7}/-\alpha^{3,7}$) para a deleção de 3,7kb, bem como um indivíduo com a deleção de 4,2kb em heterozigotia (genótipo $-\alpha^{4,2}/\alpha\alpha$). Para os quatro indivíduos que não apresentam nenhuma das duas deleções anteriores, recorreu-se à técnica de MLPA para a pesquisa de deleções grandes e/ou desconhecidas. Esta metodologia permitiu identificar quatro deleções distintas em heterozigotia, a deleção do Sudeste Asiático (genótipo $--^{SEA}/\alpha\alpha$) e três deleções que removem a região regulatório HS-40: deleção $(\alpha\alpha)^{ALT}$, deleção $(\alpha\alpha)^{TI}$ e deleção $(\alpha\alpha)^{MM}$.

O diagnóstico molecular de variantes de hemoglobina de cadeia α foi realizado para nove amostras e permitiu identificar oito variantes *missense* raras. Após análise em bases de dados e realização de estudos *in silico*, foi possível concluir que cinco são benignas (Hb J-Paris, Hb Nouakchott, Hb Brugg, Hb Oleander e Hb Stanleyville-II), uma é provavelmente patogénica (Hb Westmead) e duas são patogénicas (Hb Legnano e Hb Setif). É ainda importante salientar que a Hb Stanleyville-II foi encontrada em dois indivíduos em associação com a deleção 3,7kb, e que a Hb Westmead foi encontrada em simultâneo com a deleção do Sudeste Asiático e com HbE (uma variante de cadeia β).

Com o objetivo de identificar os haplótipos do HS-40 e determinar a ancestralidade da deleção 3,7kb, foi realizada uma técnica de PCR que permite amplificar um fragmento de DNA que contém a região de interesse, seguida de sequenciação de Sanger. Nesta parte do estudo foram analisadas 111 amostras de DNA: 61 de indivíduos diagnosticados com esta deleção [34 com o genótipo $-\alpha^{3,7}/\alpha\alpha$ (grupo II) e 27 com o genótipo $-\alpha^{3,7}/-\alpha^{3,7}$ (grupo III)], bem como 50 controlos sem α-talassémia (grupo I).

As alterações genéticas detetadas nas sequências do HS-40 revelaram quatro haplótipos distintos denominados A, B, C e D. O haplótipo A foi encontrado em 60,4% dos alelos estudados, sendo o mais comum em todos os grupos, o que era expectável considerando que esta é a sequência ancestral. Em geral, o segundo haplótipo mais comum foi o B (29,7% dos alelos), tendo sido encontrado em 39,0% dos indivíduos sem α-talassémia e em 27,9% dos portadores da deleção de 3,7kb. No entanto, nos indivíduos com a deleção de 3,7kb em homozigotia, o haplótipo D foi o segundo mais observado, em 27,8% dos alelos. Por sua vez, o haplótipo C foi apenas encontrado em 1,3% dos alelos, todos eles em indivíduos controlo.

No que diz respeito aos genótipos do HS-40, foram detetadas sete combinações diferentes designadas AA, AB, AD, BB, BC, BD e DD. As combinações AA e AB são as mais comuns em indivíduos sem α-talassémia (respetivamente 36,0% e 40,0%) e em indivíduos com o genótipo $-\alpha^{3.7}/\alpha\alpha$ (respetivamente 44,1% e 38,2%). No entanto, nos pacientes com o genótipo $-\alpha^{3.7}/-\alpha^{3.7}$ a combinação mais prevalente corresponde à AD (37,0%). Através de testes estatísticos verificou-se que a distribuição tanto de haplótipos do HS-40 como de genótipos apresenta diferenças significativas entre indivíduos com e sem α-talassémia.

O haplótipo D, primariamente descrito na população Africana, foi neste trabalho identificado em frequências crescentes ao longo dos grupos estudados (I a III) com apenas 1,0% nos indivíduos sem α-talassémia, 4,4% nos portadores da deleção de 3,7kb e 27,8% nos pacientes com esta deleção em ambos os alelos. Na sua maioria este haplótipo foi encontrado na combinação AD, em indivíduos com a deleção de 3,7kb em homozigotia (grupo III), sendo que neste grupo, também foi possível encontrar o haplótipo D nas combinações BD e DD. Assim, é possível concluir que o haplótipo D está fortemente associado em *cis* a esta deleção α-talassémica, sugerindo uma origem Africana para esta deleção observada na população Portuguesa. Para validação desta conclusão foi realizada uma análise de correspondência múltipla que revelou que os indivíduos sem α-talassémia da população Portuguesa estão agrupados com outras populações europeias, enquanto que indivíduos com a deleção de 3,7kb se encontram separados destas e mais proximamente relacionadas com a população Africana.

Pensa-se que a presença do haplótipo D possa resultar numa diminuição da expressão dos genes α-globínicos, uma vez que deriva do haplótipo A devido a uma substituição nucleotídica no local de ligação do fator transcricional NF-E2. Múltiplos estudos utilizando ratinhos revelaram que esta proteína atua como *enhancer* na expressão dos genes globínicos e, como tal, a sua ausência resulta em microcitose, anemia ligeira e numa diminuição da quantidade de hemoglobina. Após comparação dos parâmetros hematológicos de indivíduos com os genótipos AA, AD e DD deste último grupo, não foi

possível verificar diferenças significativas entre os mesmos. Estes resultados podem dever-se ao tamanho reduzido da amostra, bem como ao facto de a regulação da expressão dos genes da α-globina em humanos diferir daquela observada em ratinhos. Assim, seria de extrema importância realizar este tipo de estudo numa amostra maior, bem como recorrer a testes da expressão do gene da luciferase sob o controlo de diferentes haplótipos humanos de HS-40.

Em suma, este estudo permitiu identificar seis deleções distintas, três que removem um ou mais genes α-globínicos e três que afetam as regiões regulatórias que se encontram a montante do agrupamento génico da α-globina, bem como oito variantes raras de hemoglobina de cadeia α, das quais duas são consideradas patogénicas (Hb Legnano e Hb Setif) e uma é considerada como sendo provavelmente patogénica (Hb Westmead). Para além disso, foi possível diagnosticar um caso bastante complexo envolvendo a deleção α-talassémia do Sudeste Asiático, uma variante de cadeia α Hb Westmead e uma variante de cadeia β HbE que, segundo o nosso conhecimento, foi o primeiro caso em que estas hemoglobinopatias foram encontradas no mesmo indivíduo. Este estudo revelou ainda, pela primeira vez, uma associação entre um haplótipo específico do HS-40 com a deleção comum de 3,7kb na população Portuguesa, e a sua provável ancestralidade Africana.

Palavras-chave: haplótipos do HS-40; alfa-talassémia; genes da alfa-globina; variantes de hemoglobina; hemoglobinopatias.

Abstract

Hemoglobinopathies are genetic diseases that comprise abnormal structural hemoglobin variants and thalassemias, where α-thalassemia is included. Alpha-thalassemia is typically caused by the deletion of one or more α-globin genes present in the α-globin gene cluster on the short arm of chromosome 16. These deletions lead to a reduction of α-globin chain synthesis, resulting in an excess of unpaired β-chains. The α-MRE (Major Regulatory Element), also known as HS-40, is critical for α-globin gene expression regulation. The sequence of this enhancer presents polymorphic variability, where haplotypes may be identified, that is characteristic of each geographic population's ancestry.

The purpose of this research was to contribute to the knowledge of the spectrum of deletions and point mutations that cause α-thalassemia, or lead to α-chain hemoglobin variants in the population residing in Portugal. Furthermore, we intended to identify the different haplotypes of HS-40 associated with the 3.7kb deletion, the most common α-thalassemia deletion, and to determine the ancestry of this deletion in the Portuguese population.

Through this study, we were able to identify six different deletions, three that remove one or more α-globin genes and three that affect the upstream regulatory elements, as well as eight rare α-chain hemoglobin variants, of which two are considered pathogenic (Hb Legnano and Hb Setif) and one that is considered likely pathogenic (Hb Westmead). Furthermore, it was possible to diagnose a very complex clinical case involving the Southeast Asian deletion, the α-chain hemoglobin variant Hb Westmead, and the β-chain hemoglobin variant HbE. As far as we know, this case is the first report of the co-inheritance of these three hemoglobinopathies in the same individual. This study also revealed, for the first time in the Portuguese population, an association between the specific HS-40 haplotype D and the common 3.7kb α-thalassemia deletion, and its likely African ancestry.

Keywords: HS-40 haplotypes; alpha-thalassemia; alpha-globin genes; hemoglobin variants; hemoglobinopathies.

Table of contents

Acknowledgments II

Resumo IV

Abstract VII

Table of contents VIII

List of figures X

List of tables XII

List of abbreviations XIII

1. Introduction 1
   - 1.1. Erythropoiesis 1
   - 1.2. Hemoglobin 1
     - 1.2.1. Hemoglobin structure 1
     - 1.2.2. Hemoglobin during development 2
   - 1.3. Human globin gene clusters 2
     - 1.3.1. Alpha-globin gene cluster 3
       - 1.3.1.1. Alpha-major regulatory element 3
     - 1.3.2. Beta-globin gene cluster 5
   - 1.4. Hemoglobinopathies 6
     - 1.4.1. Hemoglobin variants 6
     - 1.4.2. Thalassemias 6
       - 1.4.2.1. Alpha-thalassemia 7
         - 1.4.2.1.1. Alpha-thalassemia diagnosis 7
         - 1.4.2.1.2. Molecular basis of α-thalassemia 8
         - 1.4.2.1.3. Diverse genetic conditions of α-thalassemia 9
         - 1.4.2.1.4. Treatment of α-thalassemia 10
         - 1.4.2.1.5. Prevalence of α-thalassemia in Portugal 11

2. Objectives 12

3. Materials and methods 13
   - 3.1. Population sample 13
   - 3.2. DNA extraction, quantification, and quality analysis 13
   - 3.3. Polymerase chain reaction (PCR) 13
     - 3.3.1. Gap-PCR 15
     - 3.3.2. Multiplex Gap-PCR 16
   - 3.4. Sanger sequencing 17
   - 3.5. Multiplex ligation-dependent probe amplification (MLPA) 17

3.6. Bioinformatics analysis of the hemoglobin variants 19
  3.6.1. Analysis of the impact of variants at the protein level 19
3.7. Statistical analysis 20
4. Results and discussion 21
4.1. Population characterization 21
4.2. Characterization of the hematological parameters according to the α-globin genotype 21
4.3. Hemoglobinopathies involving the α-globin genes 25
  4.3.1. Diagnosis of α-chain hemoglobin variants 26
    4.3.1.1. Clinical cases involving only Hb variants of α-chain 27
      4.3.1.1.1. Heterozygous for Hb Legnano 27
      4.3.1.1.2. Heterozygous for Hb Setif 28
      4.3.1.1.3. Heterozygous for Hb Oleander 30
  4.3.2. Alpha-thalassemia diagnosis 32
    4.3.2.1. Search of common α-thalassemia deletions 32
    4.3.2.2. Search of rare α-thalassemia deletions 34
      4.3.2.2.1. Southeast Asia deletion 36
      4.3.2.2.2. Deletions that remove the upstream regulatory elements 36
  4.3.3. Clinical cases involving deletions and Hb variants of α-chain 40
    4.3.3.1. Association of Hb Stanleyville-II and the 3.7kb α-thalassemia deletion 40
    4.3.3.2. Complex case of --SEA deletion, Hb Westmead, and HbE 42
4.4. Haplotypes of the distal regulatory region HS-40 44
  4.4.1. HS-40 haplotypes in the Portuguese population 45
    4.4.1.1. Ancestry of the 3.7kb α-thalassemia deletion in the Portuguese population 47
  4.4.2. Influence of haplotype D in the long-range regulation of α-globin gene expression 48
5. Conclusion 50
6. References 52
7. Supplementary materials 58

List of figures

Figure 1.1 – Hemoglobin structure……1
Figure 1.2 – Sites of erythropoiesis and pattern of globin synthesis during development……2
Figure 1.3 – Organization of the α-globin gene cluster on chromosome 16……3
Figure 1.4 – Localization of the DNase I hypersensitivity sites in the α-globin gene cluster, which constitute the distal regulatory region……3
Figure 1.5 – Schematic representation of the α-globin gene cluster and the upstream regulatory region……4
Figure 1.6 – Organization of the β-globin gene cluster on chromosome 11……5
Figure 1.7 – Schematic mechanism underlying the most common deletions that cause α+-thalassemia……8
Figure 1.8 – Classification of α-thalassemia and phenotypic expression……9

Figure 3.1 - Representation of the *HBA2* and *HBA1* genes, as well as the hybrid gene obtained in the presence of the 3.7kb α-thalassemia deletion……15
Figure 3.2 – Schematic representation of the electrophoretic profile that can be obtained for the diagnosis of 3.7kb α-thalassemia deletion……15
Figure 3.3 - Representation of the *HBA1ps*, *HBA2*, and *HBA1* genes in (A) a normal allele; (B) an allele with the 3.7kb deletion; and (C) an allele with the 4.2kb deletion……16
Figure 3.4 - Schematic representation of the electrophoretic profile that can be obtained for the diagnosis of both 3.7kb and 4.2kb α-thalassemia deletions……16
Figure 3.5 - MLPA probe……18

Figure 4.1 – Boxplot of the RBC count according to the α-globin genotype group: αα/αα, $-\alpha^{3.7}/\alpha\alpha$, and $-\alpha^{3.7}/-\alpha^{3.7}$……22
Figure 4.2 – Boxplot of the amount of Hb according to the α-globin genotype group: αα/αα, $-\alpha^{3.7}/\alpha\alpha$, and $-\alpha^{3.7}/-\alpha^{3.7}$……22
Figure 4.3 – Boxplot of the MCV according to the α-globin genotype group: αα/αα, $-\alpha^{3.7}/\alpha\alpha$, and $-\alpha^{3.7}/-\alpha^{3.7}$……23
Figure 4.4 – Boxplot of the MCH according to the α-globin genotype group: αα/αα, $-\alpha^{3.7}/\alpha\alpha$, and $-\alpha^{3.7}/-\alpha^{3.7}$……23
Figure 4.5 – Boxplot of the RDW according to the α-globin genotype group: αα/αα, $-\alpha^{3.7}/\alpha\alpha$, and $-\alpha^{3.7}/-\alpha^{3.7}$……24
Figure 4.6 – Biochemical characterization of a hemoglobin variant: presumably Hb Setif……26
Figure 4.7 – Visualization of the amplified DNA fragments corresponding to *HBA2* and *HBA1* genes……26
Figure 4.8 – Molecular identification of a hemoglobin variant: Hb Legnano……28
Figure 4.9 – Prediction of the pathogenic effect of the p.Arg141Leu in the α-globin protein through the PolyPhen-2 software……28
Figure 4.10 – Molecular identification of a hemoglobin variant: Hb Setif……29
Figure 4.11 – Prediction of the pathogenic effect of the p.Asp94Tyr in the α-globin protein through the PolyPhen-2 software……29
Figure 4.12 – Evolutionary conservation profile of the p.Asp94 amino acid in the α-globin protein……30
Figure 4.13 – Molecular identification of a hemoglobin variant: Hb Oleander……30
Figure 4.14 – Prediction of the pathogenic effect of the p.Glu116Gln in the α-globin protein through the PolyPhen-2 software……31
Figure 4.15 – Evolutionary conservation profile of the p.Glu116 amino acid in the α-globin protein……31
Figure 4.16 – Visualization of the DNA fragments obtained in a Gap-PCR experiment for the diagnosis of the 3.7kb deletion……32
Figure 4.17 – Visualization of the DNA fragments obtained in a Multiplex Gap-PCR experiment for the simultaneous diagnosis of the 3.7kb and 4.2kb deletions……33
Figure 4.18 – Schematic representation of the deletions diagnosed through MLPA……35
Figure 4.19 – MLPA results of an individual heterozygous for the Southeast Asian deletion ($--^{SEA}/\alpha\alpha$)……36
Figure 4.20 – MLPA results for patient number 14 revealing the $(\alpha\alpha)^{ALT}$ deletion in heterozygosity……37
Figure 4.21 – Visualization of the amplified DNA fragment generated from the $(\alpha\alpha)^{ALT}$ deletion by Gap-PCR...37
Figure 4.22 – MLPA results for patient number 15 revealing a deletion similar to the $(\alpha\alpha)^{TI}$ deletion in heterozygosity……38

Figure 4.23 – MLPA results for patient number 16 revealing a deletion similar to the $(\alpha\alpha)^{MM}$ deletion in heterozygosity……………………………………………………… ……………………………………………39
Figure 4.24 – Molecular identification of a hemoglobin variant: Hb Stanleyville-II in case 11……………………40
Figure 4.25 – Prediction of the pathogenic effect of the p.Asn78Lys in the α-globin protein through the PolyPhen-2 software…………………………… ……………………………………………………………………………41
Figure 4.26 – Evolutionary conservation profile of the p.Asn78 amino acid in the α-globin protein………………41
Figure 4.27 –Representation of the genotype diagnosed for (A) individual 7 and (B) individual 11……………42
Figure 4.28 – Molecular identification of a hemoglobin variant: Hb Westmead…………………………………42
Figure 4.29 – Prediction of the pathogenic effect of the p.His122Gln in the α-globin protein through the PolyPhen-2 software……………………………………………………… ……………………………………………43
Figure 4.30 – Evolutionary conservation profile of the p.His122 amino acid in the α-globin protein……………43
Figure 4.31 – Visualization of the amplified DNA fragment that includes the HS-40, the principal distal regulatory site of the α-globin genes……………………………………………………………………………………45
Figure 4.32 – Graphic distribution of the different HS-40 genotype frequencies (in %) according to the corresponding α-globin genotypes of the studied Portuguese individuals…………………………………………46
Figure 4.33 – Multiple correspondence analysis of the HS-40 genotypes in multiple geographic populations……48

Figure 7.1 - Information given to the adult participants for the molecular study of hemoglobinopathies………….58
Figure 7.2 - Informed consent signed by adult participants for the molecular study of hemoglobinopathies…….59
Figure 7.3 - Molecular markers used in electrophoresis performed in an agarose gel……………………………62
Figure 7.4 - Genetic code……………………………………………………………………………………………64
Figure 7.5 - Representation of the location of the probes in the SALSA MLPA probemix P140-C1 HBA……….65

List of tables

Table 1.1 – Reference values of the hematological parameters for normal adult individuals……………………….7
Table 1.2 – Molecular epidemiology of the α-globin gene cluster obtained from 100 blood cord samples of newborns in Lisbon……………………………………………………………………………………………..11

Table 4.1 - Association between the hematological parameters of our population with and without α-thalassemia.24
Table 4.2 – Hematological parameters available for 16 clinical cases of hemoglobinopathies involving the α-globin genes………25
Table 4.3 - Reference sequences of the *HBA2* and *HBA1* genes from *RefSeq* used in the comparative analysis of the results obtained from Sanger Sequencing……………………………………………………………………...27
Table 4.4 – Variants found through Sanger Sequencing in the *HBA2* and *HBA1* genes in nine clinical cases……27
Table 4.5 – HS-40 haplotypes in the Portuguese population without the α-thalassemia 3.7kb deletion, with the deletion in heterozygosity and in homozygosity………………………………………………………………….45
Table 4.6 – HS-40 genotypes in the Portuguese population without the α-thalassemia3.7kb deletion, with the deletion in heterozygosity and in homozygosity……………………………………………………………………..46

Table 7.1 - Conventional PCR conditions for the amplification of the genes *HBA2* and *HBA1*…………………..60
Table 7.2 - Conventional PCR conditions for the amplification of the regulatory region HS-40…………………61
Table 7.3 – Gap-PCR conditions for the amplification of a fragment containing the $(\alpha\alpha)^{ALT}$ deletion…………..61
Table 7.4 - Composition of the buffer solutions used during the study…………………………………………62
Table 7.5 - Gap-PCR conditions for the detection of the 3.7kb α-thalassemia deletion…………………………..62
Table 7.6 - Multiplex Gap-PCR conditions for the detection of the 3.7kb and 4.2kb α-thalassemia deletions…..63
Table 7.7 - Conditions for Sanger sequencing…………………………………………………………………63
Table 7.8 - Nucleotide code from the International Union of Pure and Applied Chemistry (IUPAC)……………64
Table 7.9 - Amino acids code from the International Union of Pure and Applied Chemistry (IUPAC)…………65
Table 7.10 - MLPA conditions for the detection of CNV in the α-globin gene cluster…………………………..65
Table 7.11 - MLPA probes for the α-globin gene cluster arranged according to their chromosomal location…….66
Table 7.12 - Descriptive statistics of the hematological parameters and Shapiro-Wilk's normality test for the population with α-thalassemia………………………………………………………………………………67
Table 7.13 - Descriptive statistics of the hematological parameters and Shapiro-Wilk's normality test for the population without α-thalassemia……………………………………………………………………………67
Table 7.14 - Descriptive statistics of the hematological parameters and Shapiro-Wilk's normality test for the population with the $-\alpha^{3.7}/\alpha\alpha$ genotype…………………………………………………………………67
Table 7.15 - Descriptive statistics of the hematological parameters and Shapiro-Wilk's normality test for the population with the $-\alpha^{3.7}/-\alpha^{3.7}$ genotype………………………………………………………………68
Table 7.16 - Description of the primers used for the sequencing of the *HBA2* and *HBA1* genes…………………68

List of abbreviations

A	Adenine
AP-1	Activator Protein 1
ATR	Alpha-Thalassemia Mental Retardation
bp	Base pairs
BZIP	Basic Region-leucine Zipper
C	Cytosine
CBC	Complete Blood Count
CNV	Copy Number Variation
ddNTPs	Dideoxy-nucleotides
deoxyHb	Deoxyhemoglobin
DEP	Departamento de Epidemiologia
DGH	Departamento de Genética Humana
DGS	Direção Geral de Saúde
DNA	Deoxyribonucleic Acid
dNTPs	Deoxyribonucleotides Triphosphates
DPS	Departamento de Promoção da Saúde e Prevenção de Doenças Não Transmissíveis
EDTA	Ethylenediamine Tetraacetic Acid
EPO	Erythropoietin
EtBr	Ethidium Bromide
G	Guanine
GATA-1	GATA Binding Protein 1
GTFs	General Transcription Factors
Hb	Hemoglobin
HbA	Adult hemoglobin
HBA1	Fetal/adult α1 gene
HBA1ps	Pseudogene ψα1
HbA$_2$	Hemoglobin A$_2$
HBA2	Fetal/adult α2 gene
HBA2ps	Pseudogene ψα2
HBB	Adult β gene
HBD	Adult δ gene
HBE	Embryonic ε gene
HbE	Hemoglobin E
HbF	Fetal hemoglobin
HBG1	Fetal Aγ gene
HBG2	Fetal Gγ gene
HbH	Hemoglobin H
HBQ	Pseudogene θ
HBZ	Embryonic ζ2 gene
HBZps	Pseudogene ψξ
HCT	Hematocrit
HPLC	High Performance Liquid Chromatography
HS	DNase I Hypersensitivity Site
HSCs	Hematopoietic Stem Cells
HVR	Hypervariable region

indels	Insertions/deletions
INSA	Instituto Nacional de Saúde Doutor Ricardo Jorge
IUPAC	International Union of Pure and Applied Chemistry
kb	kilobases
MCH	Mean Corpuscular Hemoglobin
MCHC	Mean Corpuscular Hemoglobin Concentration
MCS	Multispecies Conserved Sequences
MCV	Mean Corpuscular Volume
MED	Mediterranean
MEL	Murine Erythroleukemia
MLPA	Multiplex Ligation-dependent Probe Amplification
mRNA	Messenger RNA
NF-E2	Nuclear Factor Erythroid 2
nsSPNs	Non-synonymous SNPs
oxyHb	Oxyhemoglobin
PCR	Polymerase Chain Reaction
Poll II	RNA Polymerase II
RBCs	Red Blood Cells
RDW	Red Cell Distribution Width
RNA	Ribonucleic Acid
SEA	Southeast Asian
SIFT	Sorting Intolerant From Tolerant
SNPs	Single-nucleotide Polymorphisms
SNVs	Single-nucleotide Variants
T	Thymine
TBE	Tris-Borate-EDTA
UV	Ultraviolet
α-MRE	α-Major Regulatory Element
α-URE	α-Upstream Regulatory Element
β-LCR	β-locus Control Region

1. Introduction

1.1. Erythropoiesis

Erythropoiesis is a dynamic process in which the red blood cells (RBCs), also known as erythrocytes, are produced in the bone marrow. When the required nutrients, such as folate, vitamin B_{12} and iron are available, erythropoietin (EPO), a hormone produced by specific interstitial cells in the renal cortex, stimulates the hematopoietic stem cells (HSCs). This stimulation causes the maturation of erythroblasts to reticulocytes, which leave the bone marrow and enter the bloodstream. Reticulocytes mature in about 24 hours and give rise to RBCs. A normal human RBC has the mean lifespan of about 120 days, and the new RBCs replace the old ones that go through senescence or premature destruction [1,2].

Aside from nutritional factors, other aspects can influence erythropoiesis, including anemia. Anemia occurs when the number of RBCs in circulation decreases or the concentration of hemoglobin is low, which consequently leads to a decrease in oxygen delivery to the tissues. In response to this event, EPO production exponentially increases, resulting in an increased RBC production. The greater number of circulating RBCs allows for more oxygen delivery, lowering EPO levels and restoring normal RBC production. However, this is only possible if the anemia is caused by blood loss or hemolysis and both the bone marrow and kidneys are functioning normally [1,2].

1.2. Hemoglobin

Hemoglobin (Hb) is a two-way respiratory carrier that transports the oxygen in the RBCs from the lungs to various tissues and organs throughout the body while also facilitating the return transport of carbon dioxide [3].

1.2.1. Hemoglobin structure

The human hemoglobin is a globular tetrameric protein that contains two alpha-like (α or ζ) and two beta-like (β, ε, γ or δ) globin chains [4,5] (Figure 1.1). The lengths of these chains are very similar, with the alpha-like having 141 and the beta-like having 146 amino acid residues, but the amino acids are arranged in different sequences [3,6]. Each globin chain forms a pocket that houses the heme group, which is composed of a porphyrin ring with a ferrous iron atom in the center, to which oxygen binds reversibly [3,4,6].

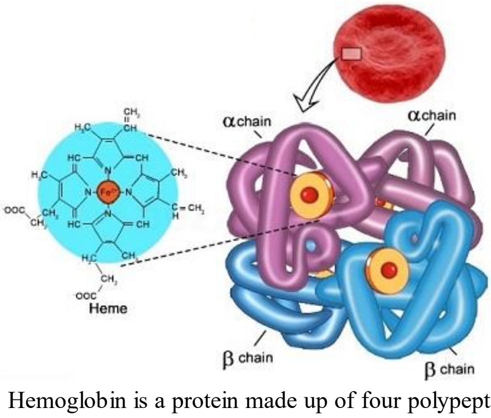


Figure 1.1 – Hemoglobin structure. Hemoglobin is a protein made up of four polypeptide chains, named globins (α1, α2, β1 and β2), each one containing a heme group composed of a porphyrin ring with a ferrous iron atom in the center, where oxygen binds. Adapted from [7].

Hemoglobin is classified into different types based on the composition of the four globin chains. Adult hemoglobin (HbA) is composed of two alpha and two beta chains ($\alpha_2\beta_2$), whereas hemoglobin A_2 (HbA_2) has two alpha and two delta chains ($\alpha_2\delta_2$). Fetal hemoglobin (HbF), on the other hand, is made up of two alpha and two gamma chains ($\alpha_2\gamma_2$) [8].

1.2.2. Hemoglobin during development

During human development, the expression of the alpha-like and beta-like chains must be tightly regulated to maintain a high and balanced concentration of hemoglobin within the RBCs [9]. This is accomplished through the complex interactions of regulatory elements and transcription factors that switch genes on and off [4].

In the first 10 to 12 weeks of gestation, the ξ and ε chains are synthesized in the erythroid cells derived from the yolk sac, giving rise to the embryonic Hb Gower I ($\xi_2\varepsilon_2$), Hb Gower II ($\alpha_2\varepsilon_2$) and Hb Portland ($\xi_2\gamma_2$). After that period, erythropoiesis shifts to the liver and spleen, and HbF becomes the most prevalent form of hemoglobin. The sites of erythropoiesis gradually shift to the bone marrow before birth, and the transition from gamma- to beta-globin synthesis begins, resulting in HbA and HbA_2 (Figure 1.2). Infants should have transitioned to predominantly HbA, a small amount of HbA_2, and negligible HbF at the age of six months [8–11].

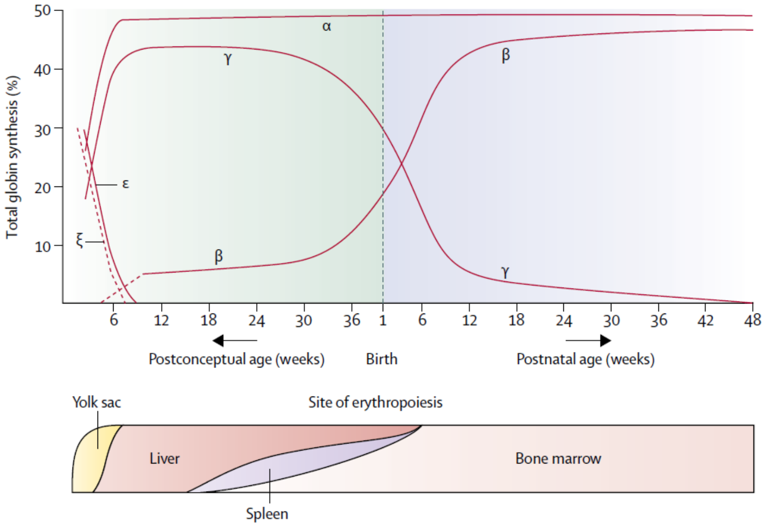


Figure 1.2 – Sites of erythropoiesis and pattern of globin synthesis during development. Fetal hemoglobin (HbF) is the most common hemoglobin type throughout the fetal period (green shade). The synthesis of β-globin chains increases just before delivery, giving rise to the adult hemoglobin HbA, which becomes the predominant form of hemoglobin in postnatal age (blue shade). Adapted from [11].

1.3. Human globin gene clusters

Two independent gene clusters on different chromosomal loci encode the alpha-like and beta-like globin chains: the α-globin gene cluster on the short arm of chromosome 16 (16p13.3) and the β-globin gene cluster on the short arm of chromosome 11 (11p15.5) [4,12,13]. In each of the clusters, the genes are organized in the 5’ to 3’ direction in the same order in which they will be expressed during human development [6,10,11].

1.3.1. Alpha-globin gene cluster

The α-globin genes are duplicated and clustered into two highly homologous units near the telomere of chromosome 16, forming an 80 kb cluster [4,13–15]. This cluster includes the embryonic ζ2 gene (*HBZ*); pseudogenes ψξ (*HBZps*), ψα2 (*HBA2ps*), and ψα1 (*HBA1ps*); two fetal/adult α genes (*HBA2* and *HBA1*) and pseudogene θ (*HBQ*) of unknown function [4,13] (Figure 1.3).

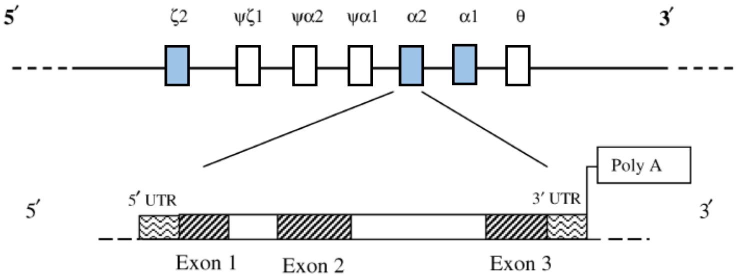


Figure 1.3 – Organization of the α-globin gene cluster on chromosome 16. Representation of the α-locus on chromosome 16 (16p13.3). The three functional genes are represented in blue rectangles; ζ2 is expressed in the fetal period, while α2 and α1 are expressed in both the fetal and adult periods. The white rectangles represent the pseudogenes (ψξ1, ψα2, ψα1 and θ). The α2 gene structure is presented below, showing the three coding exons (striped boxes), the two introns (white boxes) and the 5'UTR and 3'UTR regions (zigzag boxes) common to all globin genes. Adapted from [9].

1.3.1.1. Alpha-major regulatory element

Alpha-globin gene expression depends on both local and remote cis-acting sequences, such as the promoter sequences and the α-Upstream Regulatory Element (α-URE), respectively [16,17]. The α-URE is composed by four remote highly conserved noncoding regulatory elements called Multispecies Conserved Sequences (MCS-R1 to MCS-R4) [4,14,15,18]. These enhancers correspond to the previously identified erythroid-specific DNase I hypersensitivity sites located 48 (HS-48), 40 (HS-40), 33 (HS-33), and 10 (HS-10) kb upstream of the ζ-gene in chromosome 16 (Figure 1.4) [5,15,17,19].

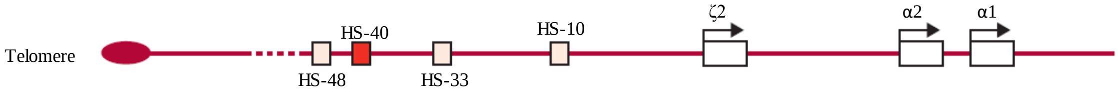


Figure 1.4 – Localization of the DNase I hypersensitivity sites in the α-globin gene cluster, which constitute the distal regulatory region. DNase I hypersensitivity sites (HS), also known as Multispecies Conserved Sequences, are located 30 to 70 kb upstream the genes and control the expression of the α-globin genes. HS-40 is the most effective regulatory element and, therefore, is represented by a red box, while HS-48, HS-33, and HS-10 are represented by cream boxes. The white boxes correspond to the ζ2, α2 and α1 genes. Adapted from [20].

MCS-R2, also known as HS-40 or α-MRE (α-Major Regulatory Element) [21], is a 350 bp enhancer whose main function is to activate and enhance the expression of the ζ-globin and α-globin promoters [22]. The functional domain of this element is composed by several conserved nuclear binding sites, such as two binding sites for the Nuclear Factor Erythroid 2 (NF-E2), four potential binding sites for GATA Binding Protein 1 (GATA-1), and four CACC boxes. However, only the two NF-E2 sites, three GATA-1 sites and one CACC box are occupied *in vivo* (black arrows in Figure 1.5). These regions recruit general transcription factors (GTFs), as well as the RNA Polymerase II (Poll II), which will bind to the promoter sequence of the α-globin genes [17,23,24].

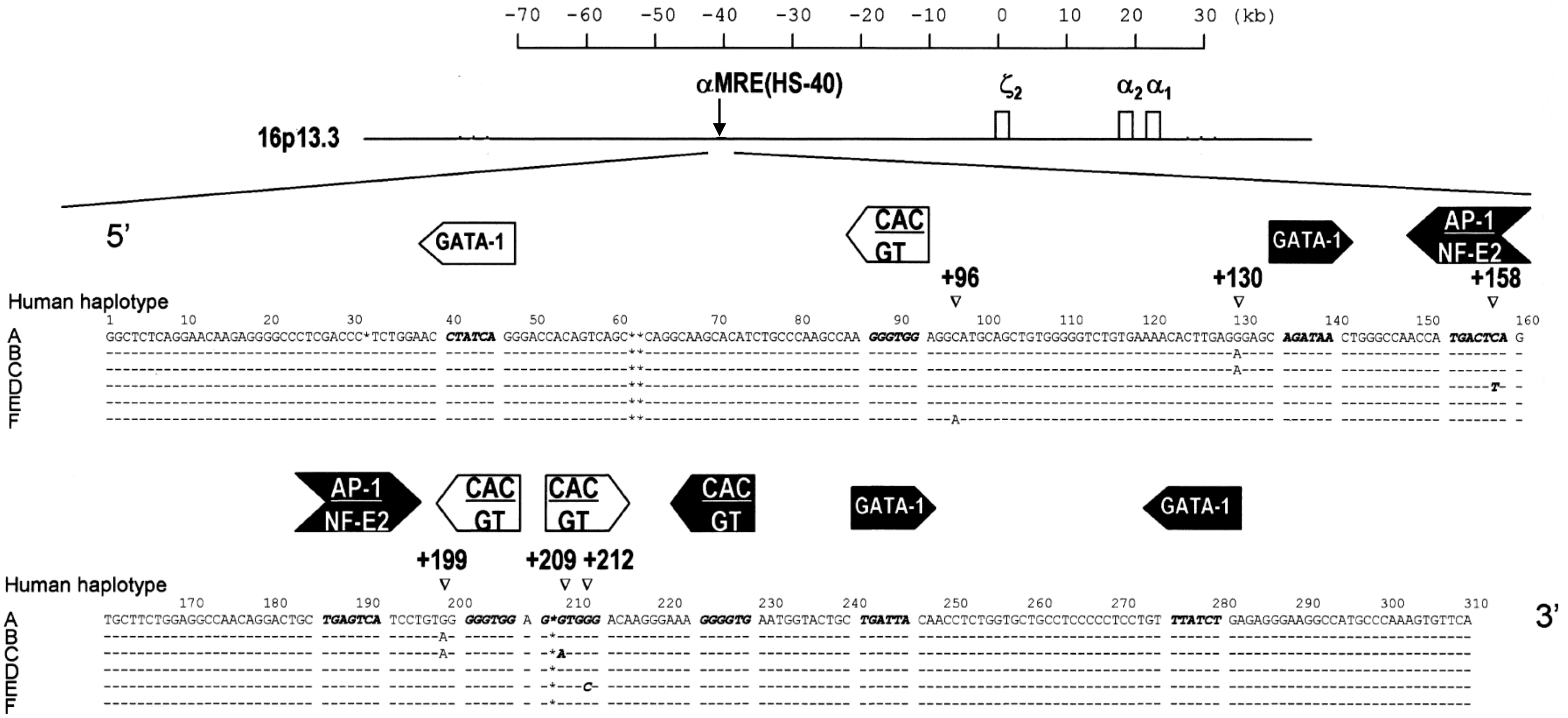


Figure 1.5 – Schematic representation of the α-globin gene cluster and the upstream regulatory region. Functional genes ζ_2, α_2 and α_1 are represented as withe boxes and the major regulatory element or HS-40 is indicated by a black arrow. The human HS-40 sequence is shown below the α-cluster. The small bars represent sequence similarity, while the asterisks represent gaps in comparison with the human sequence. The position and direction of the binding sites for nuclear factors GATA-1, AP-1/NF-E2, and the CAC/GT binding factors are indicated as arrows. Dark arrows represent the binding sites occupied *in vivo*; white arrows represent binding sites that are not occupied *in vivo*. The human HS-40 haplotypes are indicated next to each sequence as A to F. The numbers indicate the position of the polymorphic sites in the HS-40 sequence. Adapted from [25].

A variety of expression assays have been performed using different methods, including transgenic mice and transfected murine erythroleukemia (MEL) cells in order to better understand the long-range regulation of α-globin gene expression. All of these studies suggest that HS-40 is the only element of the α-URE that is capable of enhancing α-globin expression on its own and, therefore, it is considered to be critical in the regulation of the α-globin gene expression [16,22,25,26].

In a study conducted by Harteveld and colleagues in 2002, six distinct HS-40 haplotypes (labelled A to F) were discovered in European, Indonesian, Chinese and African populations. These haplotypes are characterized by the presence of one or more single-nucleotide polymorphisms (SNPs) in specific positions of the HS-40 region (Figure 1.5). In all studied populations, haplotype A was the most common and, therefore, is considered the ancestral sequence. Haplotype B, which differs from A in two substitutions at positions 130 (G→A) and 199 (G→A), was the most prevalent in the European population. Haplotype C is the least common haplotype in all groups, and it is derived from B by a substitution at position 209 (G→A). The haplotype D, which derives form A at position 158 (C→T), was the second most frequent haplotype in the African population. Pygmies are the only group who had the haplotypes E and F, which are derived from A at positions 212 (G→C) and 96 (C→A), respectively [25].

The majority of the SNPs are located either in flanking sequences of binding sites or in sites considered not to be active *in vivo,* these polymorphisms are not expected to interfere with the regulation of α-globin gene expression. However, this is not applied to haplotype D, since the base substitution at position 158 changes the first consensus binding site of NF-E2, which can affect the activity of HS-40 and, consequently, may be linked to different levels of α-globin gene expression [4,5,17,21,22].

1.3.2. Beta-globin gene cluster

The genes of the beta-like goblin chains form a cluster with approximately 70 kb on the interstitial region of the 11th chromosome [6,14]. The β-globin gene cluster is constituted by the embryonic ε gene (*HBE*), two fetal Gγ and Aγ genes (*HBG2* and *HBG1*), one pseudogene ψβ, and the adult δ and β genes (*HBD* and *HBB*) [13,14] (Figure 1.6). The expression of the β-gene cluster is controlled by the β-locus control region (β-LCR) [14].

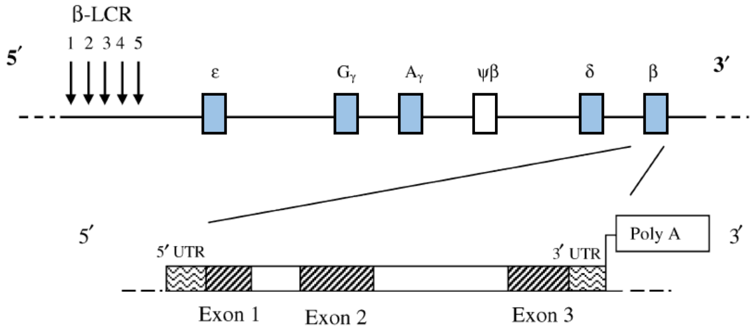


Figure 1.6 – Organization of the β-globin gene cluster on chromosome 11. Representation of the β-locus on chromosome 11 (11p15.4), with the five functional genes ε, Gγ, Aγ, δ and β represented by blue rectangles, and the pseudogene ψβ represented by a white rectangle. The black arrows indicate the location of the β locus control region (β-LCR) composed by the five DNA hypersensitivity sites. The β gene structure is presented below, showing the three coding exons (striped boxes), the two introns (white boxes) and the 5'UTR and 3'UTR regions (zigzag boxes) common to all globin genes. Adapted from [9].

1.4. Hemoglobinopathies

Hemoglobinopathies, or inherited hemoglobin disorders, are genetic diseases usually inherited in an autosomal recessive manner. There are two types of hemoglobinopathies: qualitative and quantitative. The first, also known as hemoglobin variants, are caused by defects in the structure of Hb, whereas the latter, commonly known as thalassemias, is caused by the reduction or total absence of a globin synthesis [4,13].

Nearly 1000 mutations, located in the globin genes or in their regulatory regions, have been described to alter the expression, structure or regulation of the globin genes and the hemoglobins they encode, however, only a few present with clinical manifestations [27].

1.4.1. Hemoglobin variants

Hemoglobin variants generally result from point mutations in the coding regions of the α- or β-globin genes that give rise to an amino acid substitution in the polypeptide chain of the hemoglobin molecule (missense mutations). The functional effect of each mutation depends on the characteristics of the amino acid change in the protein, its location or if affects the bonds between α- and β-chains and, consequently, has an impact in the hemoglobin tetramer formation [28].

The presence of a hemoglobin variant can result in no clinical or physiological manifestation, instability of the hemoglobin molecule, altered O_2 affinity or an increase in polymerization tendency [28]. The clinical manifestations of hemoglobin variants also depend on its quantitative proportion. Since α-globin genes are duplicated, mutations in an individual locus generally produce only 25 to 30% of an abnormal globin, while a heterozygous mutation in the β-globin locus gives rise to around 50% of the abnormal variant [27].

The biochemical study of hemoglobins using techniques such as HPLC (High Performance Liquid Chromatography) and capillary electrophoresis is usually used to diagnose these hemoglobin variants [29]; however, these methodologies only allow for a presumptive diagnosis. For a definitive diagnosis, molecular genetic methodologies, such as the amplification and sequencing of the genes suspected of having a mutation that leads to that hemoglobin variant, are required.

The most common and with a clinical importance variants are of β-chain, such as HbS, HbC, and HbE [30]. However, a study conducted in 1993 with blood samples of young Portuguese males, revealed that Hb J-Paris/J-Aljezur, a hemoglobin variant of α-chain, is common in our population [31].

1.4.2. Thalassemias

Thalassemias are one of the most common genetic diseases [5], with an estimated 7% of carriers in the whole world [13]. This disorder is characterized by the presence of mutations that cause a decrease or total absence of globin gene expression [5].

Initially, thalassemias were only found in tropical and subtropical regions, where the carrier frequency can go as high as 80-90% [4,19], including the Mediterranean where it was first discovered [5,13]. These diseases are highly prevalent in areas where malaria is or was endemic, such as Southeast Asia, the Middle East, and Africa, and carriers of hemoglobinopathies are assumed to be protected from malaria, although the mechanisms are unknown [4,5,18,19]. However, population movement increased the prevalence of thalassemias in Europe, North America, Australia, and South Korea [5,13,18,19].

Thalassemias can be classified into α-, β-, δβ-, γδβ-, δ-, and γ-thalassemia, depending on the specific type of globin chain that is affected. The two most common forms are α-thalassemia and β-thalassemia [13,32].

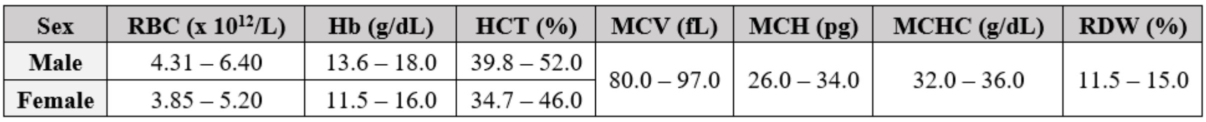
1.4.2.1. Alpha-thalassemia

Alpha-thalassemia is an autosomal recessive disorder that results from a deficiency or absence of α-globin chain synthesis, which is usually caused by the deletion of one or more α-globin genes [8,13,18,19,33]. This disease is characterized by a wide range of clinical severity and phenotypes, which vary accordingly to the degree of α-globin chain deficiency and the unbalance between α- and non-α-like globin chains [4,19].

1.4.2.1.1. Alpha-thalassemia diagnosis

The diagnosis of thalassemias starts with the characterization of the hematological phenotype, through the complete blood count (CBC) performed in a hematology analyzer equipment [5]. Among other parameters, this test evaluates the RBC count, the amount of Hb, the hematocrit (HCT), the mean corpuscular volume (MCV), the mean corpuscular hemoglobin (MCH), the mean corpuscular hemoglobin concentration (MCHC), and the red cell distribution width (RDW) [34]. In some situations, these methods may disclose varying degrees of anemia (low Hb), microcytosis (reduced MCV), and/or hypochromia (reduced MCH) [4,19]. The reference values of the RBC indices obtained in a CBC test considered for normal adult individuals by "Direção Geral de Saúde" (DGS) [35] are shown in Table 1.1.

Table 1.1 – Reference values of the hematological parameters for normal adult individuals



Sex	RBC (x 10^{12}/L)	Hb (g/dL)	HCT (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)
Male	4.31 – 6.40	13.6 – 18.0	39.8 – 52.0	80.0 – 97.0	26.0 – 34.0	32.0 – 36.0	11.5 – 15.0
Female	3.85 – 5.20	11.5 – 16.0	34.7 – 46.0				

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. Adapted from [35].

When we have a case of microcytosis (MCV < 80 fL) and/or hypochromia (MCH < 27 pg) that may or may not be accompanied by anemia, the possibility of the patient having either α- or β-thalassemia is raised. If the HbA_2 dosage is within normal levels ($HbA_2 < 3\%$), which is performed by HPLC, the diagnosis of β-thalassemia is ruled out. HPLC can also be used in combination with capillary zone electrophoresis to separate abnormal hemoglobin fractions, and thus for demonstrating the existence of any Hb variant associated with an α-thalassemia phenotype [4,19,33]. However, these approaches do not contribute to the concluding diagnosis of α-thalassemia and a molecular analysis of the α-globin genes and their gene cluster is recommended for an accurate diagnosis [4,5,33].

Molecular analysis of the α-globin genes is performed by different techniques such as Gap-Polymerase Chain Reaction (Gap-PCR), Multiplex Ligation-dependent Probe Amplification (MLPA), and Sanger Sequencing, depending on the type of mutation [4,13,15].

1.4.2.1.2. Molecular basis of α-thalassemia

More than 120 molecular defects causing α-thalassemia have been reported, with many of them resulting from deletions of various lengths in the α-globin locus [5,19]. The pathophysiology of α-thalassemia differs depending on whether these alterations occur at the level of the α1 gene or the α2 gene. Those involving the α2 gene have a greater impact on the hematological parameters because these genes encode two to three times more α-globin than the α1 gene [4,18].

Deletional α-thalassemia

As previously mentioned, α-thalassemia is most commonly caused by deletions in one α-globin gene, resulting in a decrease or complete absence of expression of that gene which, consequently, leads to a α^+-thalassemia phenotype, also known as α-thalassemia 2 [4,13,18,32]. The two most common types of α^+-thalassemia are the 3.7kb and 4.2kb deletions (Figure 1.7) [5,13]. The first deletion occurs when two homologous regions 3.7 kb apart, called Z boxes, recombine unequally during meiosis, resulting in the formation of a chromosome with one hybrid α-globin gene. Similarly, an allele with the 4.2kb deletion is obtained by the other mispaired homologous segment, known as X boxes, which are 4.2 kb apart [4,11,13,18,36].

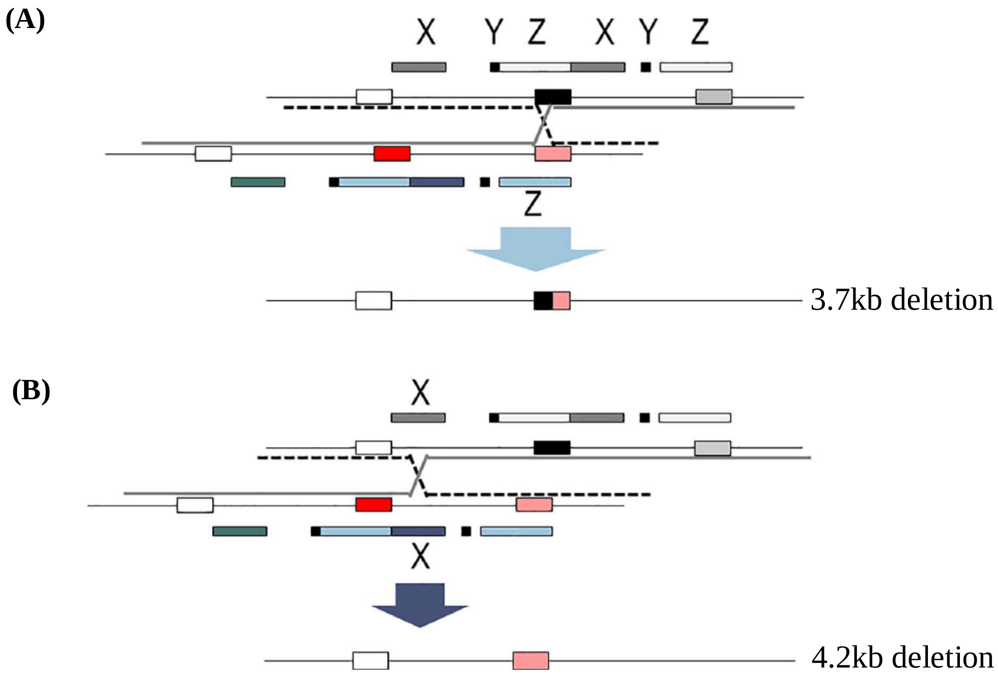


Figure 1.7 – Schematic mechanism underlying the most common deletions that cause α^+-thalassemia. The homologous duplication units X, Y and Z in which the α-globin genes are embedded are indicated as colored boxes. The pseudogene ψα1 is represented in both alleles by a white box; a black box on one allele and a red box on the other represents the α2 gene, while a grey box in one allele and a pink box on the other represent the α1 gene. **(A)** A crossover between the mispaired Z boxes gives rise to a chromosome with the 3.7kb deletion. **(B)** A crossover between the X boxes gives rise to a chromosome with the 4.2kb deletion. Adapted from [4].

Large deletions of both α-globin genes in one allele are less frequent, but they result in the complete absence of α-globin chain synthesis in one chromosome, giving rise to a α^0-thalassemia phenotype, also known as α-thalassemia 1 [4,13,18,32]. The Southeast Asian ($--^{SEA}/\alpha\alpha$) and Mediterranean ($--^{MED}/\alpha\alpha$) deletions are the most common in this group. Alpha^0-thalassemia can be caused by deletions that completely remove the MCS-R2 enhancer or small deletions restricted to this crucial region that leave all the other enhancers and genes intact, but completely inactivated [5,13].

Non-deletional α-thalassemia

Less frequently, α-thalassemia can result from non-deletional defects, such as single-nucleotide variants (SNVs) or oligonucleotides deletions/insertions in the α-globin genes or in regions that are of extreme importance for the expression of those genes. Some of these non-deletion variants can lead to a more significant decrease in α-globin synthesis than single α-gene deletions [4,13,18].

A wide range of mutations in the α-globin genes has been described. These mutations can have an effect on mRNA splicing or degradation, on globin chain stability, interfere with interactions between α- and β-globin chains or produce an unstable variant or variants with reduced or increased oxygen affinity [4].

1.4.2.1.3. Diverse genetic conditions of α-thalassemia

Normal individuals have two α-globin genes on each chromosome, however, there are four clinical conditions recognized for α-thalassemia: the silent carrier state, the α-thalassemia trait, the intermediate form of hemoglobin H disease, and the hemoglobin Bart hydrops fetalis syndrome (Figure 1.8) [4,18]. These conditions increase in severity and the phenotype that follows differs depending on the amount of α-globin chains produced in comparison to β-globin chains [18].

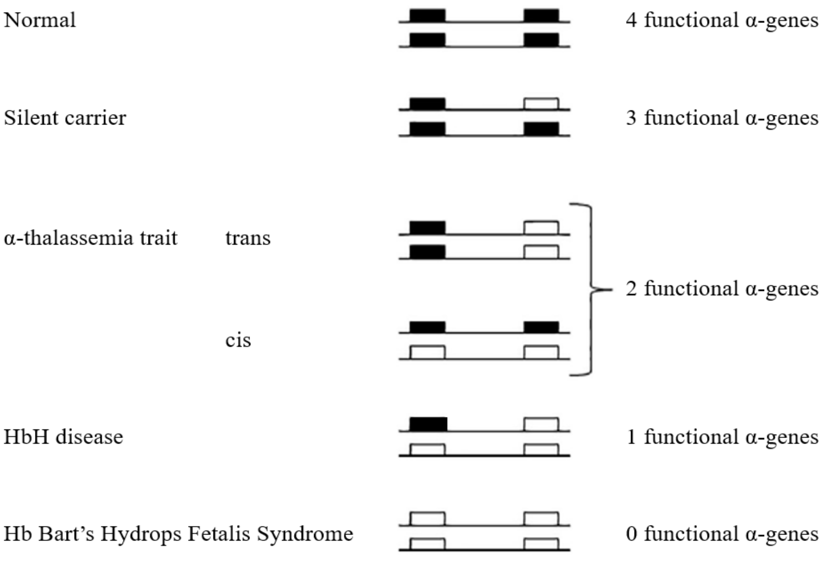


Figure 1.8 – Classification of α-thalassemia and phenotypic expression. Each allele of the chromosome 16 (represented by the black line) contains two α genes (represented by the boxes). The black boxes represent a functional gene, while the white ones represent not functional genes. Adapted from [4].

The four conditions stated above can also be classified into carrier states (silent carrier state and α-thalassemia trait) and clinically relevant forms (HbH disease and Hb Bart hydrops fetalis syndrome) [18].

A single gene deletion (-α/αα) results in the silent carrier state, which is characterized in the newborns by a slight increase in Hb Bart, a tetramer of γ-chains that forms when there is an excess of these chains relative to the α-chains [4,5,18]. In adults, this condition is usually asymptomatic with normal hematologic parameters; nevertheless, it can present with mild microcytosis and hypochromia in some cases, despite normal levels of HbA_2 and HbF [8,18].

People with the α-thalassemia trait have two functional α-genes, either in cis (--/αα) or in trans (-α/-α) [4,5,18]. These individuals usually do not have anemia or other symptoms related to it because of an increase of microcytic RBCs [4,8]. Patients with this condition are discovered by chance when their blood analyses reveal a decrease in both MCH and MCV [8,18].

Hemoglobin H disease (HbH disease), also known as α-thalassemia intermedia, is caused by a three-gene deletion (--/-α) accompanied by about 30% reduction in gene expression, resulting in an excess of β-globin chains and, consequently, in the formation of β_4-tetramers (HbH) [8,18,19]. Due to its instability, HbH precipitates within the RBCs, causing the premature destruction of these cells in the spleen and a moderate to severe hemolysis [18]. Aside from that, patients with this condition also present with microcytic and hypochromic hemolytic anemia, hepatosplenomegaly, gallstones, different forms of jaundice, and α-thalassemia-like bone modifications [4,5,8,18].

Two types of HbH disease have been reported: an acquired, which is characterized by the presence of HbH inclusion bodies in the RBCs and is associated with severe microcytic and hypochromic anemia; and α-thalassemia associated with mental retardation syndromes, that includes two forms, the ATR16 syndrome, and the ATRX syndrome. The ATR16 syndrome is caused by substantial deletions on chromosome 16 (16p13.3) that remove both α-globin genes and other flanking genes. The ATRX syndrome occurs due to the presence of point mutations in the ATRX gene, which encodes for a chromatin-associated protein that was found to be a transcriptional regulator affecting the α-globin gene expression. In this last syndrome, the mental retardation is more severe than in ATR16, and the patients have additional clinical characteristics [4,18].

The deletion of four genes (--/--) leads to a large increase in the production of hemoglobin Bart's (Hb Bart's) [5,8]. The most prevalent cause of this deletion is the inheritance of an allele that lacks functional α-globin genes from both parents [4]. This causes hemoglobin Bart hydrops fetalis syndrome, which is fatal in utero (within 20 to 38 weeks of gestation) or shortly after birth and, therefore, is the most severe form of α-thalassemia [4,5,8,18]. The fetus is unable to produce α-globin chains to make HbF or HbA, resulting in severe anemia, hydrops fetalis, hepatosplenomegaly, and heart failure [18].

1.4.2.1.4. Treatment of α-thalassemia

Individuals with the silent carrier state or with the α-thalassemia trait do not require any type of treatment or long-term surveillance [8]. However, they can present with anemia as a consequence of hematopoiesis-related nutritional deficits, in which case a ferritin level assessment is required to avoid iron overload if iron supplements are needed [4].

HbH disease causes mild to moderate hemolysis and, depending on the severity of the clinical condition, transfusions may be required [8]. Individuals with a deletion type of this disease may need intermittent blood transfusions, while those with a non-deletion type of the condition will require iron chelation therapy and more frequent transfusions [4,19]. Blood transfusions are used to give functional RBCs, reduce ineffective erythropoiesis and minimize the thalassemia's pathophysiological processes [5,11].

There is still no effective treatment for Hb Bart’s hydrops fetalis syndrome, and this condition is usually not compatible with postnatal life [18]. Despite this, several cases of intra-uterine transfusion have been performed following the diagnosis; however, the majority of surviving infants still developed limb and urogenital tract abnormalities [4].

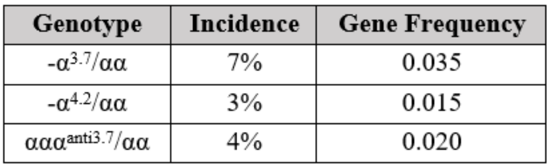
Genetic counseling before conception is advised for all persons with thalassemia [8], especially for those couples that are both heterozygous for a large cis deletion since they have a 25% risk of having an offspring with Hb Bart’s hydrops fetalis syndrome [33].

Multiple studies are currently looking into increasing the levels of α-globin gene expression or reducing β-globin or γ-globin synthesis in order to maintain the proper balance between α-like and β-like globin chains, thereby reducing the severity of α-thalassemia. One of these researches involves the transplantation of HSCs for the correction of defective HSCs in thalassemia patients; however, a matching sibling donor must be available. Another approach is to use pharmacological agents to suppress the expression of β and γ genes; however, because these medications are linked with multiple genes, they have the potential to promote tumorigenesis [5,11].

1.4.2.1.5. Prevalence of α-thalassemia in Portugal

As far as we know, the only study of the prevalence of α-thalassemia in Portugal was performed in 1995. To determine the molecular basis of this disease in our population Peres and collaborators used blood cord samples from 100 newborns of the Lisbon area. The authors concluded that 10% of the individuals had deletional α-thalassemia, with a higher frequency being detected for the 3.7kb deletion (Table 1.2). Furthermore, the authors also identified the triple α-globin gene rearrangement ($\alpha\alpha\alpha^{anti\ 3.7kb}$) and the 4.2kb deletion both in heterozygosity [37].

Table 1.2 – Molecular epidemiology of the α-globin gene cluster obtained from 100 blood cord samples of newborns in Lisbon



Genotype	Incidence	Gene Frequency
$-\alpha^{3.7}/\alpha\alpha$	7%	0.035
$-\alpha^{4.2}/\alpha\alpha$	3%	0.015
$\alpha\alpha\alpha^{anti3.7}/\alpha\alpha$	4%	0.020

Adapted from [37].

2. Objectives

For this study, three main goals were set:

1. Characterize the molecular basis, such as deletions and point mutations that cause α-thalassemia or give rise to α-chain hemoglobin variants, by analyzing the α-globin genes and their proximal and distal regulatory regions in several unusual clinical cases of hemoglobinopathies.

2. Estimate the pathophysiological mechanisms involved in the aforementioned cases and the relationship between genotypes and phenotypes.

3. Characterize the haplotypes of the distal regulatory region HS-40 (α-MRE) in individuals with and without α-thalassemia, and identify which haplotype is associated with the 3.7kb α-thalassemia deletion in Portuguese population, as well as determine the ancestry of this deletion in our population.

In order to achieve the three previous goals, three secondary objectives were also implemented:

i. Implement in the lab a methodology that allows the simultaneous detection of the 3.7kb and 4.2kb deletions, the two most frequent deletional causes of α-thalassemia, using a Multiplex Polymerase Chain Reaction Assay.

ii. Study the Portuguese cases with suspected α-thalassemia trough different methodologies, such as Gap-PCR for the detection of the 3.7kb deletion, and Multiplex Ligation-dependent Probe Amplification (MLPA) for search of unknown and/or large deletions.

iii. Amplify the DNA fragment that contains the HS-40 region and determine the variability of its sequence by Sanger sequencing.

3. Materials and methods

3.1. Population sample

In this study, we investigated 17 samples from individuals suspected of α-thalassemia or carrying a hemoglobin variant of α-chain. In addition, 111 DNAs from Portuguese individuals (61 samples with the 3.7kb α-thalassemia deletion and 50 individuals without this deletion) were also analyzed for the HS-40 sequence characterization.

The samples are part of the existing DNA biobank at “Instituto Nacional de Saúde Doutor Ricardo Jorge” (INSA), to which “Departamento de Genética Humana” (DGH), “Departamento de Promoção da Saúde e Prevenção de Doenças Não Transmissíveis” (DPS), and “Departamento de Epidemiologia” (DEP) contributed.

DNAs were obtained within the scope of scientific research projects approved by the Ethics Committee of INSA, namely “Investigação Molecular em Hemoglobinopatias” (nº 2012DGH720) and “Prevalência e identificação da base molecular da talassémia na população residente em Portugal avaliada no estudo INSEF” (nº 2019DGH1638), both taking place at DGH. In addition, clinical cases of suspected α-thalassemia were also investigated when the patient’s physician requested a genetic analysis to our lab.

All participants in this study signed an informed consent form, as exemplified for adults in Figure 7.1 and 7.2 of the Supplementary materials. All samples were characterized at a hematological and biochemical level prior to the molecular investigation, and anonymized or pseudo-anonymized when clinically relevant results were to be communicated to the clinician.

3.2. DNA extraction, quantification, and quality analysis

DNA was extracted from peripheral blood samples, collected in Ethylenediamine Tetraacetic Acid (EDTA), of the selected individuals. Our colleagues of “Unidade de Genética Molecular” extracted de DNA samples from DGH using the automatic extractor *MagNA Pure LC 2.0* (*Roche*®). On the other hand, our colleagues of DPS used a manual extraction technique based on the method described by Lahiri and his collaborators in 1991 [38] to extract DNA from their blood samples. All DNA samples were stored at 4°C.

Both DNA quantification and evaluation of purity were performed using a *NanoDrop One* (*Thermo Fisher Scientific*) spectrophotometer. First, we determined the DNA concentration (ng/µL) at 230 nm, 260 nm, and 280 nm. The ratios between absorbance at 260 and 280 nm and between 230 nm and 280 nm was calculated using these values. DNA samples were considered pure if both ratios were between 1.8 and 2.0.

3.3. Polymerase chain reaction (PCR)

Polymerase Chain Reaction, or PCR, is a molecular biology technique developed by Kary Mullis in 1983 [39]. This approach is used to exponentially amplify a specific DNA fragment, usually up to 10 kilobases (kb). A PCR mixture requires different components such as a DNA template, two primers, a

Taq DNA-Polymerase, a buffer suitable for the chosen enzyme, and deoxyribonucleotides triphosphates (dNTPs). The two primers used (forward and reverse) are short, synthetic DNA strands that are complementary to the start and end of the region to be amplified [40].

The PCR reaction is split into three main steps: (1) Denaturation; (2) Annealing; and (3) Extension. Firstly, the double-stranded DNA needs to be heated between 94ºC and 96ºC to disrupt the hydrogen bonds that connect the two strands. After dehybridization of the strands, the temperature is lowered to allow the primers to link to the single DNA strand in specific hybridization sites. Both the sequence and size of the primers determine the temperature of this step and the incorrect temperature can cause the primers to bind in unexpected sites or not bind at all. Finally, the Taq DNA-Polymerase begins extending the DNA single chain from the 3' of the annealed primer to the end of the amplicon. The extension time depends on the length of the amplified DNA fragment (for Taq DNA Polymerase is 1min/kb) and, in general, this step happens at 72ºC. It is also important to note that the PCR technique uses a series of cycles (usually between 20 to 35 cycles) of heating and cooling to make an exponential number of copies of a specific DNA region [40]. In all PCR reactions, a negative control (without DNA) is included to ensure that any contaminations are avoided.

Following the PCR, an agarose gel electrophoresis is performed to evaluate the success of the amplification. This is possible due to the visualization of the different amplified fragments as bands and the comparison of their presumed molecular weight with a molecular weight marker that comprises DNA fragments of known size. Electrophoresis requires the preparation of an agarose gel, the concentration of which depends on the size of the fragments to be separated (the bigger the fragment, the smaller the porosity of the gel and, consequently, the lower the concentration of the gel). A portion of each PCR product is injected into a well of the gel, which is electrically charged, along with a bromophenol blue solution that gives density to the amplicon. Smaller PCR products move faster toward the positive pole of the current than bigger ones, allowing them to separate. The gel also contains ethidium bromide (EtBr), which intercalates with the DNA and emits fluorescence when exposed to ultraviolet (UV) light, allowing the visualization of the bands [40].

All PCRs in this study were executed in a thermal cycler (*T1 Thermocycler, TGradient Thermocycler,* and *TProfessional Thermocycler, Biometra®*), which is a machine that heats and cools the reaction tubes to the exact temperature required for each step [40]. The separation of the amplicons was carried out in agarose gel (*SeaKem® LE Agarose, Lonza*) with different concentrations depending on the size of the fragments, and 1x Tris-Borate-EDTA (TBE) buffer containing EtBr (*Sigma*). The electrophoresis was performed in an electrophoretic system (*PowerPac 300* or *PowerPac Basic, Bio-Rad*), and the results were visualized in a UV chamber (*UVITEC*).

In this study, conventional PCR was used for the amplification of the α-globin genes, their regulatory region (HS-40), and to confirm the presence of the $(\alpha\alpha)^{ALT}$ deletion. The DNA fragments were amplified in a thermal cycler following the conditions shown in Tables 7.1 to 7.3 from the Supplementary materials, respectively.

The concentration and composition of PCR and loading buffers, as well as any solutions used in this study are described in Supplementary material's Table 7.4 and the two DNA Ladders used to compare the sizes of PCR fragments are presented in Figure 7.3 of the Supplementary materials.

3.3.1. Gap-PCR

Gap-PCR is a variation of the PCR technique implement for the detection of large deletions in a DNA sequence, using primers that flank that known deletion. For each deletion, three primers are designed (one forward and two reverses), allowing the amplification of both the normal gene and a hybrid gene that results from this deletion [41]. In this type of PCR, two different PCR's mixtures (mix I and mix II) are prepared, differing only in the reverse primer. Through this methodology, and following the PCR conditions presented in Table 7.5 from the Supplementary materials, it was possible to determine the presence of the 3.7kb α-thalassemia deletion (Figure 3.1).

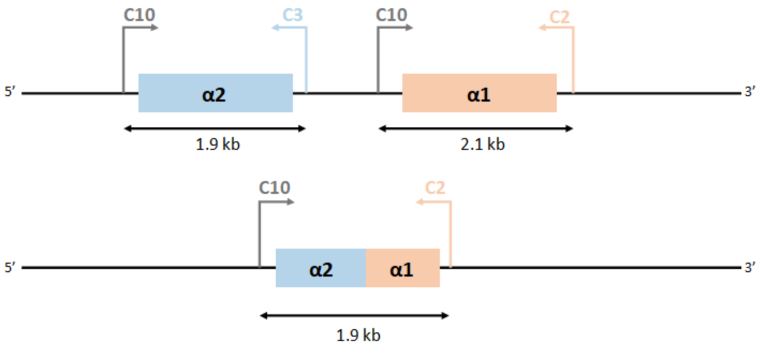


Figure 3.1 - Representation of the *HBA2* and *HBA1* genes, as well as the hybrid gene obtained in the presence of the 3.7kb α-thalassemia deletion. The *HBA2* (α2) gene is represented in blue and the *HBA1* (α1) gene is in orange. The hybrid gene α2α1 that results from the 3.7kb α-thalassemia deletion is represented in both colors. It also represented the ligation sites of the three primers (C10, C3, and C2) used for the amplification, plus the expected sizes of the three obtained fragments.

After separation of the amplified fragments by electrophoresis, the analysis of the band is used to diagnose the 3.7kb α-thalassemia deletion. In case there is a 1.9 kb band in mix I and a 2.1 kb band in II, the individual is a normal homozygous (αα/αα). If it is possible to observe a 1.9 kb band in mix I and two bands of 1.9 kb and 2.1 kb in mix II, the patient is a heterozygous for the deletion ($-\alpha^{3.7}/\alpha\alpha$). In the case of the individual having the deletion in homozygosity ($-\alpha^{3.7}/-\alpha^{3.7}$), it is possible to see one band of 1.9 kb in mix II, while no band is present in mix I (Figure 3.2).

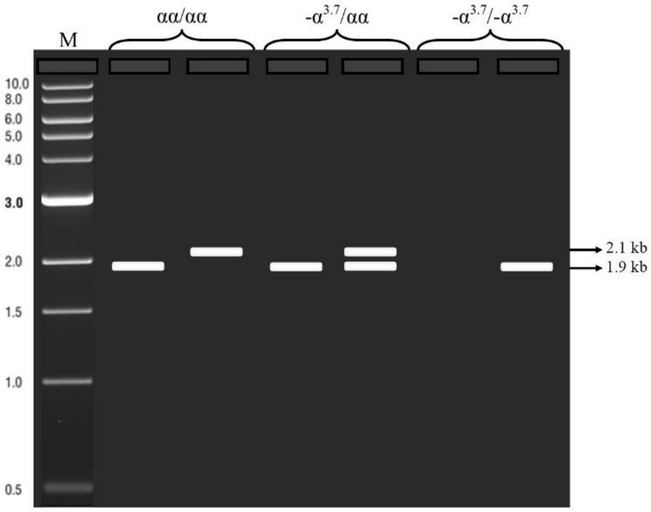


Figure 3.2 – Schematic representation of the electrophoretic profile that can be obtained for the diagnosis of 3.7kb α-thalassemia deletion. Interpretation of Gap-PCR results according to the subject genotype. **M** – molecular weight marker (1kb DNA Ladder, *BioLabs*)

3.3.2. Multiplex Gap-PCR

Multiplex Polymerase Chain Reaction Assay, also known as Multiplex Gap-PCR, allows for the simultaneous identification of multiple α-thalassemia deletions [42,43]. Because the 3.7kb and the 4.2kb single gene deletions are the most common causes of α-thalassemia, we implemented a methodology in the lab that allows the diagnosis of these deletions in a single PCR reaction (Table 7.6, Supplementary materials). This approach requires the use of three pairs of primers, two for each deletion and one for the detection of the *HBA2* gene (Figure 3.3).

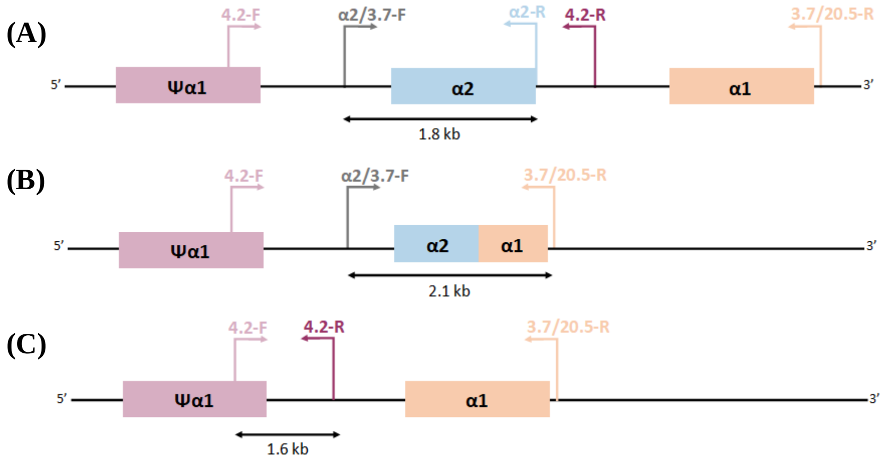


Figure 3.3 - Representation of the *HBA1ps*, *HBA2*, and *HBA1* genes in (A) a normal allele; (B) an allele with the 3.7kb deletion; and (C) an allele with the 4.2kb deletion. The *HBA1ps* (ψα1) gene is represented in pink, the *HBA2* (α2) gene in blue, and the *HBA1* (α1) gene is in orange. The hybrid gene α2α1 that results from the 3.7kb deletion is represented in both blue and orange. It also represented the ligation sites of the five primers (4.2-F, α2/3.7-F, α2-R, 4.2-R, and 3.7/20.5-R) used for the amplification, plus the expected sizes of the three obtained fragments.

Because both deletions either partially or completely remove the α2 globin gene, its positive amplification indicates heterozygosity when a deletion allele is also present (two bands of different sizes) or a normal individual if just one 1.8 kb band is present. The presence of a 2.1 kb band indicates an allele with the 3.7kb deletion while the 4.2kb deletion is noted through a smaller band of 1.6 kb (Figure 3.4).

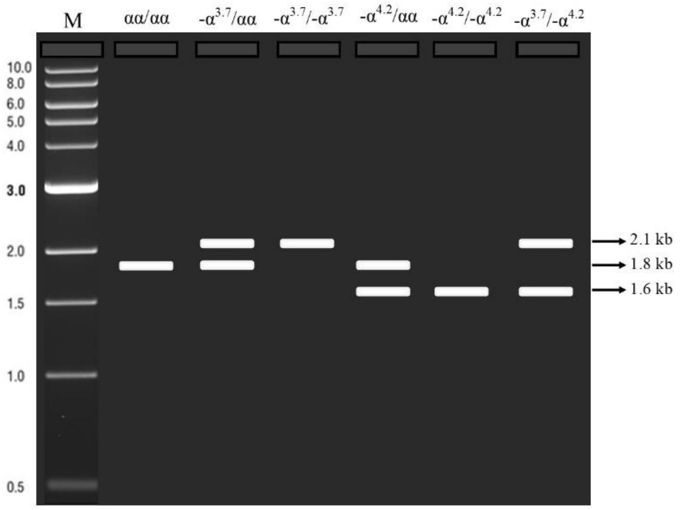


Figure 3.4 - Schematic representation of the electrophoretic profile that can be obtained for the diagnosis of both 3.7kb and 4.2kb α-thalassemia deletions. Interpretation of Gap-PCR results according to the subject genotype. **M** – molecular weight marker (1kb DNA Ladder, *BioLabs*).

3.4. Sanger sequencing

DNA sequencing is a biochemical process that determines the correct order in which the nucleotide bases of a DNA macromolecule are arranged [44]. Sanger Sequencing was the first sequencing technology developed by Sanger and collaborators in 1977 [45] and, despite being a first-generation sequencing method, it has subsequently become the most widely used sequencing technique due to its high efficiency [44].

Sanger Sequencing, also known as chain termination method, consists in the synthesis of DNA fragments using one strand of the double-stranded DNA as a template, dNTPs for the extension of the fragment, an enzyme and a suitable buffer. The sequencing process is possible due to the use of chemically modified nucleotides called dideoxy-nucleotides (ddNTPs), which are labeled with distinct colored fluorophores: ddGTP - black, ddATP - green, ddTTP - red, and ddCTP - blue. The fragments differ from one another by only one nucleotide and are separated through capillary electrophoresis according to their size [44].

In this study, the amplified PCR products were purified using the JET quick PCR Product Purification Spin Kit (*GENOMED*) according to the manufacturer's instructions, followed by a PCR reaction with the conditions described in Table 7.7 from the Supplementary materials. The sequencing reactions were carried out with the *BigDye© Terminator v1.1 Cycle Sequencing* commercial kit (*Applied Biosystem*). The nucleotide sequences were obtained through capillary electrophoresis in the automated sequencer *3130XI Genetic Analyser* (*Applied Biosystem*) and the results were analyzed with the *FinchTV v.1.4.0* (*Geospiza*) software.

In this work, Sanger sequencing was used to determine the sequence of different HS-40 haplotypes and for the molecular diagnosis of patients with suspected α-chain hemoglobin variants. For the analysis of the results, it is important to take into consideration the genetic code and the nomenclature established by the International Union of Pure and Applied Chemistry (IUPAC) (Figure 7.4, Table 7.8 and 7.9, Supplementary materials).

3.5. Multiplex ligation-dependent probe amplification (MLPA)

Although PCR-based techniques can be used to detect frequent deletions in the α-globin genes, most exons deletions and duplications are undetectable because a normal allele is present. Multiplex Ligation-dependent Probe Amplification (MLPA), first described in 2002 by Schouten and collaborators, is a simple approach used for the molecular diagnosis of a variety of genetic disorders, including deletional α-thalassemia. This technique allows the detection of deletions and duplications in the screened region and has been proven to find known and unknown deletions. This is possible through the detection of the copy number variation (CNV), a DNA sequence longer than 1 kb with a variable copy number when compared to the reference genome and is responsible for a wide range of pathologies [13,32,46–48].

MLPA is a multiplex-PCR assay that evaluates the relative copy number of a DNA sequence using up to 60 probes specific for each sequence. Each probe consists of two DNA oligonucleotides (5' and 3' half-probe) that will hybridize to specific DNA sequence targets (Figure 3.5). One or both half-probes contains a stuffer sequence that enable each probe to be distinguished during electrophoresis based on its length, allowing the determination of the amplification product size [46–48].

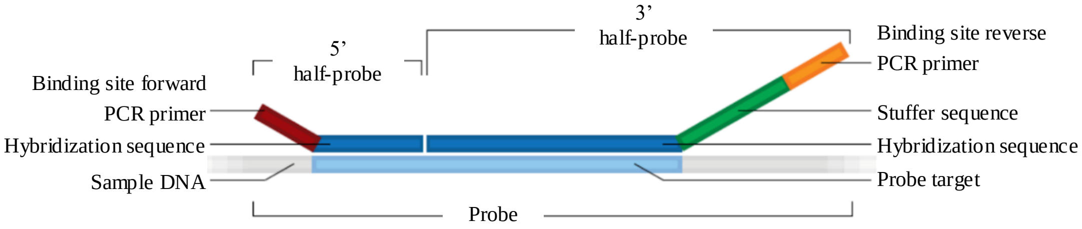


Figure 3.5 - MLPA probe. Schematic representation of an MLPA probe, which consists of two parts: a left and a right probe oligonucleotide (5’ and 3’). Both the 5’ and 3’ half-probes contain a PCR primer and a hybridization sequence. The 3’ half-probe also contains a stuffer sequence. Adapted from [46].

The MLPA reaction is split into five steps: (1) DNA denaturation and probe hybridization; (2) ligation reaction; (3) PCR amplification; (4) separation of the amplification products by capillary electrophoresis; and (5) data analysis using the *Coffalyser.Net* software (*MRC Holland*). After denaturation of the DNA sample, purified DNA is incubated with a probe mixture over night, and each probe will hybridize with its target sequence. During the subsequent PCR reaction, a universal pair of primers will be utilized for simultaneous amplification of all probes, resulting in a set of unique amplicons that are a measure of the number of target sequences in the initial sample. One of the primers used in the PCR reaction is fluorescently labeled, allowing each fragment to be detected during capillary electrophoresis, where the PCR products are separated based on their size. This separation gives rise to an eletropherogram called the sample peak pattern [46–48].

The interpretation of the results in MLPA is performed by a comparison of the sample peak pattern. When compared to the relative heights in distinct reference DNA samples, the relative height of each probe peak in the sample under analysis indicates the relative copy number of the corresponding target sequence. A decrease in the relative height of the peaks reflects deletions of a target sequence, whereas an increase in this height reflects a rise in the number of copies. It is also worth noting that homozygous or hemizygous deletions are evidenced by the absence of specific target genes peaks, while heterozygous deletions result in variable heights and/or areas of the peaks [46,48].

In this study, MLPA was performed using the SALSA® MLPA® probemix P140-C1 HBA kit (*MRC-Holland*), following the manufacturer's instructions (Table 7.10 and 7.11 and Figure 7.5, Supplementary materials). This kit is used for the detection of deletions and duplications in the α-globin gene cluster and in its regulatory regions in genomic DNA isolated from peripheral blood samples, to confirm or diagnose potential causes of α-thalassemia [49]. In an MLPA reaction, each sample is tested simultaneously with three normal controls, one positive control, and one negative control. The amplified fragments were separated by capillary electrophoresis in the *3500 Genetic Analyser*, *Abi Prism* (*Applied Biosystem*). The quality of the peaks was evaluated through the software *Microsatellite Analysis* v1.0.1 (*Applied Biosystems*) and the quantitative data were obtained with the *Coffalyser.Net* software (*MRC-Holland*). After standardization, the areas of the peaks were used to determine the relative number of copies in the genomic sequences of each sample.

3.6. Bioinformatics analysis of the hemoglobin variants

For the bioinformatics analysis of the hemoglobin variants found during this study, we used different online databases: *Ensembl* (https://www.ensembl.org/index.html); *IthaGenes* (https://www.ithanet.eu/db/ithagenes); *HbVar* (https://globin.bx.psu.edu/hbvar/hbvar.html); *ClinVar* (https://www.ncbi.nlm.nih.gov/clinvar/); *PubMed* (https://pubmed.ncbi.nlm.nih.gov/); and *UniProt* (https://www.uniprot.org/blast/). These databases allow us to obtain the genomic location of the genes that are being studied, the nucleotide sequence, the amino acid sequence that results from the translation of each gene, as well as an *in silico* prediction of the clinical importance of the variants found during the study.

In these analyses, we used as a reference the *Ensembl* sequence ENSG00000188536 for the *HBA2* gene and the transcript ENST00000251595.11. In regards to the *HBA1* gene, we used the *Ensembl* reference sequence ENSG00000206172 and the transcript ENST00000320868.9. Both genes encode the same corresponding protein sequence according to the *UniProtKB* identifier P69905.

3.6.1. Analysis of the impact of variants at the protein level

The majority of human genetic variations occur due to SNPs that lead to different phenotypes. When these SNPs are located in coding regions, they are called non-synonymous SNPs (nsSNPs) and result in amino acid changes (missense variants) in the protein sequence. Missense variants can cause human hereditary diseases and, because of that, it is important to study their impact at a structural and functional level [50].

PolyPhen-2

To study the impact of missense variants we used the online software *PolyPhen-2* (http://genetics.bwh.harvard.edu/pph2/). This automatic tool predicts the possible impact of amino acid substitutions on the function and stability of human proteins using structural and comparative evolutionary factors. Based on these properties, the software estimates the probability of the variant being pathogenic, likely pathogenic, or benign [50,51].

PolyPhen-2 uses two different models to make the predictions: *HumDiv* and *HumVar*. All damaging alleles (present in the *UniProt* database) known to cause human Mendelian diseases, as well as non-damaging differences between human proteins and their mammalian homologs are included in *HumDiv*. *HumVar* consists of all the mutations (also present in the *UniProt* database) that cause disease in humans, along with nsSNPs without annotations of being involved in human diseases. The *HumDiv* model is used for evaluating rare alleles that may be involved in complex phenotypes and natural selection analysis, while the *HumVar* model is used for Mendelian diseases diagnosis, which require the differentiation between mutations that have drastic effects from all other human variations, including numerous mildly deleterious alleles [51].

The score given by the software corresponds to the probability of the missense mutation being damaging and ranges from 0.00 to 1.00: between 0.00 and 0.50, the variants are considered benign, while between 0.50 and 1.00, they are considered likely pathogenic or pathogenic.

Sorting Intolerant From Tolerant (SIFT)

To give better support to the impact that a missense mutation can have on a protein, we also used the *SIFT* (Sorting Intolerant From Tolerant) software (https://sift.bii.a-star.edu.sg/). This program predicts if an amino acid substitution affects protein function and, thus, the phenotype, using sequence homology. Because SIFT assumes that essential amino acids would be conserved within a protein family, any alterations in well-conserved positions tend to be predicted as probably deleterious [52]. This software can also predict the impact of insertions/deletions (indels) in the coding regions of a gene, which might result in frameshift mutations or changes in amino acid sequence [53].

Although this software gives a score in the same range as *PolyPhen-2*, the classification is the opposite: between 0.00 and 0.05, the mutations are considered to be damaging and between 0.05 and 1.00, they are considered benign (tolerated).

3.7. Statistical analysis

The statistical treatment of all data obtained during this study was performed using *R* software and *RStudio* v.2021.09.1+372. For all analyses, the statistical significance was established for a *p*_value lower than 0.05.

All data were tested for its normality distribution using Shapiro-Wilk's test (Tables 7.12 to 7.15, Supplementary materials). When the normality of both populations was confirmed, the parametric test T-test was used to compare hematological parameters between patients with α-thalassemia and healthy individuals, as well as between those with the 3.7kb deletion in heterozygosity and homozygosity. The non-parametric test of Mann-Whitney was applied when there were statistically significant differences between one or both samples and a normal distribution. The same process was implemented upon the comparison of HS-40 haplotypes between the group with the 3.7kb deletion in heterozygosity and healthy individuals, or between the former and the group with the same deletion in homozygosity.

In order to determine the ancestry of the 3.7kb α-thalassemia deletion in the Portuguese population, a multiple correspondence analysis was performed using the *R* software and a specific function to draw the respective graphical representation.

4. Results and discussion

4.1. Population characterization

Based on the results of the hematological and biochemical analyses, the samples used in this study were divided considering the previously set aims: 1) to characterize the deletions and point mutations that cause α-thalassemia and give rise to α-chain hemoglobin variants, respectively; and 2) to identify the haplotypes of the distal regulatory region HS-40.

Thus, 17 individual clinical cases were selected for the diagnosis of hemoglobinopathies. For the identification of the HS-40 haplotypes, we selected **111 Portuguese individuals** that were split into two groups:

- Individuals without α-thalassemia: including **50 individuals** equally divided between males and females with a mean age of 48 ± 17 years old (range 20 to 86 years old) that presented normal hematological parameters and gave a negative result for the presence of the 3.7kb deletion.

- Individuals with the 3.7kb α-thalassemia deletion: including **61 individuals** (22 males and 39 females) with a mean age of 47 ± 14 years old (range 16 to 77 years old) that were previously studied through Gap-PCR for the presence of the 3.7kb deletion and gave a positive result (**34** have the genotype $-\alpha^{3.7}/\alpha\alpha$ and **27** with the $-\alpha^{3.7}/-\alpha^{3.7}$ genotype).

4.2. Characterization of the hematological parameters according to the α-globin genotype

In this study, all the statistical analysis were performed with the following hematological parameters: RBCs, Hb, HCT, MCV, MCH, MCHC, and RDW. It is important to note that during this study we followed the guidelines from DGS [35], which considers as adults all individuals above 11 years old because after that age the hematological parameters do not change. Considering that hematological values also change during pregnancy, four pregnant females were excluded from the statistical analysis.

Firstly, we analyzed the presence of outliers through the representation of all hematological parameters in boxplots (Figures 4.1 to 4.5) available for three groups: individuals without α-thalassemia (genotype αα/αα), individuals with the 3.7kb deletion in heterozygosity (genotype $-\alpha^{3.7}/\alpha\alpha$), and individuals with the 3.7kb deletion in homozygosity (genotype $-\alpha^{3.7}/-\alpha^{3.7}$). Since the number of RBCs and the amount of Hb varies with the individual's sex, for these two parameters they were subdivided into females and males. All the outliers observed were removed from any further statistical analysis to allow for more reliable results.

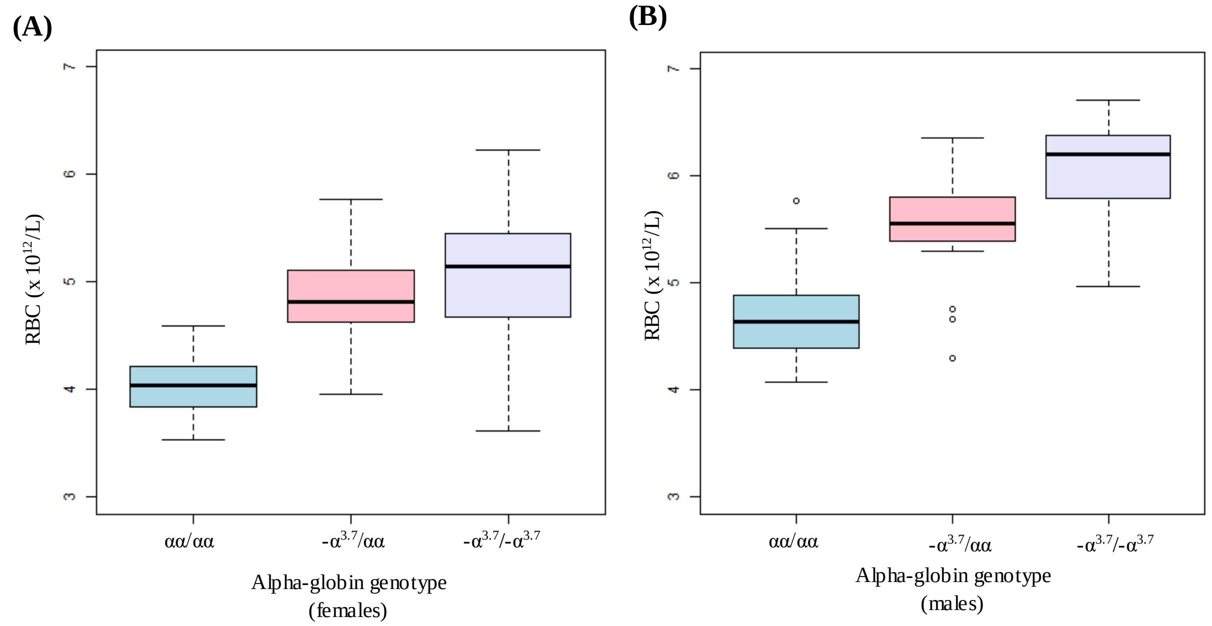


Figure 4.1 – Boxplot of the RBC count according to the α-globin genotype group: αα/αα, -$\alpha^{3.7}$/αα, and -$\alpha^{3.7}$/-$\alpha^{3.7}$. (A) females and **(B)** males. Lower and upper box boundaries represent 25th and 75th percentiles, respectively, the bold line inside the boxes represent the median, lower and upper error lines represent the 10th and 90th percentiles, and the white circles correspond to data falling out of the 10th and 90th percentile (outliers).

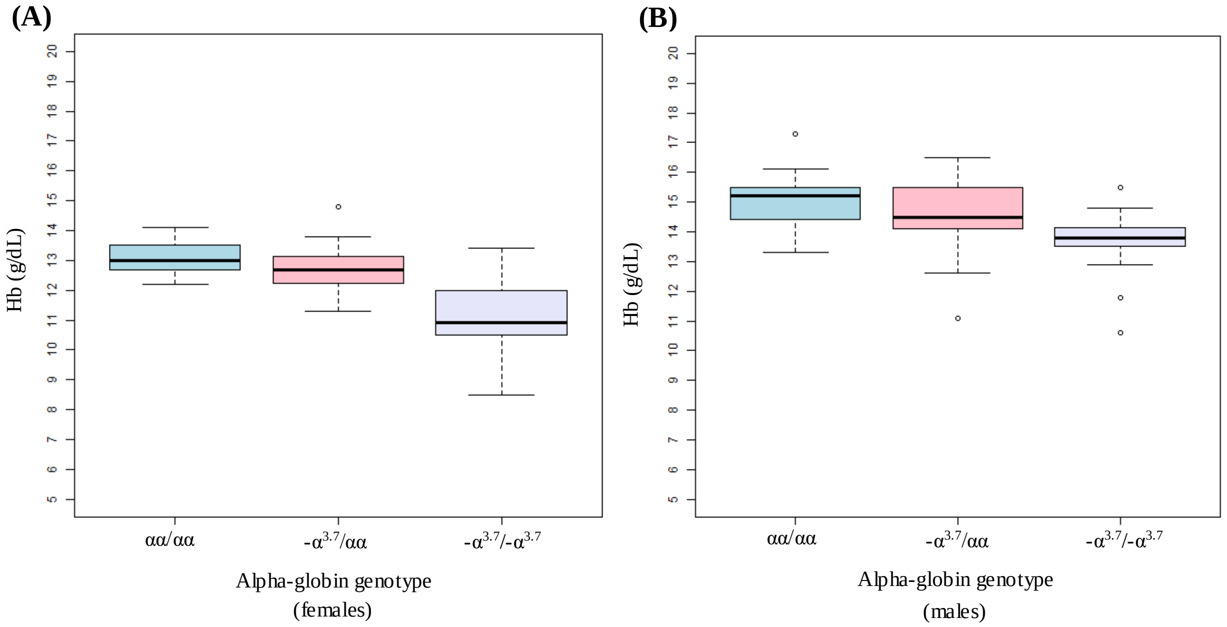


Figure 4.2 – Boxplot of the amount of Hb according to the α-globin genotype group: αα/αα, -$\alpha^{3.7}$/αα, and -$\alpha^{3.7}$/-$\alpha^{3.7}$. (A) females and **(B)** males. Lower and upper box boundaries represent 25th and 75th percentiles, respectively, the bold line inside the boxes represent the median, lower and upper error lines represent the 10th and 90th percentiles, and the white circles correspond to data falling out of the 10th and 90th percentile (outliers).

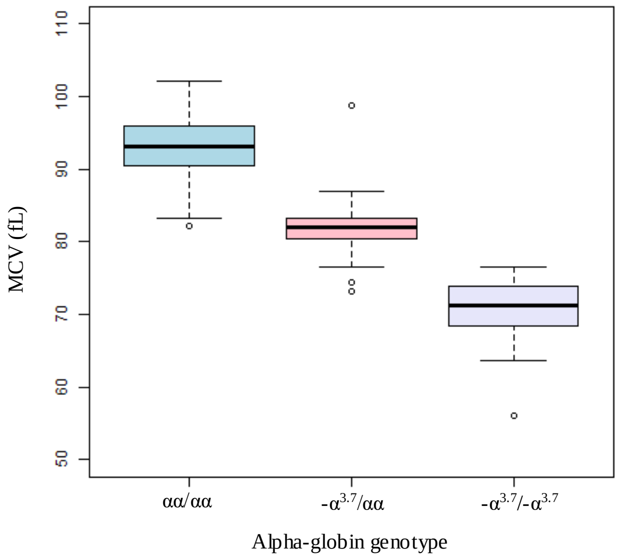


Figure 4.3 – Boxplot of the MCV according to the α-globin genotype group: αα/αα, $-\alpha^{3.7}/\alpha\alpha$, and $-\alpha^{3.7}/-\alpha^{3.7}$. These representations include both females and males. Lower and upper box boundaries represent 25th and 75th percentiles, respectively, the bold line inside the boxes represent the median, lower and upper error lines represent the 10th and 90th percentiles, and the white circles correspond to data falling out of the 10th and 90th percentile (outliers).

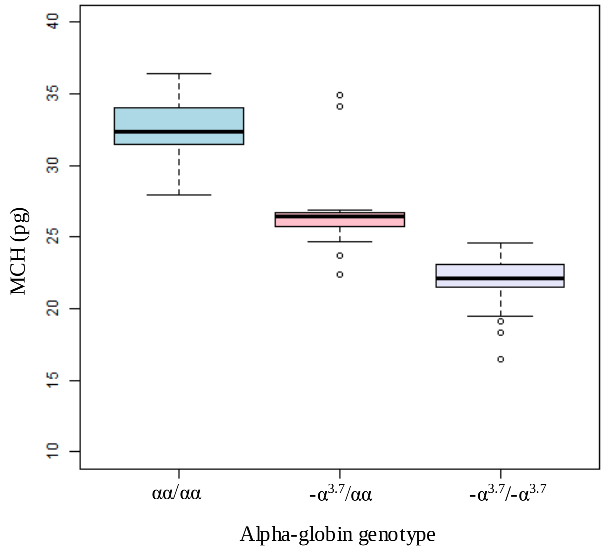


Figure 4.4 – Boxplot of the MCH according to the α-globin genotype group: αα/αα, $-\alpha^{3.7}/\alpha\alpha$, and $-\alpha^{3.7}/-\alpha^{3.7}$. These representations include both females and males. Lower and upper box boundaries represent 25th and 75th percentiles, respectively, the bold line inside the boxes represent the median, lower and upper error lines represent the 10th and 90th percentiles, and the white circles correspond to data falling out of the 10th and 90th percentile (outliers).

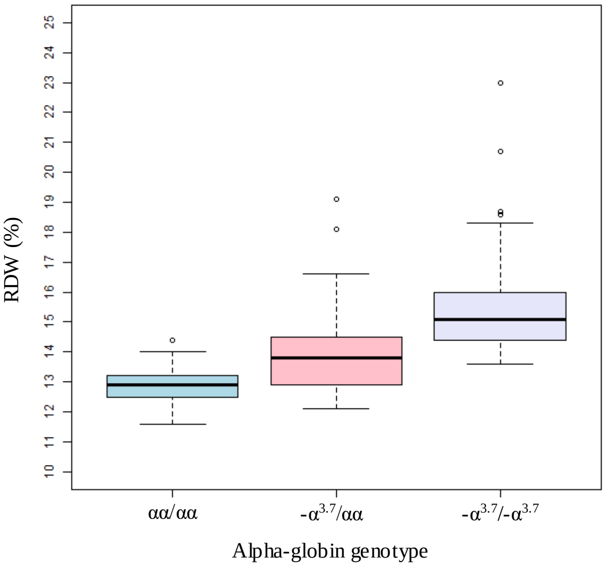


Figure 4.5 – Boxplot of the RDW according to the α-globin genotype group: αα/αα, $-\alpha^{3.7}/\alpha\alpha$, and $-\alpha^{3.7}/-\alpha^{3.7}$. These representations include both females and males. Lower and upper box boundaries represent 25th and 75th percentiles, respectively, the bold line inside the boxes represent the median, lower and upper error lines represent the 10th and 90th percentiles, and the white circles correspond to data falling out of the 10th and 90th percentile (outliers).

In order to investigate a possible association between the presence of α-thalassemia and specific alterations in the hematological parameters we compared the values available simultaneously for 83 individuals with α-thalassemia with a group of 47 individuals without this condition (Table 4.1).

Table 4.1 - Association between the hematological parameters of our population with and without α-thalassemia

Hematological parameters	Individuals with α-thalassemia* (n = 83)				Individuals without α-thalassemia (n = 47)				p_value
	Mean	Median	Min	Max	Mean	Median	Min	Max	
RBC (x 10^{12}/L)									
Females	4.99 ± 0.06	4.89	4.29	6.08	4.05 ± 0.07	4.04	3.53	4.59	< **0.001**
Males	5.90 ± 0.09	5.92	4.76	6.71	4.57 ± 0.06	4.50	4.07	5.07	< **0.001**
Hb (g/dL)									
Females	12.0 ± 0.2	12.1	9.9	14.8	13.1 ± 0.1	13.0	12.2	14.1	< **0.001**
Males	14.2 ± 0.2	14.2	11.8	16.5	14.9 ± 0.2	14.9	13.3	16.1	**0.001**
MCV (fL)	76.4 ± 0.6	75.7	65.2	86.9	92.9 ± 0.6	93.5	83.2	102.1	< **0.001**
MCH (pg)	24.2 ± 0.2	23.7	20.8	26.9	32.5 ± 0.3	32.5	27.9	36.4	< **0.001**
RDW (%)	14.4 ± 0.1	14.3	12.1	16.7	12.8 ± 0.1	12.9	11.6	14.0	< **0.001**

***This group includes 39 individuals $-\alpha^{3.7}/\alpha\alpha$ and 44 $-\alpha^{3.7}/-\alpha^{3.7}$**; **RBC** – Red Blood Cells; **Hb** – Hemoglobin; **MCV** – Mean Corpuscular Volume; **MCH** – Mean Corpuscular Hemoglobin; **RDW** – Red Cell Distribution Width; **Min** – Minimum; **Max** – Maximum. All comparisons were performed using T Test. We considered a statistically significant difference when *p*_value < 0.05 (highlighted in bold).

Through the statistical analysis, we found significant differences in RBCs of both males and females (*p*_value < 0.001), Hb for males (*p*_value = 0.001) and females (*p*_value < 0.001), MCV (*p*_value < 0.001), MCH (*p*_value < 0.001), and RDW (*p*_value < 0.001) between individuals with α-thalassemia and individuals without α-thalassemia. Besides that, we can also see that having α-thalassemia is associated with a higher count of RBCs for both males and females, which is denominated erythrocytosis. Erythrocytosis occurs when erythropoiesis is stimulated to produce a greater number of RBCs and, consequently, compensate for the fact that these cells are smaller and contain less Hb than normal [54]. Besides that, having α-thalassemia is associated with lower levels of Hb for males and females, lower MCV, and lower MCH. We can also see that individuals with α-thalassemia have a higher RDW because the RBCs present more variability in their sizes – anisocytosis.

After removing the outliers, all individuals with α-thalassemia present with hypochromia (MCH < 27 pg), with a minimum of 20.8 pg and a maximum of 26.9 pg. However, the same thing does not happen with the MCV. Although our population has a median of MCV equal to 75.7 fL, which is lower than the value considered for microcytosis (MCV < 78 fL), not all individuals present with this condition (minimum of 65.2 fL and a maximum of 86.9 fL). Thus, it is possible to conclude that, in general, the MCH is a more reliable parameter than the MCV to distinguish individuals with α-thalassemia from those without this pathology.

It is also important to note that usually individuals with the 3.7kb deletion in heterozygosity are asymptomatic and present normal hematological parameters [8,18]. However, our population with α-thalassemia only includes the individuals who arrived at our lab due to an abnormal phenotype that needed to be clarified. Thus, they are not representative of the entire population, which can explain why our results show such large differences between the individuals with and without α-thalassemia.

4.3. Hemoglobinopathies involving the α-globin genes

Seventeen clinical cases arrived at the lab for the diagnosis of complex or uncommon hemoglobinopathies. Eight of these patients were suspected of having a hemoglobin variant of α-chain (**cases 1 to 7, and 11**). Another eight were suspected of having α-thalassemia due to changes in their hematological parameters (**cases 8 to 10, and 12 to 16**), and one patient arrived at the lab with the suspicion of presenting both α-thalassemia and an α-chain hemoglobin variant (**case 17**). All of these patients' hematological parameters are included in Table 4.2.

Table 4.2 – Hematological parameters available for 16 clinical cases of hemoglobinopathies involving the α-globin genes

Case number	RBC (x 10^{12}/L)	Hb (g/dL)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)	Suspected
1	4.74	13.7	85.9	28.9	33.7	15.0	Hb Variant
2	4.29	12.3	91.1	28.7	31.5	12.9	Hb Variant
3	4.48	11.4	82.8	25.4	30.7	13.9	Hb Variant
4	5.40	11.7	74.4	21.7	29.2	20.8	Hb Variant
5	5.27	14.7	96.0	27.9	29.1	14.6	Hb Variant
6	5.02	13.1	81.0	26.0	32.1	15.4	Hb Variant
7	3.94	12.9	95.7	32.8	34.3	12.7	Hb Variant
8	4.83	12.7	79.5	26.3	33.1	13.1	α-thalassemia
9	4.60	12.9	78.0	28.0	35.9	12.5	α-thalassemia
10	2.61	8.5	92.5	32.5	35.2	16.9	α-thalassemia
11	4.52	10.3	71.8	22.9	31.9	15.2	Hb Variant
12	6.08	12.7	68.3	20.9	30.6	14.8	α-thalassemia
13	5.40	14.4	79.6	26.4	33.2	14.3	α-thalassemia
14	5.41	11.9	67.4	21.9	32.5	16.8	α-thalassemia
15	5.31	10.8	64.7	20.3	31.4	16.1	α-thalassemia
16	5.29	10.7	68.9	20.3	29.4	14.4	α-thalassemia
17	4.32	9.7	69.7	22.5	32.2	14.1	Hb Variant and α-thalassemia

RBC – Red Blood Cells; **Hb** – Hemoglobin; **MCV** – Mean Corpuscular Volume; **MCH** – Mean Corpuscular Hemoglobin; **MCHC** – Mean Corpuscular Hemoglobin Concentration; **RDW** – Red Cell Distribution Width.

In total, thirteen patients present with anemia (Hb < 12 g/dL for females and Hb < 13 g/dL for males), microcytosis (MCV < 78 fL) and/or hypochromia (MCH < 27 pg) (Table 4.2). Each clinical case will be presented either individually or in groups based on the outcome of the diagnosis.

4.3.1. Diagnosis of α-chain hemoglobin variants

The initial presumptive diagnosis of α-chain hemoglobin variants was performed for nine samples using biochemical methodologies, such as HPLC and/or capillary zone electrophoresis. These diagnoses were performed using peripheric blood from the individuals in "Unidade de Diagnóstico Laboratorial e Referência" of DPS/INSA, and revealed abnormal hemoglobin migration patterns. In Figure 4.6 it is possible to see an example of both biochemical techniques (executed in DPS) for an individual with an α-chain Hb variant, presumably called Hb Setif, which will be discussed later on this dissertation.

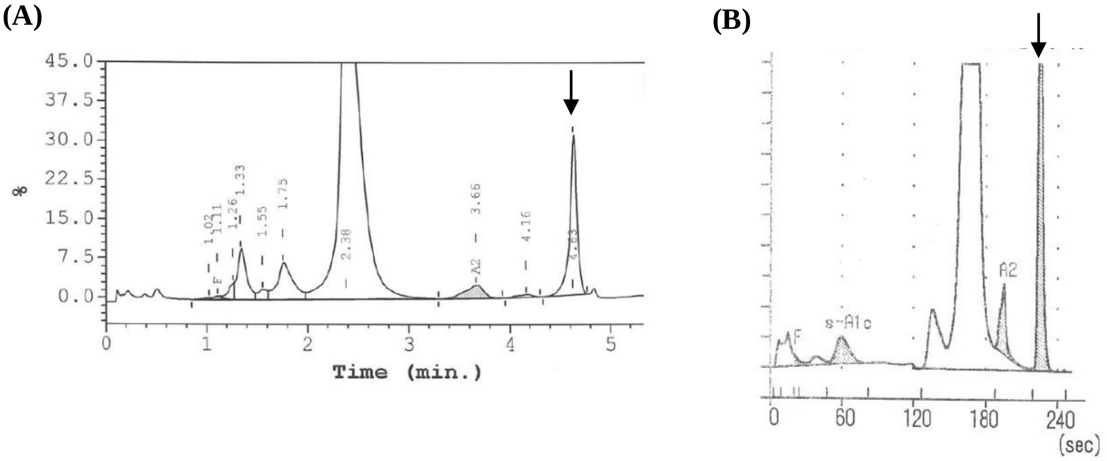


Figure 4.6 – Biochemical characterization of a hemoglobin variant: presumably Hb Setif. Example of biochemical results obtained through **(A)** High Performance Liquid Chromatography and **(B)** capillary zone electrophoresis. A black arrow highlights the Hb variant detected in both profiles. These figures were kindly provided by DPS for illustration purposes.

To perform the molecular diagnosis of these uncommon cases, each of the α-globin genes were amplified individually in a conventional PCR technique as described in the Materials and methods section. This methodology uses two distinct mixtures, only differing in the reverse primer, and therefore allowing the *HBA2* and *HBA1* genes to be amplified with mix I and mix II, respectively. The amplicons were visualized in a 1% agarose gel through the presence of a 1 kb band in both mixtures (Figure 4.7).

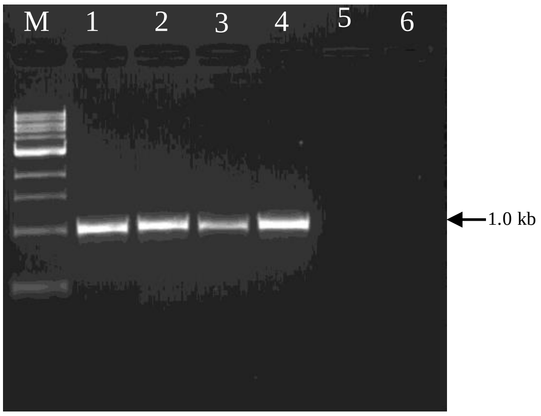
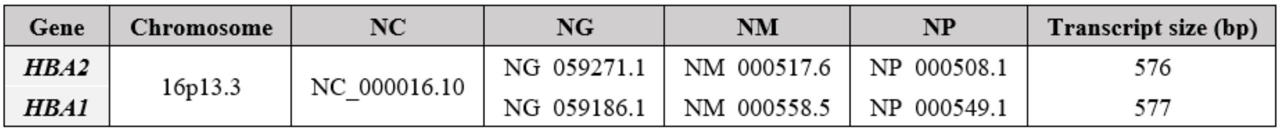


Figure 4.7 – Visualization of the amplified DNA fragments corresponding to *HBA2* and *HBA1* genes. Electrophoretic profile of the 1 kb amplified fragment in a 1% (m/v) agarose gel, in an electrophoretic system at 70V during 60 minutes. **M** – molecular weight marker (1 kb DNA Ladder, BioLabs); **1 and 3** – PCR product corresponding to the *HBA2* gene using mix I; **2 and 4** – PCR product corresponding to the *HBA1* gene using mix II; **5 and 6** – negative control for mix I and II, respectively. The size of the fragment in kb is included on the right side of the image.

Following the PCR reaction and purification of the PCR products, Sanger sequencing was used to sequence both amplicons using different primers from those used in the amplification (Table 7.16, Supplementary materials). The obtained *HBA2* and *HBA1* sequences were compared to their references (Table 4.3) in the *RefSeq* database (https://www.ncbi.nlm.nih.gov/refseq/).

Table 4.3 - Reference sequences of the *HBA2* and *HBA1* genes from *RefSeq* used in the comparative analysis of the results obtained from Sanger Sequencing

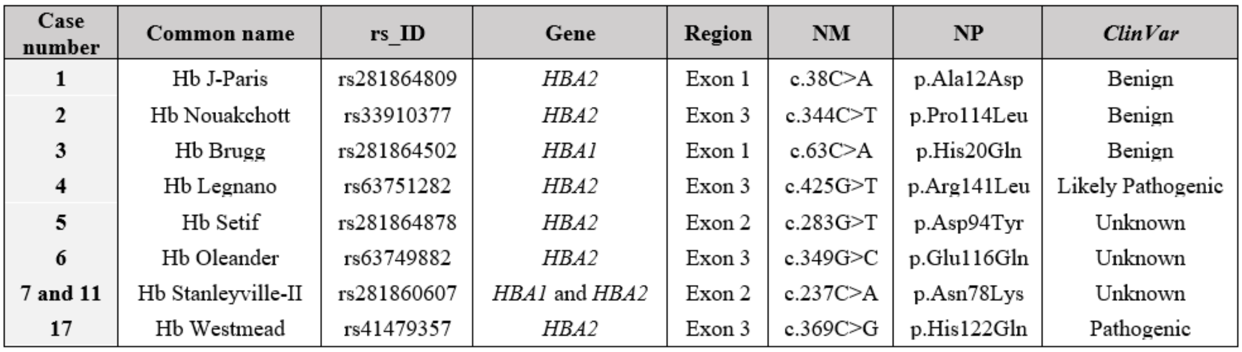


Gene	Chromosome	NC	NG	NM	NP	Transcript size (bp)
HBA2	16p13.3	NC_000016.10	NG 059271.1	NM 000517.6	NP 000508.1	576
HBA1			NG 059186.1	NM 000558.5	NP 000549.1	577

NC – chromosome reference sequence; **NG** – gene reference sequence; **NM** – mRNA reference sequence; **NP** – protein reference sequence; **bp** – base pairs.

In the nine investigated cases, *HBA2* and *HBA1* Sanger sequencing analyses revealed eight missense variants in coding regions that confirmed the suspected α-chain hemoglobin variants (Table 4.4). In a first approach, the clinical significance of all variants was evaluate using the *ClinVar* database, which classifies the variants as benign, unknown pathogenicity, likely pathogenic or pathogenic, according to previous described clinical cases.

Table 4.4 – Variants found through Sanger Sequencing in the *HBA2* and *HBA1* genes in nine clinical cases



Case number	Common name	rs_ID	Gene	Region	NM	NP	*ClinVar*
1	Hb J-Paris	rs281864809	*HBA2*	Exon 1	c.38C>A	p.Ala12Asp	Benign
2	Hb Nouakchott	rs33910377	*HBA2*	Exon 3	c.344C>T	p.Pro114Leu	Benign
3	Hb Brugg	rs281864502	*HBA1*	Exon 1	c.63C>A	p.His20Gln	Benign
4	Hb Legnano	rs63751282	*HBA2*	Exon 3	c.425G>T	p.Arg141Leu	Likely Pathogenic
5	Hb Setif	rs281864878	*HBA2*	Exon 2	c.283G>T	p.Asp94Tyr	Unknown
6	Hb Oleander	rs63749882	*HBA2*	Exon 3	c.349G>C	p.Glu116Gln	Unknown
7 and 11	Hb Stanleyville-II	rs281860607	*HBA1* and *HBA2*	Exon 2	c.237C>A	p.Asn78Lys	Unknown
17	Hb Westmead	rs41479357	*HBA2*	Exon 3	c.369C>G	p.His122Gln	Pathogenic

rs_ID – rs identification; **NM** – mRNA reference sequence; **NP** – protein reference sequence.

In this dissertation we will go into further detail in the variants that were classified as pathogenic, likely pathogenic or with unknown pathogenicity. Thus, Hb J-Paris, Hb Nouakchott, and Hb Brugg, corresponding to **cases 1, 2, and 3** in Table 4.2 will not be discussed. It is important to note that Hb Stanleyville-II and Hb Westmead were found simultaneously with other hemoglobinopathies and, therefore, will be discussed in detail in the "Clinical cases involving deletions and Hb variants of α-chain" section of this dissertation.

In addition, we also detected one substitution at the 3'UTR of the *HBA2* gene (*HBA2*:c.*107A>G) that was present in all the nine samples studied. This alteration is a polymorphism not associated with any thalassemia features.

4.3.1.1. Clinical cases involving only Hb variants of α-chain

4.3.1.1.1. Heterozygous for Hb Legnano

A 65-year-old male (**case 4** in Table 4.2) exhibiting microcytosis (MCV = 74.4 fL), hypochromia (MCH = 21.7 pg), anemia (Hb = 11.7 g/dL), low MCHC (29.2 g/dL), and elevated RDW (20.8%) was studied by genetic molecular methodologies variant after HPLC revealed 25% of a suspected Hb variant.

Sanger sequencing of the *HBA2* gene revealed a mutation in heterozygosity in codon 141 (*HBA2*:c.425G>T) (Figure 4.8) leading to a substitution of arginine for leucine, which gives rise to a hemoglobin variant called **Hb Legnano** [55–57].

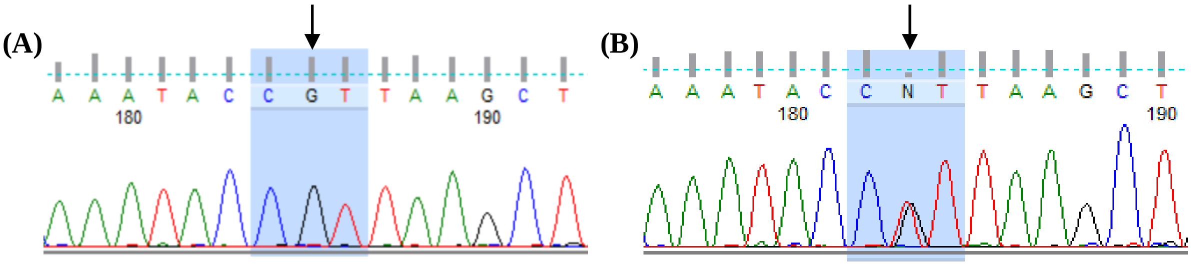
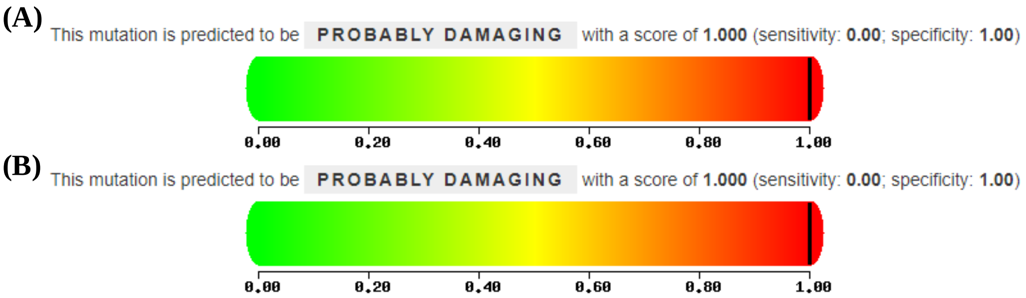


Figure 4.8 – Molecular identification of a hemoglobin variant: Hb Legnano. Partial Sanger sequencing eletropherogram of the *HBA2* gene at exon 3 showing **(A)** the wild-type and **(B)** the mutation *HBA2*:c.425G>T (p.Arg141Leu) in heterozygosity responsible for Hb Legnano. The black arrows indicate the wild type and mutated positions and codon 141 is highlighted in a blue shaded box.

In silico studies using software *PolyPhen-2* and SIFT, predict this variant to be probably damaging at a protein level, with a score of 1.0 (Figure 4.9) and 0.0, respectively.


(A) This mutation is predicted to be PROBABLY DAMAGING with a score of **1.000** (sensitivity: **0.00**; specificity: **1.00**)

(B) This mutation is predicted to be PROBABLY DAMAGING with a score of **1.000** (sensitivity: **0.00**; specificity: **1.00**)

Figure 4.9 – Prediction of the pathogenic effect of the p.Arg141Leu in the α-globin protein through the *PolyPhen-2* software. (A) The *HumDiv* model predicts the mutation to be probably damaging. **(B)** The *HumVar* model predicts the mutation to be probably damaging.

In Hb Legnano the amino acid substitution happens in the α-globin chain C-terminus, a critical position responsible for the formation of salt bridges with neighboring amino acids involved in the structure stabilization of the deoxyhemoglobin (deoxyHb). Thus, a disruption in these salt bridges reduces the stability of deoxyHb, resulting in an increased oxygen affinity. Due to a lower oxygen delivery in the tissues, a compensatory erythrocytosis takes place in order to rise the hemoglobin concentration [55–57].

This Hb variant was previously described in individuals from the Italian [55], Japanese, and Chinese [57] populations. In these clinical cases, the presence of Hb Legnano in heterozygosity was associated with abnormal RBC indices, as well as microcytosis (low MCV), anisocytosis (high RDW), hypochromia (low MCH) and mild polycythemia (high RBC) [55]. Therefore, the altered hematological parameters in our patient can be explained by the presence of Hb Legnano, which has now been found in a patient with Portuguese nationality.

4.3.1.1.2. Heterozygous for Hb Setif

A 51-year-old male (**case 5** in Table 4.2) arrived at the lab for the molecular identification of an Hb variant of α-chain. His HPLC profile analysis revealed an abnormal peak of a putative 12.1% structural Hb variant. This patient presented all hematological parameters within the normal range, except for a low MCHC of 29.1 g/dL.

The obtained eletropherogram of Sanger sequencing for the α2 gene revealed a substitution in heterozygosity in codon 94 (*HBA2*:c.283G>T) (Figure 4.10), which leads to an alteration of the aspartic acid in that position for a tyrosine, and, consequently, gives rise to a Hb variant designated **Hb Setif** [58–60].

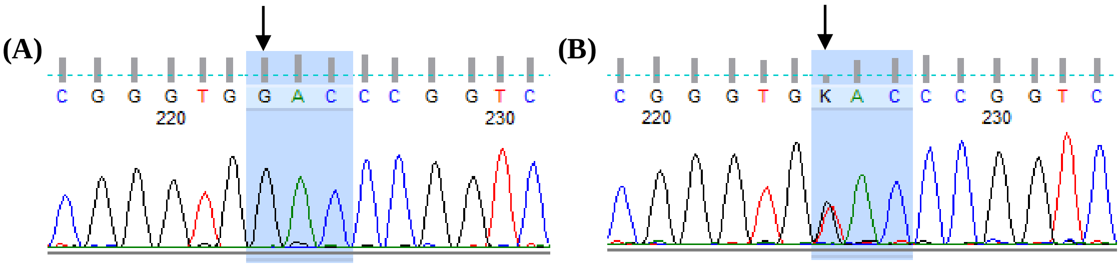
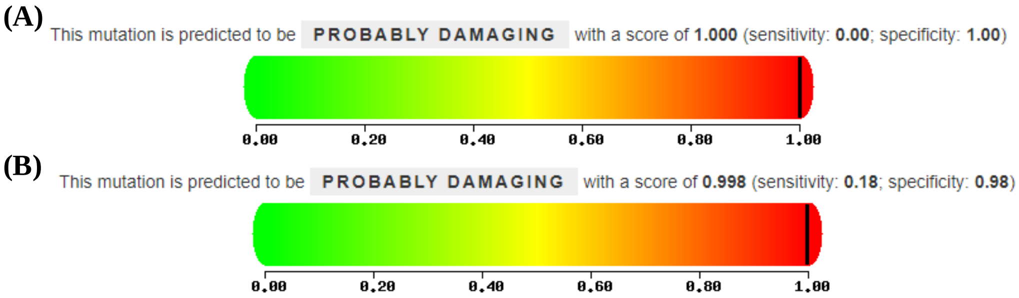


Figure 4.10 – Molecular identification of a hemoglobin variant: Hb Setif. Partial Sanger sequencing eletropherogram of the *HBA2* gene at exon 2 showing **(A)** the wild-type and **(B)** the mutation *HBA2*:c.283G>T (p.Asp94Tyr) in heterozygosity responsible for Hb Setif. The black arrows indicate the wild type and mutated positions and codon 94 is highlighted in a blue shaded box.

PolyPhen-2 and SIFT software were used to perform studies *in silico* to understand the impact of this nucleotide substitution at the protein level, and both predict Hb Setif to be probably damaging, with a score of 1.0 (Figure 4.11) and 0.0, respectively.


(A) This mutation is predicted to be **PROBABLY DAMAGING** with a score of **1.000** (sensitivity: **0.00**; specificity: **1.00**)

(B) This mutation is predicted to be **PROBABLY DAMAGING** with a score of **0.998** (sensitivity: **0.18**; specificity: **0.98**)

Figure 4.11 – Prediction of the pathogenic effect of the p.Asp94Tyr in the α-globin protein through the *PolyPhen-2* software. (A) The *HumDiv* model predicts the mutation to be probably damaging. **(B)** The *HumVar* model predicts the mutation to be probably damaging.

Hb Setif was first described in an Algerian family, and since then, there have been several reports from the Middle East and European populations, including Greece [58], Iran [59,61], Saudi Arabia [62,63], Turkey [64], Malta [65,66], Italy [67], and Spain [68].

The evolutionary conservation profile of the acid aspartic at position 94 (Figure 4.12) revealed that this amino acid is conserved within similar mammal species. Codon 94 of the α2-globin gene is located in a position prone to the formation of a DNA hairpin, which potentially can lead to a variety of acid aspartic substitutions with other amino acids [59]. Besides that, this position is very important in the $\alpha_1\beta_2$ contact [61,69], because it is the only polar interaction in the $\alpha_1\beta_2$ interface of the oxyhemoglobin (oxyHb) conformation. Therefore, a substitution of the acid aspartic for tyrosine leads to an unstable hemoglobin with oxygen affinity below the normal range [59–61].

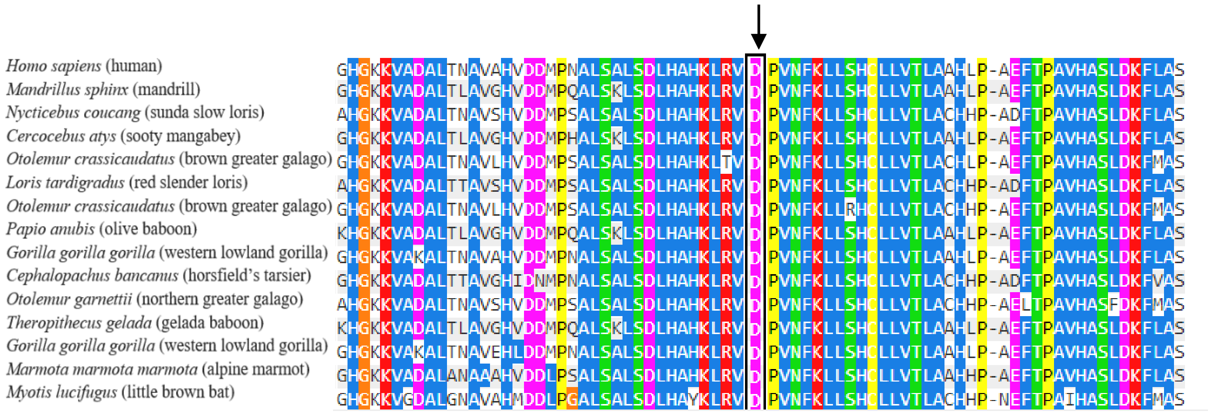


Figure 4.12 – Evolutionary conservation profile of the p.Asp94 amino acid in the α-globin protein. Multiple amino acid sequence alignment of homolog proteins of 15 related mammal species, obtained through the *PolyPhen-2* software. The amino acid substitution pattern is highlighted with an arrow and inside a black box, in which the letter D represents aspartic acid.

Despite *in silico* studies and the evolutionary conservation profile indicate that Hb Setif can have an adverse clinical impact, carriers usually present with hematological indices on the low limit of the normal range [59]. This is probably due to the low percentage of Hb variant observed in heterozygous individuals (between 12.0 and 15.0%) [60,61]. Therefore, our patient's phenotype can be explained by the presence of Hb Setif.

4.3.1.1.3. Heterozygous for Hb Oleander

A 39-year-old female (**case 6** in Table 4.2) arrived at the lab after the analysis of her HPLC profile revealed an abnormal peak of a 23.3% structural Hb variant. Besides that, she presented all hematological parameters within the normal range, with the exception for a slight hypochromia (MCH = 26.0 pg).

Sanger sequencing eletropherogram of the *HBA2* gene revealed a substitution in heterozygosity of nucleotide 349 (*HBA2*:c.349G>C) (Figure 4.13). This substitution occurs in codon 116 and leads to a substitution of glutamic acid for glutamine in the α-globin chain, which consequently gives rise an Hb variant called **Hb Oleander** [70].

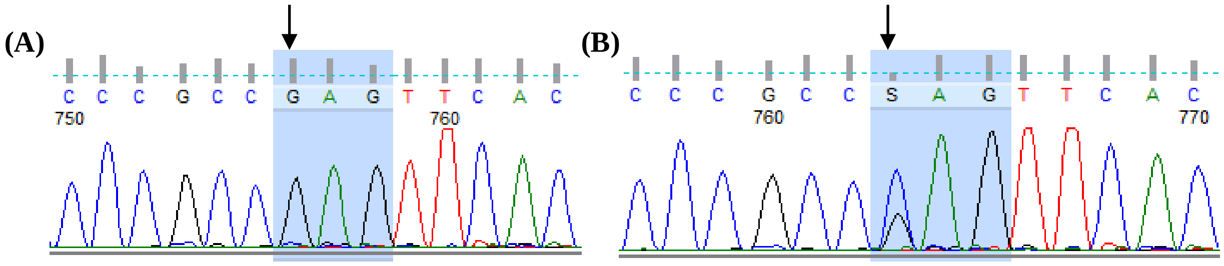
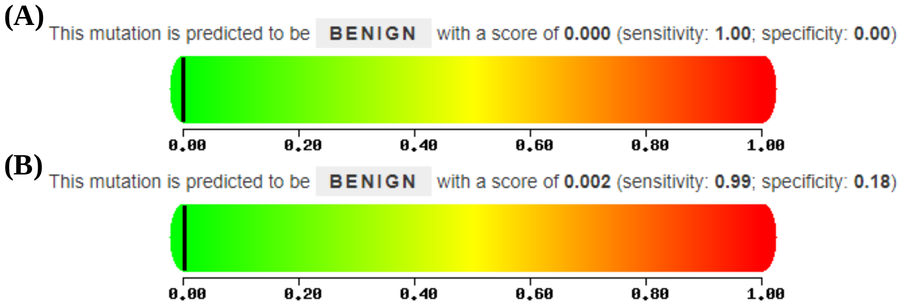


Figure 4.13 – Molecular identification of a hemoglobin variant: Hb Oleander. Partial Sanger sequencing eletropherogram of the *HBA2* gene at exon 3 showing **(A)** the wild-type and **(B)** the mutation *HBA2*:c.349G>C (p.Glu116Gln) in heterozygosity responsible for Hb Oleander. The black arrows indicate the wild type and mutated positions and codon 116 is highlighted in a blue shaded box.

In silico studies using the software *PolyPhen-2* and SIFT predict Hb Oleander as benign at the protein level with a score of 0.0 (Figure 4.14) and 0.056, respectively.



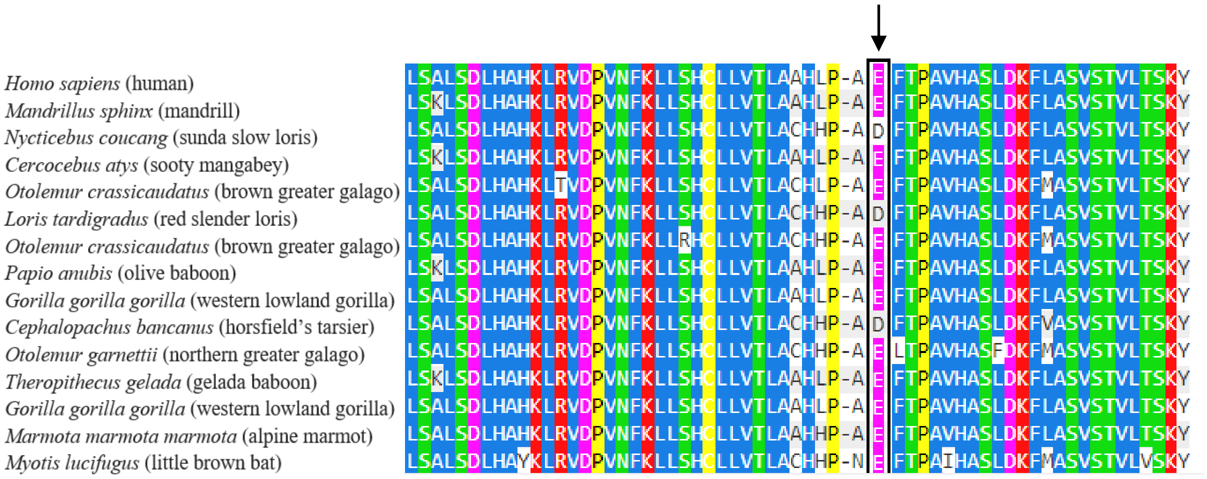
This mutation is predicted to be **BENIGN** with a score of **0.000** (sensitivity: **1.00**; specificity: **0.00**)

This mutation is predicted to be **BENIGN** with a score of **0.002** (sensitivity: **0.99**; specificity: **0.18**)

Figure 4.14 – Prediction of the pathogenic effect of the p.Glu116Gln in the α-globin protein through the *PolyPhen-2* software. (A) The *HumDiv* model predicts the mutation to be benign. **(B)** The *HumVar* model predicts the mutation to be benign.

The amino acid substitution that happens in Hb Oleander results in the loss of one carboxyl group in each α-chain, affecting the electrophoretic mobility of this Hb variant. Despite this, oxygen equilibrium curve studies of RBCs containing Hb Oleander revealed no differences between this fraction and those of normal hemoglobin [70].

The evolutionary conservation profile of the glutamic acid at position 116 (Figure 4.15) revealed that this residue is conserved among similar mammal species and that for some of those it can be replaced by an aspartic acid, due to the similar biochemical properties of both amino acids (negative charged side chains). The polar properties and uncharged side chains of glutamine can explain why the substitution of the glutamic acid for that residue is considered benign in terms of pathogenicity.



Homo sapiens (human)	LSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A E FTPAVHASLDKFLASVSTVLTSKY
Mandrillus sphinx (mandrill)	LSKLSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A E FTPAVHASLDKFLASVSTVLTSKY
Nycticebus coucang (sunda slow loris)	LSALSDLHAHKLRVDPVNFKLLSHCLLVTLACHHP-A D FTPAVHASLDKFLASVSTVLTSKY
Cercocebus atys (sooty mangabey)	LSKLSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A E FTPAVHASLDKFLASVSTVLTSKY
Otolemur crassicaudatus (brown greater galago)	LSALSDLHAHKLTVDPVNFKLLSHCLLVTLACHLP-A E FTPAVHASLDKFMASVSTVLTSKY
Loris tardigradus (red slender loris)	LSALSDLHAHKLRVDPVNFKLLSHCLLVTLACHHP-A D FTPAVHASLDKFLASVSTVLTSKY
Otolemur crassicaudatus (brown greater galago)	LSALSDLHAHKLRVDPVNFKLLRHCLLVTLACHHP-A E FTPAVHASLDKFMASVSTVLTSKY
Papio anubis (olive baboon)	LSKLSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A E FTPAVHASLDKFLASVSTVLTSKY
Gorilla gorilla gorilla (western lowland gorilla)	LSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A E FTPAVHASLDKFLASVSTVLTSKY
Cephalopachus bancanus (horsfield's tarsier)	LSALSDLHAHKLRVDPVNFKLLSHCLLVTLACHHP-A D FTPAVHASLDKFVASVSTVLTSKY
Otolemur garnettii (northern greater galago)	LSALSDLHAHKLRVDPVNFKLLSHCLLVTLACHHP-A E LTPAVHASFDKFMASVSTVLTSKY
Theropithecus gelada (gelada baboon)	LSKLSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A E FTPAVHASLDKFLASVSTVLTSKY
Gorilla gorilla gorilla (western lowland gorilla)	LSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A E FTPAVHASLDKFLASVSTVLTSKY
Marmota marmota marmota (alpine marmot)	LSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHP-A E FTPAVHASLDKFLASVSTVLTSKY
Myotis lucifugus (little brown bat)	LSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHP-N E FTPAIHASLDKFMASVSTVLVSKY

Figure 4.15 – Evolutionary conservation profile of the p.Glu116 amino acid in the α-globin protein. Multiple amino acid sequence alignment of homolog proteins of 15 related mammal species, obtained through the *PolyPhen-2* software. The amino acid substitution pattern is highlighted with an arrow and inside a black box, in which the letter E represents glutamic acid and D represents aspartic acid.

Hb Oleander was first described in a newborn black infant and his mother [70] and, since then, was discovered in a 16-year-old black female from Georgia [71]. Clinical and functional studies show that Hb Oleander has no adverse biological effects in these cases, besides minor morphologic changes in RBCs [70]. Therefore, the slight hypochromia presented by our patient can be explained by the presence of Hb Oleander. It is important to note that our patient was born in Guinea, which explains the diagnosis of an Hb variant with African ancestry.

4.3.2. Alpha-thalassemia diagnosis

The diagnosis of α-thalassemia was performed using three different techniques: Gap-PCR, Multiplex Gap-PCR, and MLPA, as described in the Materials and methods section. These analyses were applied to nine samples suspected of having α-thalassemia and two individuals with the Hb Stanleyville-II, a hemoglobin variant of α-chain that can be associated with the 3.7kb deletion.

4.3.2.1. Search of common α-thalassemia deletions

Firstly, we performed Gap-PCR to detect the presence of the 3.7kb deletion, the most common α-thalassemia deletion, which removes part of the α2- and α1-globin genes and gives rise to a functional hybrid $\alpha_2\alpha_1$ gene, as represented in Figure 3.1 of the Materials and methods section. This technique requires two different PCR mixtures, called mix I and mix II, that only differ in the reverse primer. Mix I is used for the amplification of the *HBA2* gene, while mix II allows both the *HBA1* gene and the hybrid gene to be amplified.

After the PCR reaction, the electrophoretic profile of each amplified PCR product was visualized in a 1% agarose gel (Figure 4.16). In a normal homozygous individual (αα/αα), there is a 1.9 kb band in mix I and a 2.1 kb band in mix II. A carrier of the 3.7kb deletion ($-\alpha^{3.7}/\alpha\alpha$) will give rise to a 1.9 kb band in mix I and two bands of 1.9 kb and 2.1 kb in mix II. Lastly, if a patient has the deletion in homozygosity ($-\alpha^{3.7}/-\alpha^{3.7}$), it is possible to see one band of 1.9 kb in mix II, while there is no band in mix I.

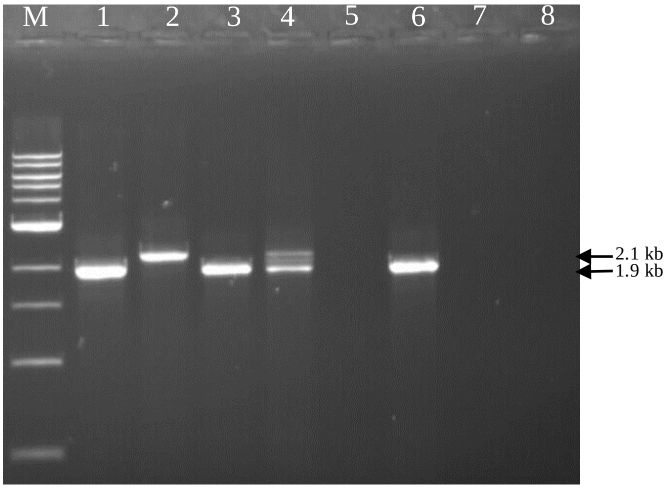


Figure 4.16 – Visualization of the DNA fragments obtained in a Gap-PCR experiment for the diagnosis of the 3.7kb deletion. Electrophoretic profile of the three possible genotypes that can be obtained through the detection of the 3.7kb deletion in a 1% (m/v) agarose gel, in an electrophoretic system at 70V during 75 minutes. **M** – molecular weight marker (1kb DNA Ladder, BioLabs); **1 and 2** - αα/αα individual with a 1.9 kb band in Mix I and a 2.1 kb band in Mix II; **3 and 4** - $-\alpha^{3.7}/\alpha\alpha$ individual with a 1.9 kb band in Mix I and two bands of 1.9 and 2.1 kb in Mix II; **5 and 6** - $-\alpha^{3.7}/-\alpha^{3.7}$ individual with a 1.9 kb band in Mix II; **7 and 8** – negative control. The size of the fragment in kb is included on the right side of the image.

One of the secondary aims of this study was to implement in our lab a Multiplex Polymerase Chain Reaction Assay that allows the simultaneous detection of the 3.7kb and 4.2kb deletions, since these are the two most frequent deletional causes of α-thalassemia in our population. This methodology is described in the Material and Methods section and was implemented based on the technique described by Tan and collaborators in 2001 [43]. This approach requires the use of three pairs of primers, two for each deletion and one for the detection of the *HBA2* gene as described in Figure 3.3 of the Materials and methods section.

The amplified PCR products obtained in this technique were visualized in a 1% agarose gel (Figure 4.17). In this methodology, the presence of a single 1.8 kb fragment indicates a normal individual (genotype αα/αα). If an allele with a deletion is also presence, it is possible to observe an extra fragment of 2.1 kb in the presence of a heterozygosity for the 3.7kb deletion, and a fragment of 1.6 kb in the presence of the 4.2kb deletion, both in heterozygosity.

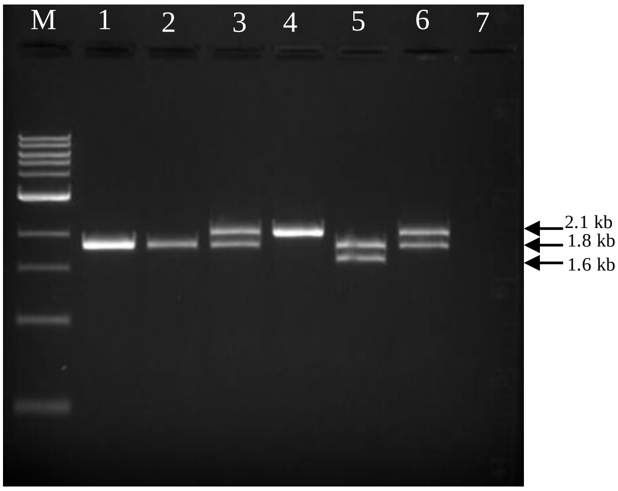


Figure 4.17 – Visualization of the DNA fragments obtained in a Multiplex Gap-PCR experiment for the simultaneous diagnosis of the 3.7kb and 4.2kb deletions. Electrophoretic profile of four genotypes obtained through the detection of the 3.7kb and 4.2kb deletions in a 1% (m/v) agarose gel, in an electrophoretic system at 70V during 90 minutes. **M** – molecular weight marker (1 kb DNA Ladder, BioLabs); **1 and 2** – αα/αα individuals with a 1.8 kb fragment; **3 and 6** – $-\alpha^{3.7}/\alpha\alpha$ individuals with two fragments of 1.8 kb and 2.1 kb; **4** – $-\alpha^{3.7}/-\alpha^{3.7}$ individual with a single fragment of 2.1kb; **5** – $-\alpha^{4.2}/\alpha\alpha$ individual with two fragments of 1.6 kb and 1.8 kb; **7** – negative control. The size of the fragments in kb is included on the right side of the image.

After optimization of the Multiplex Gap-PCR, we were able to use it in the diagnosis of two individuals, one with the 3.7kb deletion and other with the 4.2kb deletion, both in heterozygosity. Although this technique requires a larger number of primers, it only calls for the use of one mixture, and a single PCR reaction allows the detection of multiple common deletions, which is a great advantage when compared to Gap-PCR.

In the 11 clinical cases studied for α-thalassemia through these PCR techniques, we found four heterozygous ($-\alpha^{3.7}/\alpha\alpha$) (**cases 7 to 10** in Table 4.2) and two homozygous ($-\alpha^{3.7}/-\alpha^{3.7}$) (**cases 11 and 12** in Table 4.2) for the 3.7kb deletion, as well as one heterozygous ($-\alpha^{4.2}/\alpha\alpha$) for the 4.2kb deletion (**case 13** in Table 4.2).

The 3.7kb and 4.2kb α-thalassemia deletions are the two most common types of α^{+}-thalassemia [5,13]. Both of these deletions are prevalent in Middle Eastern, Mediterranean, and Asian populations, but the 3.7kb deletion is also very common in African populations [72]. A study conduct in 1995 revealed that these deletions are the two most frequent in Portugal [37]. The 3.7kb deletion occurs due to an unequal recombination between homologous segments called Z boxes that are 3.7 kb apart, resulting in the formation of a chromosome with a hybrid α-globin gene. On the other hand, the 4.2kb deletion happens through the same mechanism but between X boxes that are 4.2 kb apart [4,13,18].

Individuals with a single α-globin gene deletion are typically asymptomatic [8,18]; however this was only the case for one of our clinical cases, who was diagnosed with the 3.7kb deletion in heterozygosity and had all hematological parameters within normal range (**case 7** in Table 4.2). Despite this, individuals with a -α/αα genotype can present with mild microcytosis and/or hypochromia [8,18], as was the case for four of our patients with a single gene deletion (three with a $-\alpha^{3.7}/\alpha\alpha$ genotype and one with a $-\alpha^{4.2}/\alpha\alpha$ genotype).

When it comes to individuals with two functional α-genes, the CBC usually reveals both microcytosis and hypochromia [8,18]. This was also true for our patients that were diagnosed with the 3.7kb deletion in homozygosity, as both of them presented with low levels of MCV and MCH.

4.3.2.2. Search of rare α-thalassemia deletions

For the four individuals that did not present with the 3.7kb nor the 4.2kb deletions we searched for CNVs in the α-globin gene cluster through MLPA using the SALSA® MLPA® probemix P140-C1 HBA kit (*MRC-Holland*). After the analysis of the obtained MLPA profiles, we used the *IthaCNVs* database (https://www.ithanet.eu/db/ithacnv) to determine which deletion was more similar to the ones we had found. Thus, with this technique we found four different deletions in heterozygosity similar to deletions that had already been described (Figure 4.18). One was similar to the known Southeast Asian ($--^{SEA}$) deletion, and three that completely remove the regulatory region HS-40 and were similar to the already described $(\alpha\alpha)^{ALT}$, $(\alpha\alpha)^{TI}$, and $(\alpha\alpha)^{MM}$ deletions.

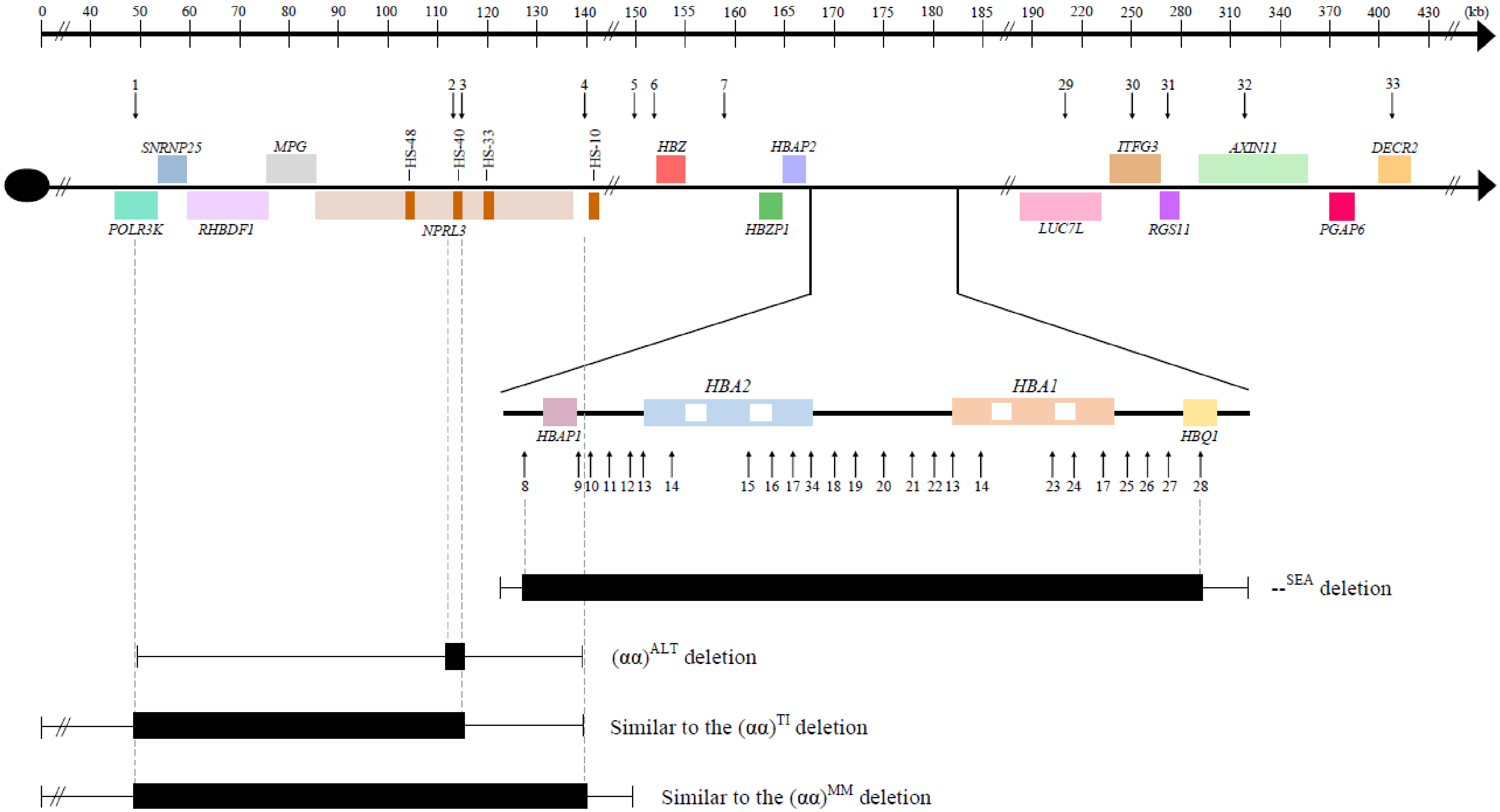


Figure 4.18 – Schematic representation of the deletions diagnosed through MLPA. Representation of a 500 kb from the sub-telomeric region of chromosome 16, containing the α-globin gene cluster, its distal regulatory regions (HS-48, -40, -33, and -10), as well as other non-globinic genes located in this sub-telomeric region. Black arrows indicate MLPA probe hybridization sites and each probe is numbered according to their sequential order of chromosomal hybridization. Black bars represent deleted DNA sequence as determined by MLPA analysis. Thin lines indicate the region of uncertainty for deletion breakpoints. The oval shape represents the telomere.

4.3.2.2.1. Southeast Asia deletion

The first MLPA profile (Figure 4.19) revealed that the proportion of MLPA probes 8 to 28 was reduced to 0.5. This indicates the presence of a large deletion in heterozygosity that completely removes *HBAP1*, *HBA2*, *HBA1*, and *HBQ* genes. As the 5'-breakpoint occurs in an uncertainty region between probes 7 and 8, in which the *HBZP1* and *HBAP2* genes are located is not possible to know if these genes are also removed. This deletion has its 3'-breakpoint in an uncertainty region between probes 28 and 29, and is similar to the previously described **Southeast Asian deletion** (--^{SEA}). Our colleagues from DPS later confirmed this diagnosis through a Gap-PCR methodology.

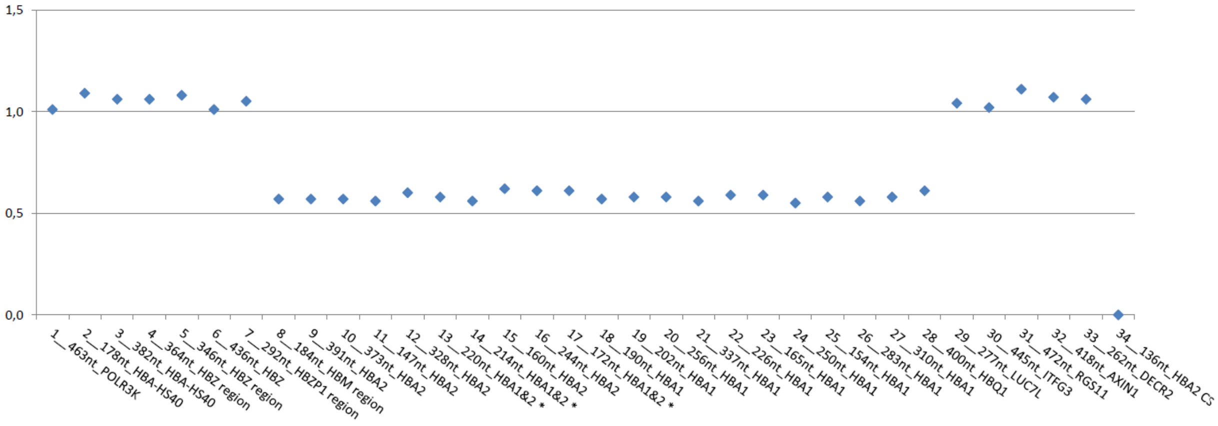


Figure 4.19 – MLPA results of an individual heterozygous for the Southeast Asian deletion (--^{SEA}/αα). Graphic representation of the search for SNVs in the α-globin gene cluster through MLPA. The MLPA probe numbers are indicated on the abscissa. The proportion of each MLPA probe, determined through the comparison of their relative peak with three different reference DNA samples, is indicated in the ordinate. Probe number 1 to 7 and 29 to 33 have the normal ratio of 1.0, while probe 8 to 28 have a hybridization ratio of 0.5 revealing the presence of a deletion in heterozygosity. Probe number 34 only amplifies in the presence of Hb Constant Spring and, therefore, in normal individuals the proportion is zero.

The Southeast Asian deletion (--^{SEA}) is the most common and severe form of α-thalassemia in China and multiple Southeast Asian populations, such as Thailand, Vietnam, The Philippines, Laos, and Cambodia [73–75]. This α^0-thalassemia deletion of about 20.5 kb occurs on chromosome 16 and removes the *HBA2ps*, *HBA1ps*, *HBA2*, *HBA1*, and *HBQ* genes, with its 5'-breakpoint in the 5' end of the *HBA2ps* gene and the 3'-breakpoint upstream of the 3'hypervariable region (HVR) [73,75].

It is important to note that this deletion was found in **individual 17**, who have co-inherited other hemoglobinopathies and, therefore, will be discussed into further detail in the "Clinical cases involving deletions and Hb variants of α-chain" section.

4.3.2.2.2. Deletions that remove the upstream regulatory elements

Three unrelated individuals (**cases 14 to 16** in Table 4.2) arrived at the lab with abnormal hematological parameters but with normal biochemical analysis. Because of that, the phenotype could only be explained by the presence of an α-thalassemia deletion. MLPA results revealed three different deletions that remove one or more of the α-globin gene cluster upstream regulatory elements and that are similar to the (αα)^{ALT}, (αα)^{TI}, and (αα)^{MM} deletions.

Individual 14 is a 5-year-old boy presenting with microcytosis (MCV = 67.4 fL), hypochromia (MCH = 21.9 pg), low levels of HCT (36.5%), and high levels of RDW (16.8%). MLPA results revealed that the proportions of probes 2 and 3 were reduced to 0.5 (Figure 4.20), which indicates the presence of a small heterozygous deletion with both breakpoints within uncertainty regions. The 5'-breakpoint occurs between probes 1 and 2, and the 3'-breakpoint between probes 3 and 4. These results suggest that only the HS-40 site was removed. This deletion of at least 300 bp is similar to the **(αα)^{ALT} deletion** previous described by Coelho *et al.* [12] in three individuals from the Alentejo region in Portugal and by Phylipsen *et al.* [76] in one Australian individuals.

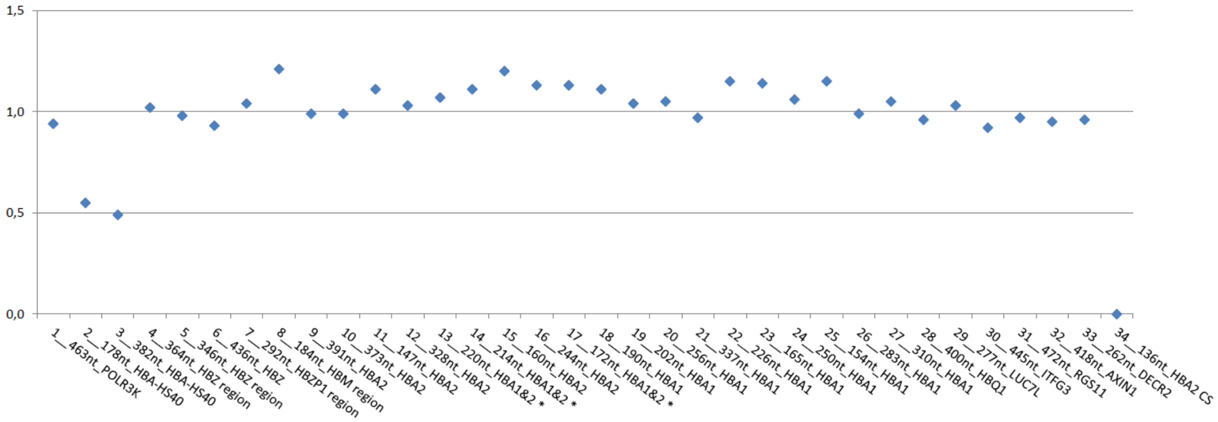


Figure 4.20 – MLPA results for patient number 14 revealing the (αα)^{ALT} deletion in heterozygosity. Graphic representation of the search for CNVs in the α-globin gene cluster through MLPA. The MLPA probe numbers are indicated on the abscissa. The proportion of each MLPA probe, determined through the comparison of their relative peak with three different reference DNA samples, is indicated in the ordinate. Probe number 1 and 4 to 33 have the normal ratio of 1.0, while probe 2 and 3 have a hybridization ratio of 0.5 revealing the presence of a deletion in heterozygosity. Probe number 34 only amplifies in the presence of Hb Constant Spring and, therefore, in normal individuals the proportion is zero.

To confirm the diagnosis of the (αα)^{ALT} deletion, a specific Gap-PCR reaction was performed using the conditions shown in Table 7.5 from the Supplementary materials, and the resulting fragments were visualized in a 1% agarose gel (Figure 4.21). In this technique, the presence of a 1.1 kb fragment indicates an individual with the (αα)^{ALT} deletion. In a normal individual, no PCR fragment is obtained as the primers are located 4.4 kb apart.

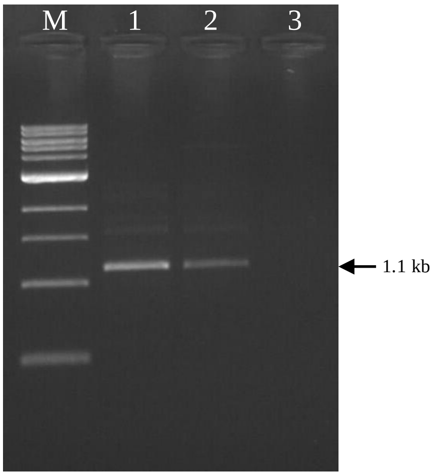


Figure 4.21 – Visualization of the amplified DNA fragment generated from the (αα)^{ALT} deletion by Gap-PCR. Electrophoretic profile of the 1.1 kb fragment in a 1.0% (m/v) agarose gel, in an electrophoretic system at 70V during 60 minutes. **M** – molecular weight marker (1 kb DNA Ladder, BioLabs); **1** – PCR product generated from a previously diagnosed homozygous (αα)^{ALT} deletion case; **2** – PCR product generated from individual 14 revealing the (αα)^{ALT} deletion; **3** – negative control. The size of the fragment in kb is included on the right side of the image.

Through the previous electrophoretic profile it is possible to confirm the diagnosis of (αα)$^{\mathrm{ALT}}$ deletion on our patient (well number 2). This small deletion of 3.3 kb happens within the *NPRL3* gene and only removes the HS-40 region. It is expected that the deletion of this element is replaced by a 39 nt insertion [12,15,76], which is hypothesized to be a sequence that resembles a functionally important motif inside the HS-40 and could be responsible for the levels of α-globin chain synthesis seen in another Portuguese individuals with this deletion [12,15].

Individual 15 is a 10-year-old boy that arrived at the lab with hematological parameters showing mild anemia (Hb = 10.8 g/dL), microcytosis (MCV = 64.7 g/dL), hypochromia (MCH = 20.3 fL), low levels of HCT (34.4%), and high levels of RDW (16.1%). MLPA results revealed that MLPA probes 1 to 3 have a 0.5 proportion of hybridization in this individual (Figure 4.22).

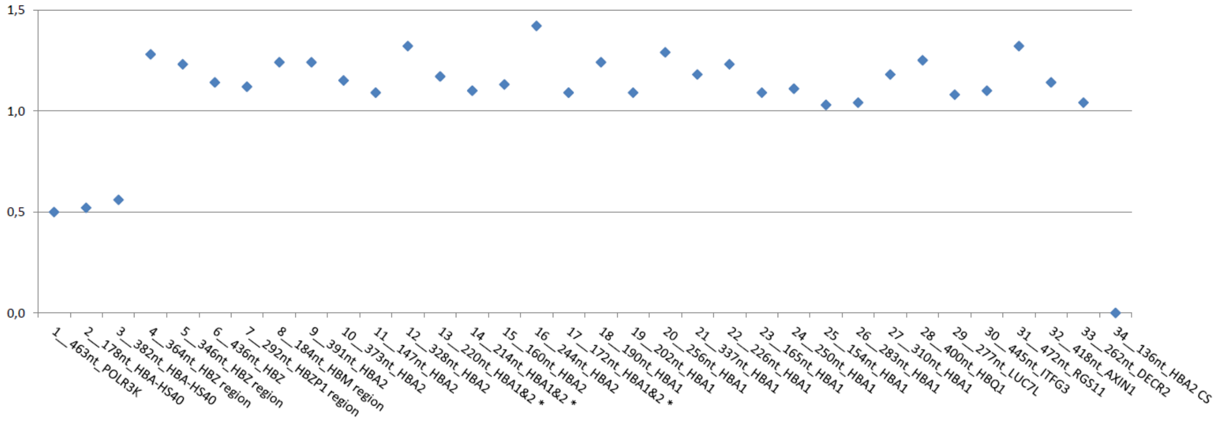


Figure 4.22 – MLPA results for patient number 15 revealing a deletion similar to the (αα)$^{\mathrm{TI}}$ deletion in heterozygosity. Graphic representation of the search for CNVs in the α-globin gene cluster through MLPA. The MLPA probe numbers are indicated on the abscissa. The proportion of each MLPA probe, determined through the comparison of their relative peak with three different reference DNA samples, is indicated in the ordinate. Probe number 4 to 33 have the normal ratio of 1.0, while probe 1 to 3 have a hybridization ratio of 0.5 revealing the presence of a deletion in heterozygosity. Probe number 34 only amplifies in the presence of Hb Constant Spring and, therefore, in normal individuals the proportion is zero.

The MLPA profile indicates the presence of a deletion in heterozygosity with at least 66 kb that starts in the sub-telomeric region of chromosome 16 and ends in an uncertainty region between probes 3 and 4. It removes at least two of the upstream regulatory elements (HS-48 and HS-40) located within the *NPRL3* gene. As HS-33 is located within the uncertain 3'-breakpoint region, it is not known if it is also removed by this deletion. Using the database *IthaCNVs*, it is possible to conclude that this deletion is similar to the **(αα)$^{\mathrm{TI}}$ deletion** previously described by Wilkie *et al.* [77] in two British individuals, by Sollaino *et al.* [78] in an Italian family, and by Ferrão *et al.* [15] in two Portuguese individuals.

The (αα)$^{\mathrm{TI}}$ deletion is a large deletion, not less than 88.1 kb long, that involves the sub-telomeric region of chromosome 16. This deletion removes three upstream regulatory elements (HS-48, HS-40, and HS-33) located within the NPRL3 gene. In previously described clinical cases with this deletion, the individuals presented with microcytosis, hypochromia and, in some cases, anemia [15,78]. In the (αα)$^{\mathrm{TI}}$ chromosome, the 5'-HVR is deleted, which suggests that the α-thalassemia phenotype could be caused by the loss of multiple upstream regulatory elements [77].

Individual 16 is a 31-year-old female with erythrocytosis (RBC = 5.29 x 10^{12}/L), anemia (Hb = 10.7 g/dL), microcytosis (MCV = 68.9 fL), hypochromia (MCH = 20.3 pg), and low levels of MCHC (29.4 g/dL). The MLPA profile showed that the hybridization of MLPA probes 1 to 4 were reduced to a 0.5 proportion (Figure 4.23).

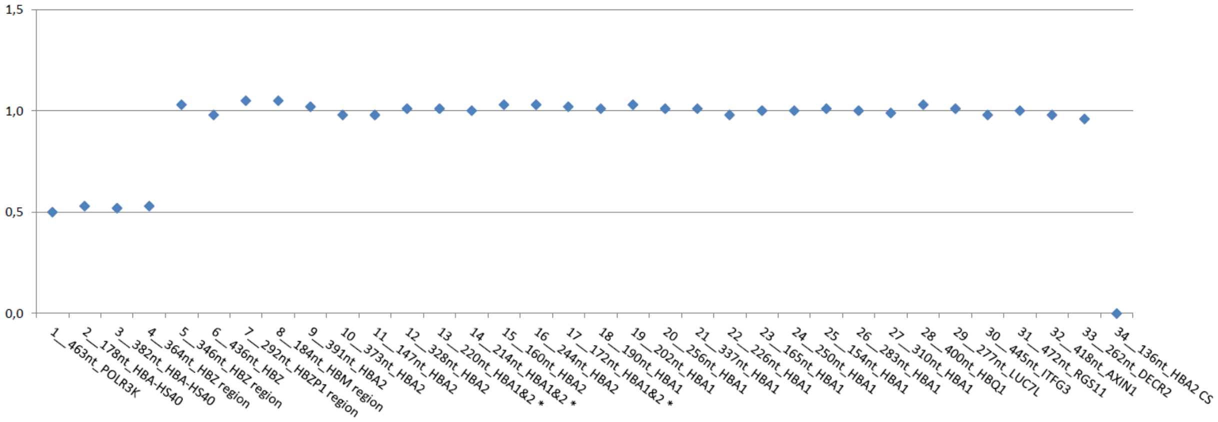


Figure 4.23 – MLPA results for patient number 16 revealing a deletion similar to the (αα)^{MM} deletion in heterozygosity. Graphic representation of the search for CNVs in the α-globin gene cluster through MLPA. The MLPA probe numbers are indicated on the abscissa. The proportion of each MLPA probe, determined through the comparison of their relative peak with three different reference DNA samples, is indicated in the ordinate. Probe number 5 to 33 have the normal ratio of 1.0, while probe 1 to 4 have a hybridization ratio of 0.5 revealing the presence of a deletion in heterozygosity. Probe number 34 only amplifies in the presence of Hb Constant Spring and, therefore, in normal individuals the proportion is zero.

The results indicate the presence of a deletion of at least 96 kb in heterozygosity that starts in the sub-telomeric region of chromosome 16. The 3'-breakpoint of this deletion is in an uncertainty region between probes 4 and 5, where the HS-10 site is located and, therefore, it is not possible to know if this element is removed by this deletion. Despite that, it completely removes three upstream regulatory elements (HS-48, HS-40, and HS-33) located within the *NPRL3* gene. This deletion is similar to the **(αα)^{MM} deletion** previously described by Romão *et al.* [79] in a child from Azores, and by Wenning *et al.* [80] and Mota *et al.* [81] in Brazilian individuals.

The (αα)^{MM} deletion is a large deletion of at least 97 kb that starts in the sub-telomeric region of chromosome 16 and with the 3'-breakpoint located upstream of the *HBZ* gene [12]. On the allele that has the deletion, both *HBA2* and *HBA1* genes are unexpressed [79]. A study using hybrids constructs containing the human chromosome 16 in murine erythroleukemia (MEL) cells revealed that deletions that remove HS-48, HS-40, and HS-33 all at once cause a more complete silencing of α-globin expression (less than 0.1%) [24]. In the absence of HS-40 alone, the remaining interactions can recruit the GTFs to the α-globin gene promoters, but not PolII [16,24]. Because all four regulatory elements are missing in the (αα)^{MM} deletion, both GTFs and PolII recruitment are affected, which may explain why **individual 16** has a more severe phenotype than the other previous two patients.

Experiments with stable transfectants, transgenic mice, and MEL cells revealed that when α-globin genes are not linked to the upstream regulatory elements, α-globin expression occurs at very low levels (less than 1% to 5% of normal) [16,17,24,82]. As a result, deletions that remove these elements, such as those found in these individuals, can cause severe down regulation of α-globin genes [12,23]. These deletions are not very common and comprise a particular group of deletional α-thalassemia, in which the α-globin genes are intact but functionally inactive [17].

In previously described clinical cases, heterozygotes for a deletion that removes the HS-40 are indistinguishable from individuals with two functional α-genes in terms of hematological parameters, as they present with low MCV and low MCH [17,83]. There have been multiple reports of individuals with these deletions combined with the 3.7kb α-thalassemia deletion that manifest clinically as HbH disease [17,79,80], and those with these deletions combined with the Southeast Asian deletion having Bart's hemoglobin [83]. Despite the fact that none of our patients have these genotypes, it is crucial in genetic counseling to analyze the α-globin cluster of their potential partners as both of the conditions mentioned can have serious consequences in the life of a future child.

4.3.3. Clinical cases involving deletions and Hb variants of α-chain

4.3.3.1. Association of Hb Stanleyville-II and the 3.7kb α-thalassemia deletion

Two individuals (case number 7 and 11) arrived at the lab for the molecular identification of an Hb variant of α-chain after their hemoglobin biochemical analyses performed by HPLC revealed an abnormal peak.

Case 7 (Table 4.2), a 66-yearl-old male, had all the hematological parameters within normal limits except for a low RBC count (RBC = 3.94 x 10^{12}/L), and a 27.7% of an Hb variant. **Case 11** (Table 4.2), a 31-year old female, presented with anemia (Hb = 10.3 g/dL), microcytosis (MCV = 71.8 fl), and hypochromia (MCH = 22.9 pg), as well as an HPLC profile that revealed a 45.1% of an Hb variant.

Sanger sequencing eletropherograms revealed that **case 7** had a heterozygous substitution in codon 78 of the *HBA1* gene (*HBA1*:c.237C>A), and that **case 11** had the same mutation but in codon 78 of *HBA2* gene (*HBA2*:c.237C>A) (Figure 4.24). This nucleotide substitution results in the alteration of asparagine for a lysine residue, and, consequently, gives rise to **Hb Stanleyville-II** [84,85].

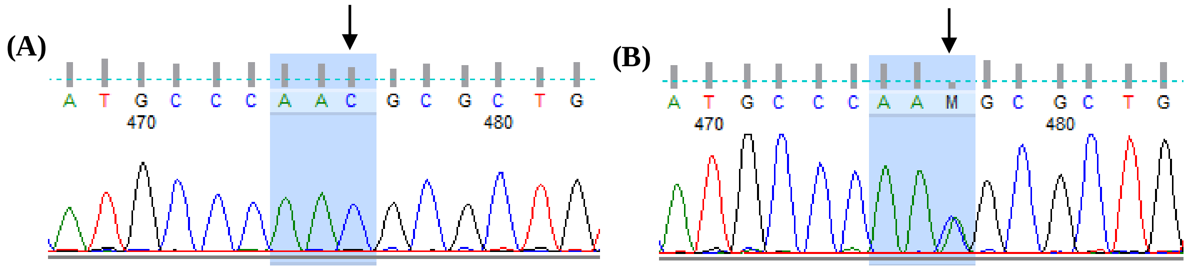
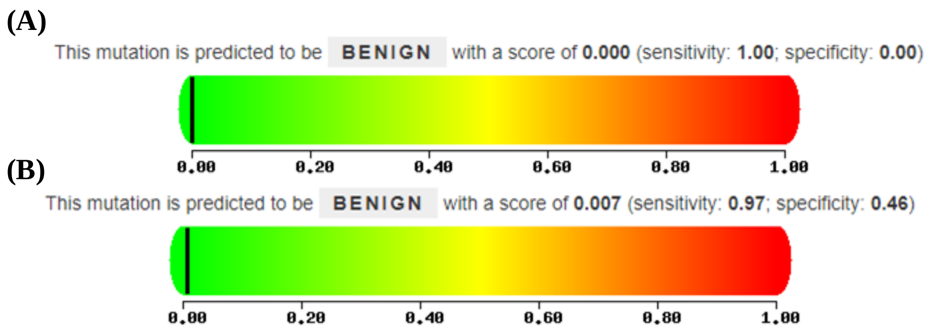


Figure 4.24 – Molecular identification of a hemoglobin variant: Hb Stanleyville-II in case 11. Partial Sanger sequencing eletropherogram of the *HBA2* gene at exon 2 showing **(A)** the wild-type and **(B)** the mutation *HBA2*:c.237C>A (p.Asn78Lys) in heterozygosity responsible for Hb Stanleyville-II. The black arrows indicate the wild type and mutated positions and codon 78 is highlighted in a blue shaded box.

This variant is predicted to be benign at the protein level by *in silico* studies performed using the software *PolyPhen-2* and SIFT, which gave a score of 0.0 (Figure 4.25) and 0.105, respectively.



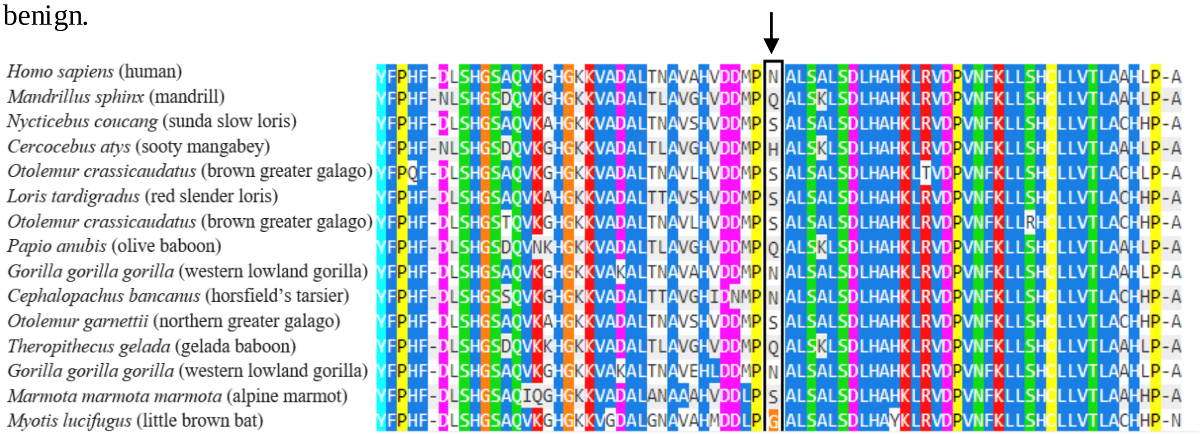
(A) This mutation is predicted to be **BENIGN** with a score of **0.000** (sensitivity: **1.00**; specificity: **0.00**)

(B) This mutation is predicted to be **BENIGN** with a score of **0.007** (sensitivity: **0.97**; specificity: **0.46**)

Figure 4.25 – Prediction of the pathogenic effect of the p.Asn78Lys in the α-globin protein through the *PolyPhen-2* software. (A) The *HumDiv* model predicts the mutation to be benign. **(B)** The *HumVar* model predicts the mutation to be benign.

Hb Stanleyville-II was first described in two Congolese families and, thereafter, has been identified in a third family from Uganda, as well as in a Chinese woman [84,85]. In these previous cases, there were no hematological changes that could be linked with this hemoglobin variant [84]. It is important to note that **individual 7** is of Cape Verdean nationality and **individual 11** is of Angolan nationality, which could explain the diagnosis of an Hb variant with African ancestry.

The evolutionary conservation profile of the asparagine residue at position 78 revealed that this amino acid is not conserved among similar mammal species (Figure 4.26). The fact that lysine has the same biochemical properties has histidine, one of the residues that can replace asparagine in other mammals, can explain why the amino acid substitution observed in Hb Stanleyville-II is considered benign.



Homo sapiens (human)	YFPHF-DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A
Mandrillus sphinx (mandrill)	YFPHF-NLSHGSDQVKGHGKKVADALTLAVGHVDDMPQALSKLSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A
Nycticebus coucang (sunda slow loris)	YFPHF-DLSHGSAQVKAHGKKVADALTNAVSHVDDMPSALSALSDLHAHKLRVDPVNFKLLSHCLLVTLACHHP-A
Cercocebus atys (sooty mangabey)	YFPHF-NLSHGSDQVKGHGKKVADALTLAVGHVDDMPHALSKLSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A
Otolemur crassicaudatus (brown greater galago)	YFPQF-DLSHGSAQVKGHGKKVADALTNAVLHVDDMPSALSALSDLHAHKLTVDPVNFKLLSHCLLVTLACHLP-A
Loris tardigradus (red slender loris)	YFPHF-DLSHGSAQVKAHGKKVADALTTAVSHVDDMPSALSALSDLHAHKLRVDPVNFKLLSHCLLVTLACHHP-A
Otolemur crassicaudatus (brown greater galago)	YFPHF-DLSHGSTQVKGHGKKVADALTNAVLHVDDMPSALSALSDLHAHKLRVDPVNFKLLRHCLLVTLACHHP-A
Papio anubis (olive baboon)	YFPHF-DLSHGSDQVNKHGKKVADALTLAVGHVDDMPQALSKLSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A
Gorilla gorilla gorilla (western lowland gorilla)	YFPHF-DLSHGSAQVKGHGKKVAKALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A
Cephalopachus bancanus (horsfield's tarsier)	YFPHF-DLSHGSSQVKGHGKKVADALTTAVGHIDNMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLACHHP-A
Otolemur garnettii (northern greater galago)	YFPHF-DLSHGSAQVKAHGKKVADALTNAVSHVDDMPSALSALSDLHAHKLRVDPVNFKLLSHCLLVTLACHHP-A
Theropithecus gelada (gelada baboon)	YFPHF-DLSHGSDQVKKHGKKVADALTLAVGHVDDMPQALSKLSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A
Gorilla gorilla gorilla (western lowland gorilla)	YFPHF-DLSHGSAQVKGHGKKVAKALTNAVEHLDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLP-A
Marmota marmota marmota (alpine marmot)	YFPHF-DLSHGSAQIQGHGKKVADALANAAAHVDDLPSALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHP-A
Myotis lucifugus (little brown bat)	YFPHF-DLSHGSAQVKGHGKKVGDALGNAVAHMDDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHP-N

Figure 4.26 – Evolutionary conservation profile of the p.Asn78 amino acid in the α-globin protein. Multiple amino acid sequence alignment of homolog proteins of 15 related mammal species, obtained through the *PolyPhen-2* software. The amino acid substitution pattern is highlighted with an arrow and inside a black box, in which the letter N represents asparagine, Q represents glutamine, S represents serine, H represents histidine, and G represents glycine.

A studied conducted by Costa and collaborators in 1991 revealed that Hb Stanleyville-II may be associated with the 3.7kb α-thalassemia deletion. The authors discovered that individuals who were homozygous for Hb Stanleyville-II also had the 3.7kb deletion in homozygosity, whereas those who had the Hb variant in heterozygosity were also heterozygous for the gene deletion [86]. As a result, we decided to amplify the α-globin genes using Gap-PCR in order to evaluate if this was the case for our samples. This technique revealed that **individual 7** has the 3.7kb deletion in heterozygosity, while **individual 11** has the same deletion in homozygosity. In Figure 4.27, it is possible to see a representation of the genotypes diagnosed for both of these individuals, to better understand the location of Hb Stanleyville-II in relation to the 3.7 kb deletion. Given that Hb Stanleyville-II is a benign Hb variant, the presence of the 3.7kb α-thalassemia deletion explains the phenotypes presented by these patients.

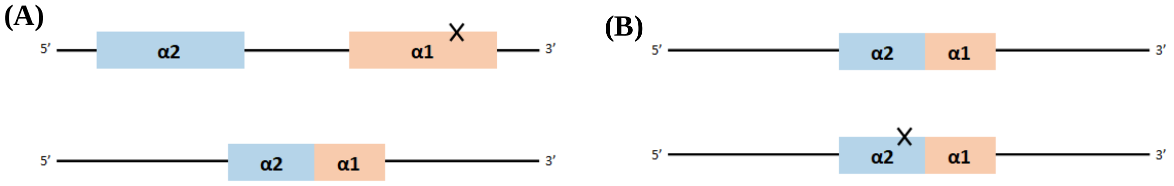


Figure 4.27 –Representation of the genotype diagnosed for (A) individual 7 and (B) individual 11. The *HBA2* (α2) gene is represented in blue and the *HBA1* (α1) gene is in orange. The hybrid gene α2α1 that results from the 3.7kb α-thalassemia deletion is represented in both colors. A black X marks the location of the mutation that leads to Hb Stanleyville-II.

4.3.3.2. Complex case of --^{SEA} deletion, Hb Westmead, and HbE

A 47-year-old male (**case 17** in Table 4.2) arrived at the lab with a complex hematological and biochemical phenotype. His hemoglobin analysis by HPLC revealed an abnormal peak suggesting the presence of a structural Hb variant with the same eletrophoretic migration as HbA$_2$ (15.9%), which suggests a probable presence of HbE (a variant of β-chain). This patient presented with anemia (Hb = 9.7 g/dL), microcytosis (MCV = 69.7 fL), hypochromia (MCH = 22.5 pg), and low levels of HCT (30.1%), which led to the suspicion of a deletion afecting the α-globin gene cluster. Besides that, a reverse-phase HPLC analysis revealed the presence of another unknown variant (of α-chain), which is indistinguishable from HbA by HPLC.

As previously mentioned, we performed a MLPA technique on a DNA sample from this individual and found out that he has the Southeast Asian deletion in heterozygosity (--^{SEA}/αα genotype). Previously described clinical cases with this genotype presented with microcytosis, hypochromia, anisocytosis, and reduced α- and β-chain synthesis ratio [73]. It is worth noting that, not only, an association of this deletion with a milder form of α^{+}-thalassemia leads to HbH disease (--^{SEA}/-α), but a homozygous for α^{0}-thalassemia (--^{SEA}/--^{SEA}) can cause Hb Bart's hydrops fetalis [75]. Therefore, it is of extreme importance to analyze the α-globin gene cluster of this individual's potential partner for any deletions and to perform prenatal diagnosis during pregnancy.

To perform the molecular identification of the unknown α-chain variant, the α-globin genes were amplified and sequenced. The eletropherogram of the *HBA2* gene revealed a substitution at position 369 (*HBA2*:c.369C>G) (Figure 4.28), which appears to be in homozygosity; however, given the previous diagnosis of the SEA deletion, we can conclude that this change occurs in hemizygosity. This substitution affects codon 122 and results in the replacement of histidine for glutamine, which consequently gives rise to an Hb variant known as **Hb Westmead** [87,88].

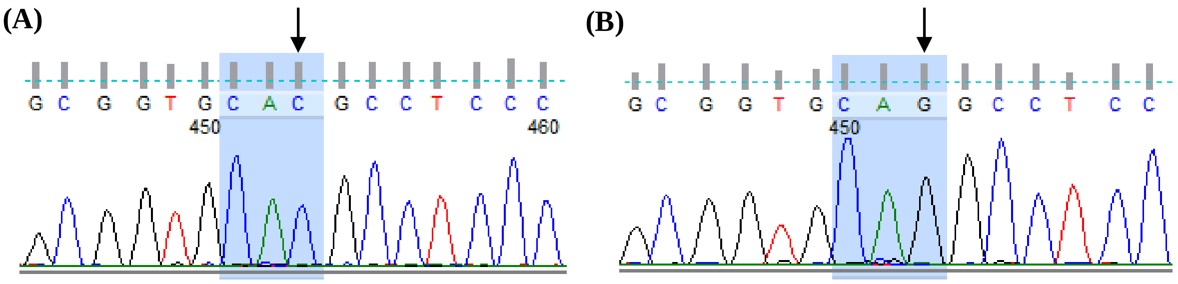
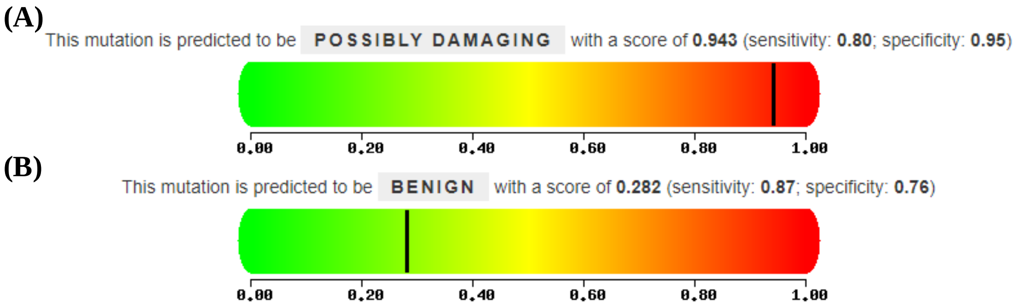


Figure 4.28 – Molecular identification of a hemoglobin variant: Hb Westmead. Partial Sanger sequencing eletropherogram of the *HBA2* gene at exon 3 showing **(A)** the wild-type and **(B)** the mutation *HBA2*:c.369C>G (p.His122Gln) in hemizygosity responsible for Hb Westmead. The black arrows indicate the wild type and mutated positions and codon 122 is highlighted in a blue shaded box.

In silico studies of Hb Westmead revealed discordant impacts at the protein level. SIFT predicts this variant to be tolerated with a score of 0.178, whereas *PolyPhen-2* (Figure 4.29) classifies this variant as possibly damaging by the *HumDiv* model and benign by the *HumVar* with scores of 0.943 and 0.282, respectively.



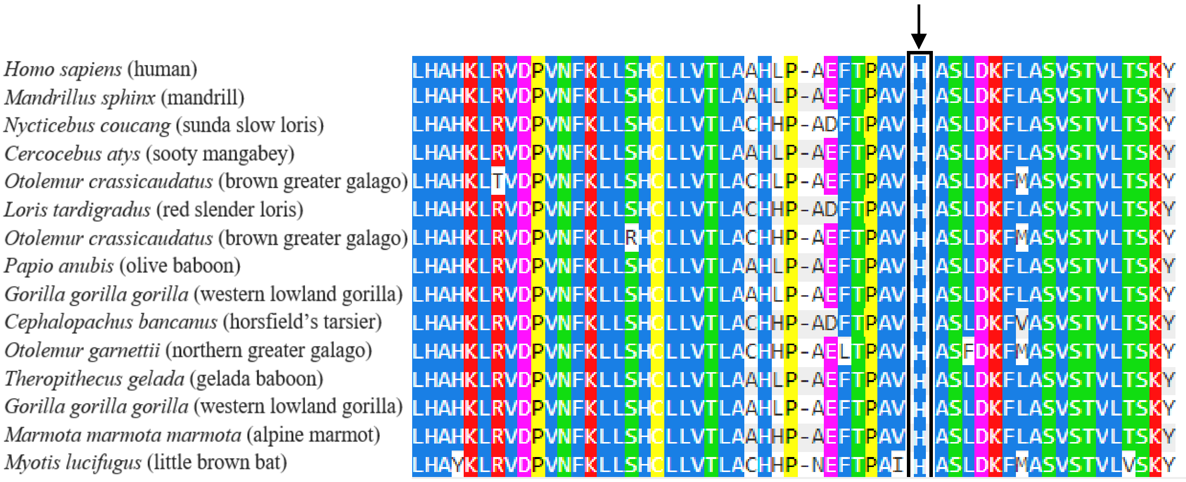
(A) This mutation is predicted to be **POSSIBLY DAMAGING** with a score of **0.943** (sensitivity: **0.80**; specificity: **0.95**)

(B) This mutation is predicted to be **BENIGN** with a score of **0.282** (sensitivity: **0.87**; specificity: **0.76**)

Figure 4.29 – Prediction of the pathogenic effect of the p.His122Gln in the α-globin protein through the *PolyPhen-2* software. (A) The *HumDiv* model predicts the mutation to be possibly damaging. **(B)** The *HumVar* model predicts the mutation to be benign.

The amino acid substitution in Hb Westmead occurs in the fifth amino acid residue of the α-chain helix H, where the histidine residue is involved in van der Waals contact between α1 and β1 chains. Despite this, heat denaturation and modified isopropanol precipitation tests revealed that the substitution for a glutamine residue does not increase the instability of the $\alpha_1\beta_1$ contact area [89].

The evolutionary profile of the histidine residue in position 122 of the α-globin chain revealed that this amino acid residue is highly conserved among similar mammal species (Figure 4.30), suggesting that the histidine substitution can result in a pathogenic Hb variant. Furthermore, in Hb Westmead there is a change of peptide charge and, consequently, in its acid-base properties, which can also explain why this variant is considered to be pathogenic.



Homo sapiens (human)	LHAHKLRVDPVNFKLLSHCLLVTLAAHLP-AEFTPAVHASLDKFLASVSTVLTSKY
Mandrillus sphinx (mandrill)	LHAHKLRVDPVNFKLLSHCLLVTLAAHLP-AEFTPAVHASLDKFLASVSTVLTSKY
Nycticebus coucang (sunda slow loris)	LHAHKLRVDPVNFKLLSHCLLVTLACHHP-ADFTPAVHASLDKFLASVSTVLTSKY
Cercocebus atys (sooty mangabey)	LHAHKLRVDPVNFKLLSHCLLVTLAAHLP-AEFTPAVHASLDKFLASVSTVLTSKY
Otolemur crassicaudatus (brown greater galago)	LHAHKLTVDPVNFKLLSHCLLVTLACHLP-AEFTPAVHASLDKFMASVSTVLTSKY
Loris tardigradus (red slender loris)	LHAHKLRVDPVNFKLLSHCLLVTLACHHP-ADFTPAVHASLDKFLASVSTVLTSKY
Otolemur crassicaudatus (brown greater galago)	LHAHKLRVDPVNFKLLRHCLLVTLACHHP-AEFTPAVHASLDKFMASVSTVLTSKY
Papio anubis (olive baboon)	LHAHKLRVDPVNFKLLSHCLLVTLAAHLP-AEFTPAVHASLDKFLASVSTVLTSKY
Gorilla gorilla gorilla (western lowland gorilla)	LHAHKLRVDPVNFKLLSHCLLVTLAAHLP-AEFTPAVHASLDKFLASVSTVLTSKY
Cephalopachus bancanus (horsfield's tarsier)	LHAHKLRVDPVNFKLLSHCLLVTLACHHP-ADFTPAVHASLDKFVASVSTVLTSKY
Otolemur garnettii (northern greater galago)	LHAHKLRVDPVNFKLLSHCLLVTLACHHP-AELTPAVHASFDKFMASVSTVLTSKY
Theropithecus gelada (gelada baboon)	LHAHKLRVDPVNFKLLSHCLLVTLAAHLP-AEFTPAVHASLDKFLASVSTVLTSKY
Gorilla gorilla gorilla (western lowland gorilla)	LHAHKLRVDPVNFKLLSHCLLVTLAAHLP-AEFTPAVHASLDKFLASVSTVLTSKY
Marmota marmota marmota (alpine marmot)	LHAHKLRVDPVNFKLLSHCLLVTLAAHHP-AEFTPAVHASLDKFLASVSTVLTSKY
Myotis lucifugus (little brown bat)	LHAYKLRVDPVNFKLLSHCLLVTLACHHP-NEFTPAIHASLDKFMASVSTVLVSKY

Figure 4.30 – Evolutionary conservation profile of the p.His122 amino acid in the α-globin protein. Multiple amino acid sequence alignment of homolog proteins of 15 related mammal species, obtained through the *PolyPhen-2* software. The amino acid substitution pattern is highlighted with an arrow and inside a black box, in which the letter H represents histidine.

Hb Westmead, one of the most common Hb variants in China [87,90], was first described in a 30-year-old Chinese female [89]. It has since then been detected in a Laotian family [91] and four Thai individuals [90]. In these previously described clinical cases, the individuals presented with multiple abnormal hematological conditions, including anemia, microcytosis, hypochromia, anisocytosis, poikilocytosis, and polychromasia [88,89,91]. Furthermore, this Hb variant has already been found in association with HbE and the 3.7kb α-thalassemia deletion [91], in compound heterozygosity with the SEA deletion [88], and in coinheritance with β-thalassemia major [88,90].

As previously mentioned, this patient's HPLC profile revealed a 15.9% peak suggestive of HbE, because it has a similar retention time as HbA_2 [92]. Therefore, one of our colleagues used a PCR technique to amplify the *HBB* gene and sequenced the resulting fragments through Sanger sequencing. The obtained results revealed a heterozygous substitution in *HBB* gene (*HBB*:c.79G>A) codon 26 that confirms the presence of HbE.

Hb E results from the alteration of the glutamic acid at position 26 for a lysine residue. Althought HbE has no clinical significance on its own, its interactions with forms of α-thalassemia can cause a wide range of clinical symptoms of varying severity. HbE is extremely common in many Asian countries, including India, Laos, Cambodia, and Thailand [92–94].

To summarize, our 47-year-old male patient represents a very complex clinical case involving the Southeast Asian α-thalassemia deletion, the α-chain variant Hb Westmead, and the β-chain variant HbE. The combination of these three modifications can explain the patient's phenotype. Aside from that, the fact that our patient is of Asian descent may help to explain why these alterations were discovered all at once. Although Hb Westmead has previously been linked to HbE and the SEA deletion separately, as far as we know, this is the first case in which all three alterations were found in the same individual.

4.4. Haplotypes of the distal regulatory region HS-40

One of the primary goals of this study was to identify the distal regulatory region HS-40 haplotypes that are associated with the 3.7kb deletion in the Portuguese population with α-thalassemia, as well as in the general population without this deletion. For that, we performed the identification of the HS-40 haplotypes in **111 Portuguese individuals** that were subdivided in three different groups according to their α-globin genotype:

(I) Individuals without α-thalassemia: including **50 individuals** with normal hematological parameters and with no common deletions in the α-globin gene cluster.

(II) Carriers of the 3.7kb deletion in heterozygosity: including **34 individuals** with the 3.7kb deletion in heterozygosity (this individuals had been previously studied in our lab for the presence this deletion).

(III) Individuals with the 3.7kb deletion in homozygosity: including **27 individuals** with a $-\alpha^{3.7}/-\alpha^{3.7}$ genotype (this diagnosis was previously determined by our colleagues from DPS/INSA).

To determine the sequence of the different HS-40 haplotypes, firstly we amplified a DNA fragment of 400 bp containing the α-MRE, using conventional PCR as described in Table 7.2 of the Supplementary materials. The amplicons were then visualized in a 2.5% agarose gel (Figure 4.31).

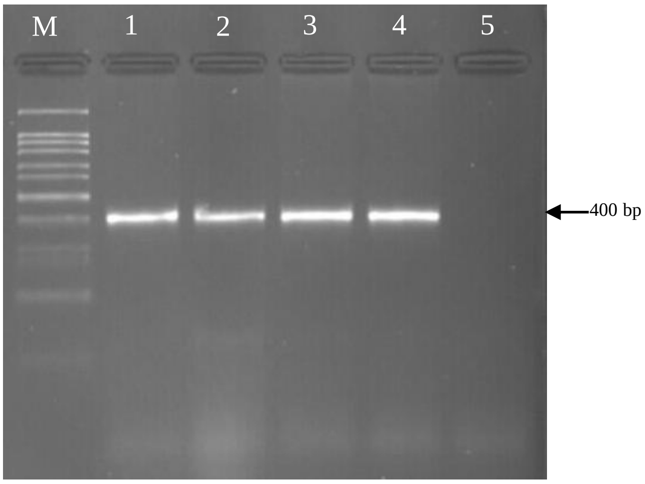


Figure 4.31 – Visualization of the amplified DNA fragment that includes the HS-40, the principal distal regulatory site of the α-globin genes. Electrophoretic profile of the 400 bp fragment in a 2.5% (m/v) agarose gel, in an electrophoretic system at 70V during 60 minutes. **M** – molecular weight marker (100 bp plus DNA Ladder, BIORON); **1, 2, 3 and 4** – PCR product including the HS-40 region of four individuals; **5** – negative control. The size of the fragment in bp is included on the right side of the image.

Following the PCR reaction, Sanger sequencing was used to sequence the DNA fragment containing the region we intend to study and the results were compared to those described by Harteveld and his collaborators in 2002 [25].

4.4.1. HS-40 haplotypes in the Portuguese population

The HS-40 sequences of the **111 Portuguese individuals** revealed four distinct haplotypes across all studied groups, labelled A, B, C, and D (Table 4.5).

Table 4.5 – HS-40 haplotypes in the Portuguese population without the α-thalassemia 3.7kb deletion, with the deletion in heterozygosity and in homozygosity

Individuals	**Alpha-globin genotype**	**HS-40 haplotypes**			
		A **x (%)**	**B** **x (%)**	**C** **x (%)**	**D** **x (%)**
Group I (n = 50; x = 100)	αα/αα	57 (57.0)	39 (39.0)	3 (3.0)	1 (1.0)
Group II (n = 34; x = 68)	$-\alpha^{3.7}/\alpha\alpha$	46 (67.7)	19 (27.9)	0 (0.0)	3 (4.4)
Group III (n = 27; x = 54)	$-\alpha^{3.7}/-\alpha^{3.7}$	31 (57.4)	8 (14.8)	0 (0.0)	15 (27.8)
Total (n = 111; x = 222)		134 (60.4)	66 (29.7)	3 (1.3)	19 (8.6)

n – number of Portuguese individuals; **x** – number of alleles.

Haplotype A, was found in 134 of the 222 studied alleles (60.4%) and is the most common across all studied groups, which was expected given that this is the ancestral sequence [25]. In general, the second most frequent haplotype is B, found in 66 out of the 222 alleles (29.7%). This is the second most frequent haplotype in individuals without α-thalassemia (group I) and in carriers of the 3.7kb deletion (group II) where it represents 39.0% and 27.9% of the alleles, respectively. However, in group III, where all alleles have the 3.7kb deletion, haplotype D is the second most prevalent (27.8%). Haplotype C, in general, was least frequent as it was found in only 1.3% of the 222 studied alleles, all of them in the group with the wild-type genotype (group I).

When it comes to the HS-40 genotypes, we found seven different combinations designated AA, AB, AD, BB, BC, BD, and DD. The number of HS-40 genotypes and their distribution within each group are shown in Table 4.6 and Figure 4.32.

Table 4.6 – HS-40 genotypes in the Portuguese population without the α-thalassemia3.7kb deletion, with the deletion in heterozygosity and in homozygosity

Individuals	Alpha-globin genotype	HS-40 genotypes						
		AA n (%)	AB n (%)	AD n (%)	BB n (%)	BC n (%)	BD n (%)	DD n (%)
Group I (n = 50)	αα/αα	18 (36.0)	20 (40.0)	1 (2.0)	8 (16.0)	3 (6.0)	0 (0.0)	0 (0.0)
Group II (n = 34)	$-\alpha^{3.7}/\alpha\alpha$	15 (44.1)	13 (38.2)	3 (8.8)	3 (8.8)	0 (0.0)	0 (0.0)	0 (0.0)
Group III (n = 27)	$-\alpha^{3.7}/-\alpha^{3.7}$	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 Portuguese individuals.

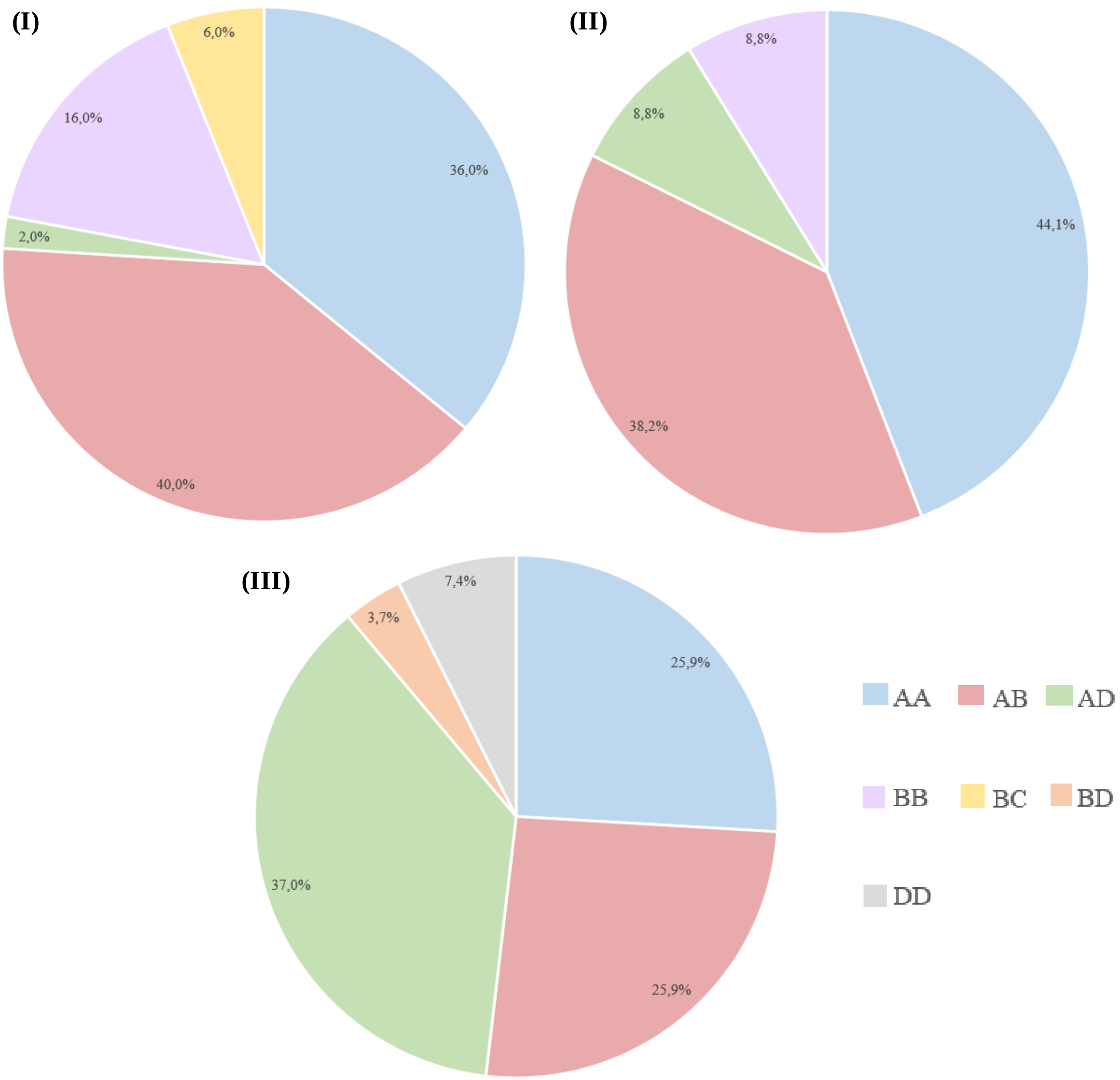


Figure 4.32 – Graphic distribution of the different HS-40 genotype frequencies (in %) according to the corresponding α-globin genotypes of the studied Portuguese individuals. (I) in the group without the 3.7kb α-thalassemia deletion (αα/αα); **(II)** in the group with the 3.7kb deletion in heterozygosity ($-\alpha^{3.7}/\alpha\alpha$); **(III)** in the group with the 3.7kb deletion in homozygosity ($-\alpha^{3.7}/-\alpha^{3.7}$).

The graphic representations above show that the AA and AB combinations are the two most common in patients without α-thalassemia (36.0% and 40.0%, respectively), as well as in those with the $-\alpha^{3.7}/\alpha\alpha$ genotype (44.1% and 38.2%, respectively). However, the same cannot be said for patients with a $-\alpha^{3.7}/-\alpha^{3.7}$ genotype, which have a higher prevalence of the AD combination (37.0%).

Haplotype B presents with a decrease in frequencies across studied groups, with 39.0% in the individuals without α-thalassemia, 27.9% in carriers of the 3.7kb deletion, and 14.8% in patients with the same deletion in homozygosity. This is especially noticeable when it comes to the BB genotype (represented in purple), which is the third most frequent genotype in the first group mentioned (16.0%), but was not found in patients with the 3.7kb deletion in homozygosity.

On the other hand, haplotype C was only found in individuals without α-thalassemia, in combination with B, resulting in the BC genotype (represented in yellow).

Haplotype D shows a tendency contrary to that displayed in haplotype B. In this case, it is possible to observe an increase in the frequency along the studied groups, with only 1.0% in the individuals without α-thalassemia, 4.4% in individuals with a $-\alpha^{3.7}/\alpha\alpha$ genotype, and 27.8% in patients with the 3.7kb deletion in both alleles. This haplotype was found in three different combinations: AD, BD and DD. The majority of haplotype D was found in combination with haplotype A, giving rise to genotype AD (represented in green). This genotype was found in a total of 14 individuals, with 10 of them being homozygous for the 3.7kb deletion (71.4%). Thus, this is the most frequent genotype in the group with the 3.7kb deletion in homozygosity and, through the graphic representations shown above, it is possible to observe an increase of the frequency in which this combination was detected across the studied groups. Both genotype BD (represented in orange) and genotype DD (represented in grey) were only found in individuals diagnosed with the 3.7kb deletion in homozygosity.

In order to evaluate if the distribution of the HS-40 haplotypes and genotypes is different between groups and is related with the presence of 3.7kb α-thalassemia deletion, we compared the proportions of haplotypes and genotypes between group I (without α-thalassemia) and group III (individuals with the 3.7kb deletion in homozygosity). Through this statistical analysis (using the Test of Equal or Given Proportions), it was possible to conclude that the distribution of HS-40 haplotypes and genotypes are significantly different between individuals with and without the 3.7kb α-thalassemia deletion **(*p*_value < 0.001)** and, consequently, there is an association between specific HS-40 haplotypes 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 in group III.

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

After determining that specific HS-40 haplotypes are associated with the 3.7kb α-thalassemia deletion in the Portuguese population, we wanted to investigate the ancestry of this deletion in our population.

As previously revealed by our results, in the Portuguese population, haplotype D is the second most common (27.8%) in the group of individuals with the 3.7kb deletion in homozygosity, and was found in three different combinations: AD, BD, and DD. Initially, these genotypes were only reported in African people [25]; however, in a more recent research, they were also detected in Uruguayans [95]. In these two populations, genotypes AD, BD, and DD have been found mostly in individuals with the

3.7kb deletion. Thus, there is a strong possibility that haplotype D is associated *in cis* with this α-thalassemia deletion.

In order to better visualize the similarities between the Portuguese population with the 3.7kb deletion and other populations, we performed a multiple correspondence analysis using *R* software. This analysis revealed that individuals who do not have α-thalassemia (PRT Normal) are grouped with other European populations, while samples with the 3.7kb deletion (PRT -α3.7/αα and PRT -α3.7/-α3.7) are isolated from these and found to be more closely related to the African population (Figure 4.33).

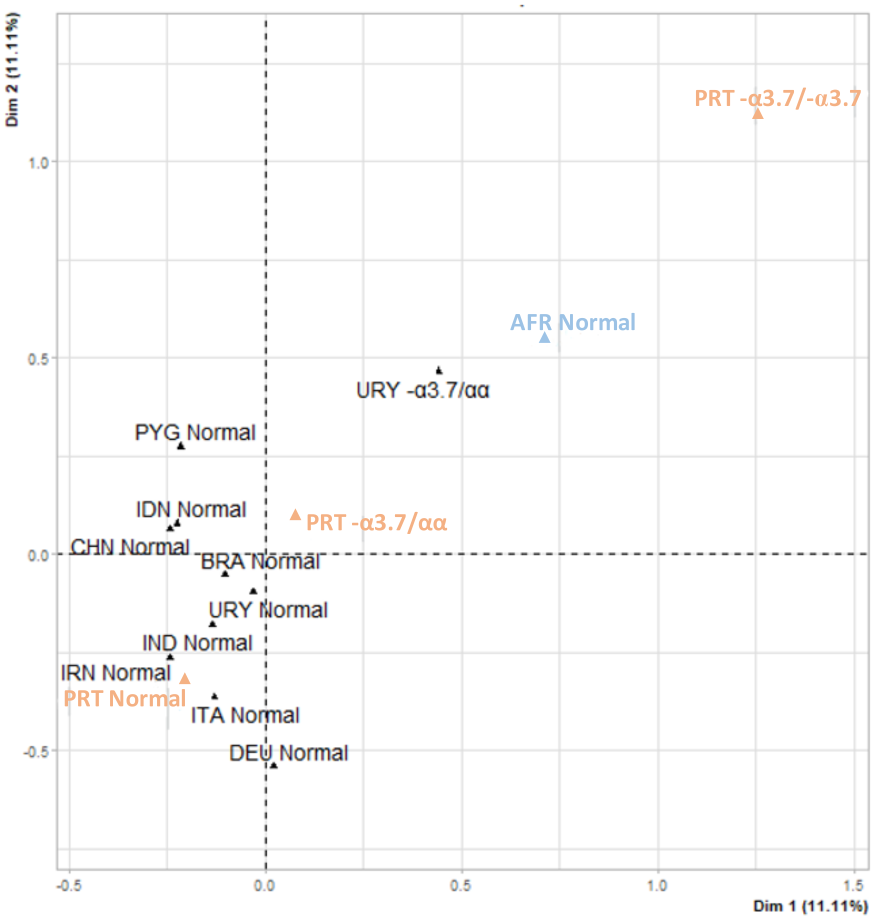


Figure 4.33 – Multiple correspondence analysis of the HS-40 genotypes in multiple geographic populations. AFR: African; BRA: Brazilian; CHN: Chinese; DEU: Dutch; IDN: Indonesian; IND: Indian; IRN; Iranian; ITA: Italian; PRT: Portuguese; PYG: Pygmies; URY: Uruguayan. All the genotypes from foreign populations were collected from [25,26,95,96]. The Portuguese populations investigated in this study are marked as PRT normal (group I); PRT –α3.7/αα (group II) and PRT –α3.7/–α3.7" (group III).

In conclusion, our results show a predominant African origin of the 3.7kb α-thalassemia deletion in the Portuguese population.

4.4.2. Influence of haplotype D in the long-range regulation of α-globin gene expression

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 [25], a composite binding site that is recognized by the transcription factor NF-E2 [97,98].

NF-E2 belongs to the basic region-leucine zipper (BZIP) family and is composed by two protein dimers: a widely expressed 18 kDa (p18Nfe2) and a hematopoietic-specific 45 kDa subunit (p45Nfe2). The protein p45Nfe2 is expressed in hematopoietic cells of the erythroid, megakaryocytic, and mast cell lineages and, therefore, its expression is considered hematopoietic restricted [98–100].

Multiple studies have been performed in order to understand the role of NF-E2 in the regulation of globin gene expression. Studies using MEL 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 [100]. Analysis of mice lacking NF-E2 showed that erythroid development is only subtly disturbed, with red cells presenting a slightly decrease in hemoglobin content and exhibit some microcytosis. These mice were also associated with an increased erythropoiesis and mild anemia [97,98]. Besides that, the same study showed that a mutation in NF-E2 leads to a 25% reduction in α-globin gene expression [98].

As a result, interference in the NF-E2 binding site, as seen in haplotype D, may result in decreased α-globin gene expression; even so, the presence of this HS-40 haplotype in heterozygosity is not enough to cause α-thalassemia. Moreover, the interference with this 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 3.7kb deletion. When it comes to individuals that are both homozygotes for the HS-40 haplotype D and for the 3.7kb deletion, α-globin gene expression may reduce below a critical level and result in the formation of HbH (β4 tetramers), due to an excess of unpaired β-globin chains [25]. However, due to the small number of homozygous individuals for the HS-40 DD genotype and the 3.7kb deletion that have been found, this association is yet to be established.

Considering the effects that the interference in the NF-E2 binding site had in the hematological parameters of NF-E2 deficient mice (mild anemia, slight microcytosis and hypochromia), we would like to investigate if this could be applied to our studied population. For that, we compared the hematological parameters available for individuals with the HS-40 AA, AD, and DD genotypes, in the group with the 3.7kb α-thalassemia deletion. However, according to the results of the statistical analysis, there are no significant differences (p_value > 0.05) for any hematological parameters. These results may be justified by many reasons. One of them could be that the sample size is too small to draw any conclusions. Another reason could be that long-range regulation of α-globin gene expression in mice differs from that in humans, as suggested by other studies [12,101,102] and is probably under a more complex mechanism.

To go further into this subject, it would be essential to carry out this investigation in a larger population sample. Furthermore, *in vitro* expression experiments using constructs with the luciferase gene under the control of different human HS-40 haplotypes would be extremely beneficial in better understanding the influence of the D-allele in α-globin gene expression.

5. Conclusion

One of the primary goals of this study was to contribute to the knowledge of the spectrum of deletions and point mutations that cause α-thalassemia or give rise to α-chain hemoglobin variants in the Portuguese population. For that, 17 samples from individuals suspected of having one or both conditions were analyzed through different molecular methodologies.

In the 11 samples studied for the diagnosis of deletional α-thalassemia using Gap-PCR and Multiplex Gap-PCR techniques, we were able to identify four heterozygous ($-\alpha^{3.7}/\alpha\alpha$ genotype) and two homozygous individuals ($-\alpha^{3.7}/-\alpha^{3.7}$ genotype) for the 3.7kb deletion, as well as one carrier of the 4.2kb deletion ($-\alpha^{4.2}/\alpha\alpha$ genotype). For the four individuals that did not have neither the 3.7kb nor the 4.2kb deletions, an MLPA technique was used to search for large and/or unknown deletions. This methodology allow the identification of four distinct deletions in heterozygosity, the Southeast Asian deletion ($--^{SEA}/\alpha\alpha$ genotype), and three very rare deletions that remove the upstream regulatory element HS-40: $(\alpha\alpha)^{ALT}$ deletion, $(\alpha\alpha)^{TI}$ deletion, and $(\alpha\alpha)^{MM}$ deletion.

The molecular diagnosis of the α-chain hemoglobin variants was performed for nine samples and allowed the identification of eight different missense variants in coding regions of *HBA* genes. These structural hemoglobin variants have different pathogenicity classifications: five are benign (Hb J-Paris, Hb Nouakchott, Hb Brugg, Hb Oleander, and Hb Stanleyville-II), one is likely pathogenic (Hb Westmead), and two are pathogenic (Hb Legnano and Hb Setif). It is worth noting that Hb Stanleyville-II was discovered in two individuals alongside with the common 3.7kb α-thalassemia deletion, and that Hb Westmead was found simultaneously with the Southeast Asian deletion and with HbE (a hemoglobin variant of β-chain).

The second main goal of this study was to identify the different HS-40 haplotypes (A to F) that are associated with the 3.7kb deletion, the most common α-thalassemia deletion, in the Portuguese population, and determine its ancestry. In this part of the study, 61 samples from individuals diagnosed with this deletion (34 with the $-\alpha^{3.7}/\alpha\alpha$ genotype and 27 with the $-\alpha^{3.7}/-\alpha^{3.7}$ genotype), as well as 50 controls without α-thalassemia were analyzed.

The sequence of HS-40 haplotypes revealed four distinct haplotypes labelled A, B, C, and D. Haplotype A was the most common in all studied groups, and in individuals without α-thalassemia and in the carriers of the 3.7kb deletion the second most frequent haplotype was B. However, in the individuals with this deletion in homozygosity, haplotype D was the second most common.

When it comes to the HS-40 genotypes, we found seven different combinations designated AA, AB, AD, BB, BC, BD, and DD. 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, while in patients with the $-\alpha^{3.7}/-\alpha^{3.7}$ genotype the most prevalent combination was AD. Using statistical tests, we verified that the distribution of HS-40 haplotypes and genotypes has significant differences between individuals with and without α-thalassemia.

In individuals with the 3.7kb deletion in homozygosity, the AD genotype was the most frequent and, in this group, it was also possible to find the BD and DD combinations. Therefore, it is possible to conclude that haplotype D is associated in cis with the 3.7kb α-thalassemia deletion. These results may have clinical importance as this haplotype has an alteration in the consensus sequence of the NF-E2 binding site and *in vitro* experiments showed a decrease in its enhancer activity on α-globin genes. Furthermore, a multiple correspondence analysis revealed that Portuguese individuals without α-thalassemia are grouped with other European populations, while samples with the 3.7kb deletion are separated from these and more closely related to the African population.

To summarize, this study allowed the identification of six different deletions, three that remove one or more of the α-globin genes and another three, very rare, that affect the upstream regulatory elements, as well as eight α-chain hemoglobin variants, two of which are considered pathogenic (Hb Legnano and Hb Setif) and one that is likely pathogenic (Hb Westmead). This way contributing to the knowledge of deletions and point mutations that cause α-thalassemia and lead to α-chain hemoglobin variants, which allows a better referral to genetic counseling, especially for couples with a chance of having a child with HbH disease or Hb Bart's hydrops fetalis.

Furthermore, this study revealed for the first time an association between a specific HS-40 haplotype and the common 3.7kb α-thalassemia deletion in the Portuguese population, and its likely African ancestry. To go further into this subject, it would be extremely beneficial to perform *in vitro* experiments using constructs with the luciferase gene under control of different HS-40 haplotypes, in order to better understand the effect of haplotype D in the expression of α-globin genes.

6. References

[1] M.J. Koury, P. Ponka, New Insights Into Erythropoiesis: The Roles of Folate, Vitamin B 12 , and Iron, Annu. Rev. Nutr. 24 (2004) 105–131. https://doi.org/10.1146/annurev.nutr.24.012003.132306.

[2] D.R. Higgs, N. Roy, D. Hay, Erythropoiesis, in: A.V. Hoffbrand, D.R. Higgs, D.M. Keeling, A.B. Mehta (Eds.), Postgrad. Haematol., 7th ed., Wiley, 2015: pp. 11–20. https://doi.org/10.1002/9781118853771.ch2.

[3] A.J. Marengo-Rowe, Structure-function relations of human hemoglobins, Baylor Univ. Med. Cent. Proc. 19 (2006) 239–245. https://doi.org/10.1080/08998280.2006.11928171.

[4] S. Farashi, C.L. Harteveld, Molecular basis of α-thalassemia, Blood Cells, Mol. Dis. 70 (2018) 43–53. https://doi.org/10.1016/j.bcmd.2017.09.004.

[5] I. Kalle Kwaifa, M.I. Lai, S. Md Noor, Non-deletional alpha thalassaemia: A review, Orphanet J. Rare Dis. 15 (2020) 1–12. https://doi.org/10.1186/s13023-020-01429-1.

[6] S. Fucharoen, P. Winichagoon, Thalassemia and abnormal hemoglobin, Int. J. Hematol. 76 (2002) 83–89. https://doi.org/10.1007/BF03165094.

[7] B. Dyson, The Pharmacist’s Guide to Sickle Cell Disease, (2017). https://www.tldrpharmacy.com/content/the-pharmacists-guide-to-sickle-cell-disease. (accessed January 4, 2022)

[8] H.L. Muncie, J.S. Campbell, Alpha and beta thalassemia, Am. Fam. Physician. 80 (2009) 339–344.

[9] B.E. Clark, S.L. Thein, Molecular diagnosis of haemoglobin disorders, Clin. Lab. Haematol. 26 (2004) 159–176. https://doi.org/10.1111/j.1365-2257.2004.00607.x.

[10] S. Mettananda, R.J. Gibbons, D.R. Higgs, α-globin as a molecular target in the treatment of β-thalassemia.pdf, Blood. 125 (2015) 3694–3701. https://doi.org/10.1182/blood-2015-03-633594.

[11] A.T. Taher, D.J. Weatherall, M.D. Cappellini, Thalassaemia, Lancet. 391 (2018) 155–167. https://doi.org/10.1016/S0140-6736(17)31822-6.

[12] A. Coelho, I. Picanço, F. Seuanes, M.T. Seixas, P. Faustino, Novel large deletions in the human α-globin gene cluster: Clarifying the HS-40 long-range regulatory role in the native chromosome environment.pdf, Blood Cells, Mol. Dis. 45 (2010) 147–153. https://doi.org/10.1016/j.bcmd.2010.05.010.

[13] J.S. Lee, S.I. Cho, S.S. Park, M.W. Seong, Molecular basis and diagnosis of thalassemia, Blood Res. 56 (2021) 39–43. https://doi.org/10.5045/br.2021.2020332.

[14] S. Kukreti, H. Kaur, M. Kaushik, A. Bansal, S. Saxena, S. Kaushik, R. Kukreti, Structural polymorphism at LCR and its role in beta-globin gene regulation, Biochimie. 92 (2010) 1199–1206. https://doi.org/10.1016/j.biochi.2010.04.020.

[15] J. Ferrão, M. Silva, L. Gonçalves, S. Gomes, P. Loureiro, A. Coelho, A. Miranda, F. Seuanes, A.B. Reis, F. Pina, R. Maia, P. Kjöllerström, E. Monteiro, J.F. Lacerda, J. Lavinha, J. Gonçalves, P. Faustino, Widening the spectrum of deletions and molecular mechanisms underlying alpha-thalassemia, Ann. Hematol. 96 (2017) 1921–1929. https://doi.org/10.1007/s00277-017-3090-y.

[16] D.R. Higgs, W.G. Wood, Long-range regulation of α globin gene expression during erythropoiesis, Curr. Opin. Hematol. 15 (2008) 176–183. https://doi.org/10.1097/MOH.0b013e3282f734c4.

[17] D.M. Ribeiro, M.F. Sonati, Regulation of human alpha-globin gene expression and alpha-thalassemia, Genet. Mol. Res. 7 (2008) 1045–1053. https://doi.org/10.4238/vol7-4gmr472.

[18] R. Galanello, A. Cao, Alpha-thalassemia, Genet. Med. 13 (2011) 83–88. https://doi.org/10.1097/GIM.0b013e3181fcb468.

[19] C.L. Harteveld, D.R. Higgs, α-Thalassaemia, Orphanet J. Rare Dis. 5 (2010) 2–21. https://doi.org/10.1186/1750-1172-5-13.

[20] D.R. Higgs, J.D. Engel, G. Stamatoyannopoulos, Thalassaemia, Lancet. 379 (2012) 373–383. https://doi.org/10.1016/S0140-6736(11)60283-3.

[21] D.M. Ribeiro, T.R. Zaccariotto, M.N.N. Santos, F.F. Costa, M.F. Sonati, Influence of the polymorphisms of the α-major regulatory element HS-40 on in vitro gene expression, Brazilian J. Med. Biol. Res. 42 (2009) 783–786. https://doi.org/10.1590/S0100-879X2009005000014.

[22] C.N. Suemasu, E.M. Kimura, D.M. Oliveira, M.A.C. Bezerra, A.S. Araújo, F.F. Costa, M.F. Sonati, Characterization of alpha thalassemic genotypes by multiplex ligation-dependent probe amplification in the Brazilian population, Brazilian J. Med. Biol. Res. 44 (2011) 16–22. https://doi.org/10.1590/S0100-

879X2010007500144.

[23] J.A. Sharpe, P.S. Chan-Thomas, J. Lida, H. Ayyub, W.G. Wood, D.R. Higgs, Analysis of the human alpha globin upstream regulatory element (HS-40) in transgenic mice., EMBO J. 11 (1992) 4565–4572. https://doi.org/10.1002/j.1460-2075.1992.tb05558.x.

[24] D.R. Higgs, D. Vernimmen, B. Wood, Long-Range Regulation of α-Globin Gene Expression, in: Adv. Genet., 2008: pp. 143–173. https://doi.org/10.1016/S0065-2660(07)00005-3.

[25] C.L. Harteveld, M. Muglia, G. Passarino, M.F. Kielman, L.F. Bernini, Genetic polymorphism of the major regulatory element HS-40 upstream of the human α-globin gene cluster, Br. J. Haematol. 119 (2002) 848–854. https://doi.org/10.1046/j.1365-2141.2002.03917.x.

[26] D.M. Ribeiro, M.S. Figueiredo, F.F. Costa, M.F. Sonati, Haplotypes of α-globin gene regulatory element in two Brazilian native populations, Am. J. Phys. Anthropol. 121 (2003) 58–62. https://doi.org/10.1002/ajpa.10193.

[27] E.J. Benz, B.L. Ebert, Hemoglobin Variants Associated With Hemolytic Anemia, Altered Oxygen Affinity, and Methemoglobinemias, 7th ed., Elsevier Inc., 2018. https://doi.org/10.1016/B978-0-323-35762-3.00043-3.

[28] S.A. Sarnaik, Thalassemia and related hemoglobinopathies, Indian J. Pediatr. 72 (2005) 319–324. https://doi.org/10.1007/BF02724015.

[29] S.K. Hartwell, B. Srisawang, P. Kongtawelert, G.D. Christian, K. Grudpan, Review on screening and analysis techniques for hemoglobin variants and thalassemia, Talanta. 65 (2005) 1149–1161. https://doi.org/10.1016/j.talanta.2004.09.013.

[30] C.S. Thom, C.F. Dickson, D.A. Gell, M.J. Weiss, Hemoglobin Variants: Biochemical Properties and Clinical Correlates, Cold Spring Harb. Perspect. Med. 3 (2013) 1–22. https://doi.org/10.1101/cshperspect.a011858.

[31] M.C. Martins, G. Olim, J. Melo, H.A. Magalhaes, M.O. Rodrigues, Hereditary anaemias in Portugal: Epidemiology, public health significance, and control, J. Med. Genet. 30 (1993) 235–239. https://doi.org/10.1136/jmg.30.3.235.

[32] T. Munkongdee, P. Chen, P. Winichagoon, S. Fucharoen, K. Paiboonsukwong, Update in laboratory diagnosis of thalassemia, Front. Mol. Biosci. 7 (2020) 1–12. https://doi.org/10.3389/fmolb.2020.00074.

[33] D.E. Sabath, Molecular diagnosis of thalassemias and hemoglobinopathies: An ACLPS critical review, Am. J. Clin. Pathol. 148 (2017) 6–15. https://doi.org/10.1093/AJCP/AQX047.

[34] T.T. Celkan, What does a hemogram say to us?, Turkish Arch. Pediatr. 55 (2020) 103–116. https://doi.org/10.14744/TurkPediatriArs.2019.76301.

[35] Prescrição e Determinação do Hemograma, Direção-Geral da Saúde, 2011. https://ordemdosmedicos.pt/wp-content/uploads/2017/09/Prescricao_e_Determinacao_do_Hemograma.pdf. (accessed May 17, 2022).

[36] C. Dodé, R. Krishnamoorthy, J. Lamb, J. Rochette, Rapid analysis of -α3.7 thalassaemia and αααanti 3.7 triplication by enzymatic amplification analysis, Br. J. Haematol. 83 (1993) 105–111. https://doi.org/10.1111/j.1365-2141.1993.tb04639.x.

[37] M.J. Peres, M.H. Carreiro, M.C. Machado, T. Seixas, I. Picanço, L. Batalha, J. Lavinha, M.C. Martins, Rastreio neonatal de hemoglobinopatias numa população residente em Portugal, Acta Med. Port. 9 (1996) 135–139. https://doi.org/10400.17/633.

[38] D.K. Lahiri, J.I. Numberger, A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies, Nucleic Acids Res. 19 (1991) 5444. https://doi.org/10.1093/nar/19.19.5444.

[39] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, H. Erlich, Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction, Cold Spring Harb. Symp. Quant. Biol. 51 (1986) 263–273. https://doi.org/10.1101/sqb.1986.051.01.032.

[40] M.T. Rahman, M.S. Uddin, R. Sultana, A. Moue, M. Setu, Polymerase Chain Reaction (PCR): A Short Review, Anwer Khan Mod. Med. Coll. J. 4 (2013) 30–36. https://doi.org/10.3329/akmmcj.v4i1.13682.

[41] S. Leng Kho, K. Heng Chua, E. George, J.A.M.A. Tan, A novel gap-PCR with high resolution melting analysis for the detection of α-thalassaemia Southeast Asian and Filipino β0-thalassaemia deletion, Sci. Rep. 5 (2015) 1–7. https://doi.org/10.1038/srep13937.

[42] S.S. Chong, C.D. Boehm, D.R. Higgs, G.R. Cutting, Single-tube multiplex-PCR screen for common deletional determinants of α-thalassemia, Blood. 95 (2000) 360–362.

https://doi.org/10.1182/blood.V95.1.360.001k03_360_362.

[43] A.S.-C. Tan, T.C. Quah, P.S. Low, S.S. Chong, A rapid and reliable 7-deletion multiplex polymerase chain reaction assay for α-thalassemia, Blood. 98 (2001) 250–251. https://doi.org/10.1182/blood.V98.1.250.

[44] M. Kchouk, J.F. Gibrat, M. Elloumi, Generations of sequencing technologies: From first to next generation, Biol. Med. 9 (2017). https://doi.org/10.4172/0974-8369.1000395.

[45] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U. S. A. 74 (1977) 5463–5467. https://doi.org/10.1073/pnas.74.12.5463.

[46] MRC-Holland, MLPA (Multiplex Ligation-dependent Probe Amplification) General Protocol Instructions For Use, 2019. https://www.mrcholland.com/technology/mlpa/technique. (accessed June 15, 2022).

[47] J.P. Schouten, C.J. McElgunn, R. Waaijer, D. Zwijnenburg, F. Diepvens, G. Pals, Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification, Nucleic Acids Res. 30 (2002) 1–13. https://doi.org/10.1093/nar/gnf056.

[48] L. Stuppia, I. Antonucci, G. Palka, V. Gatta, Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases, Int. J. Mol. Sci. 13 (2012) 3245–3276. https://doi.org/10.3390/ijms13033245.

[49] MRC-Holland, Product Description SALSA MLPA Probemix P140-C1 HBA, 2008. https://www.mrcholland.com/product/P140/1740?. (accessed May 16, 2022).

[50] I. Adzhubei, D.M. Jordan, S.R. Sunyaev, Predicting functional effect of human missense mutations using PolyPhen-2, 2013. https://doi.org/10.1002/0471142905.hg0720s76.

[51] I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov, S.R. Sunyaev, A method ans server for predicting damaging missense mutations, Nat. Methods. 7 (2010) 248–249. https://doi.org/10.1038/nmeth0410-248.

[52] P.C. Ng, S. Henikoff, SIFT: Predicting amino acid changes that affect protein function, Nucleic Acids Res. 31 (2003) 3812–3814. https://doi.org/10.1093/nar/gkg509.

[53] J. Hu, P.C. Ng, SIFT Indel: Predictions for the functional effects of amino acid insertions/deletions in proteins, PLOS ONES. 8 (2013) 1–8. https://doi.org/10.1371/journal.pone.0077940.

[54] C. Keohane, M.F. McMullin, C. Harrison, The diagnosis and management of erythrocytosis, BMJ. 347 (2013) 1–6. https://doi.org/10.1136/bmj.f6667.

[55] F. Mavilio, M. Marinucci, L. Tentori, P.P. Fontanarosa, U. Rossi, S. Biagiotti, Hemoglobin legnano (α2 141 (HC3) arg → leu β2): A new abnormal human hemoglobin with high oxygen affinity, Hemoglobin. 2 (1978) 249–259. https://doi.org/10.3109/03630267809007070.

[56] A. Giuliani, D. Maffi, M.P. Cappabianca, L. Tentori, Hemoglobin Legnano (α2 141 (HC3) Arg→Leu β2) A New High Oxygen Affinity Variant, J. Biochem. 88 (1980) 1233–1237. https://doi.org/10.1093/oxfordjournals.jbchem.a133091.

[57] W.B. Qin, X.L. Yue, L.Y. Qin, T.L. Jü, J.B. Wilson, L.H. Gu, T.H.J. Huisman, Two rare hemoglobin variants: Hb Pyrgos [β83(ef7)gly->asp] and hb legnano [α141(hc3)arg->leu] found in inner Mongolia, P.R. China, Hemoglobin. 18 (1994) 343–345. https://doi.org/10.3109/03630269408996201.

[58] V. Douna, I. Papassotiriou, A. Stamoulakatou, A. Metaxotou-Mavrommati, E. Kanavakis, J. Traeger-Synodinos, Association of Mild and Severely Unstable α Chain Variants: The First Observation of a Compound Heterozygote with Hb Setif [α94(G1)Asp→Tyr (α2)] and Hb Agrinio [α29(B10)Leu→Pro (α2)] in a Greek Family, Hemoglobin. 32 (2008) 592–595. https://doi.org/10.1080/03630260802507964.

[59] M.R. Mahdavi, M. Karimi, M. Yavarian, P. Roshan, M. Kosaryan, R. Siami, Detection of Hb Setif in north Iran and the question of its origin: Iranian or multiethnic?, Hemoglobin. 35 (2011) 152–156. https://doi.org/10.3109/03630269.2011.557461.

[60] S. Farashi, N.F. Garous, S. Vakili, M. Ashki, H. Imanian, A. Azarkeivan, H. Najmabadi, Characterization of homozygous Hb Setif (HBA2: c.283G>T) in the Iranian population, Hemoglobin. 40 (2016) 53–55. https://doi.org/10.3109/03630269.2015.1091357.

[61] G. Nozari, S. Ralthar, P. Darbre, H. Lehmann, Hemoglobin Setif (α94 (G1) ASP → TYR) in Iran a Report of 9 Cases, Hemoglobin. 1 (1977) 289–292. https://doi.org/10.3109/03630267709003412.

[62] B. Al-Awamy, G.A. Niazi, J.B. Wilson, T.H.H. Huisman, Hb Setif or α294(G1)Asp→Tyrβ2 Observed in a Saudi Arabian Family, Hemoglobin. 9 (1985) 87–90. https://doi.org/10.3109/03630268508996987.

[63] M.Z. Abdo, HB Setif [α94(Gl)ASP→TYR) in a Saudi Arabian Family, Hemoglobin. 13 (1989) 737–742. https://doi.org/10.3109/03630268908998847.

[64] G. Dinçol, D. Elam, A. Kutlar, F. Kutlar, Hb Setif [α94(G1)Asp→Tyr (α2)] Detected in a Turkish Family, Hemoglobin. 27 (2003) 249–252. https://doi.org/10.1081/HEM-120026050.

[65] I. Borg, M. Valentino, A. Fiorini, A.E. Felice, Hb Setif [alpha 94(G1)Asp-->Tyr] in Malta., Hemoglobin. 21 (1997) 91–6. https://doi.org/10.3109/03630269708997513.

[66] S.B. Wettinger, R. Galdies, C. Scerri, A.E. Felice, Characterization and locus assignment of two α-globin variants present in the maltese population: Hb St. Luke's [α95(G2)Pro→Arg] and Hb Setif [α94(G1)Asp→Tyr], Hemoglobin. 23 (1999) 145–157. https://doi.org/10.3109/03630269908996159.

[67] G. Schilirò, G. Russo-Mancuso, S.P. Dibenedetto, P. Samperi, A. Di Cataldo, R. Ragusa, R. Testa, Six Rare Hemoglobin Variants Found in Sicily, Hemoglobin. 15 (1991) 431–437. https://doi.org/10.3109/03630269108998862.

[68] J.M. de Pablos, M.A. de las Nieves, A. Romero, M. Jurado, A. García, W.F. Moo Penn, [Hemoglobin Setif (alpha 94 (G1) Asp----Tyr) in a Spanish family]., Sangre (Barc). 35 (1990) 201–4.

[69] H. Wajcman, O. Belkhodja, D. Labie, Hb Setif: G1 (94)α Asp → Tyr a new α chain hemoglobin variant with substitution of the residue involved in a hydrogen bond between unlike subunits, FEBS Lett. 27 (1972) 298–300. https://doi.org/10.1016/0014-5793(72)80645-8.

[70] R.G. Schneider, B. Hightower, U. Carpentieri, M.L. Duerst, T.B. Shin, R.T. Jones, Hemoglobin oleander [α2116(GH4)GLU → GLNβ2]: Structural and functional characterization, Hemoglobin. 6 (1982) 465–480. https://doi.org/10.3109/03630268209083760.

[71] J.B. Wilson, R.N. Wrightstone, T.H.J. Huisman, HB Oleander, α116(GH4)GLLàGLN, Found in a Black Female in Georgia, Hemoglobin. 8 (1984) 239–240. https://doi.org/10.3109/03630268408996972.

[72] H. Wajcman, J. Traeger-Synodinos, I. Papassotiriou, P.C. Giordano, C.L. Harteveld, V. Baudin-Creuza, J. Old, Unstable and Thalassemic α Chain Hemoglobin Variants: A Cause of Hb H Disease and Thalassemia Intermedia, Hemoglobin. 32 (2008) 327–349. https://doi.org/10.1080/03630260802173833.

[73] T.M. Ko, L.H. Tseng, F.J. Hsieh, P.M. Hsu, T.Y. Lee, Carrier detection and prenatal diagnosis of alpha-thalassemia of Southeast Asian deletion by polymerase chain reaction, Hum. Genet. 88 (1992) 245–248. https://doi.org/10.1007/BF00197254.

[74] Q.-W. Qiu, D.-D. Wu, L.-H. Yu, T.-Z. Yan, W. Zhang, Z.-T. Li, Y.-H. Liu, Y.-P. Zhang, X.-M. Xu, Evidence of recent natural selection on the Southeast Asian deletion (--SEA) causing α-thalassemia in South China, BMC Evol. Biol. 13 (2013) 63. https://doi.org/10.1186/1471-2148-13-63.

[75] W. Jomoui, G. Fucharoen, K. Sanchaisuriya, P. Charoenwijitkul, J. Maneesarn, X. Xu, S. Fucharoen, Genetic origin of α 0-thalassemia (SEA deletion) in Southeast Asian populations and application to accurate prenatal diagnosis of Hb Bart's hydrops fetalis syndrome, J. Hum. Genet. 62 (2017) 747–754. https://doi.org/10.1038/jhg.2017.41.

[76] M. Phylipsen, J.F. Prior, E. Lim, N. Lingam, I.P. Vogelaar, P.C. Giordano, J. Finlayson, C.L. Harteveld, Thalassemia in Western Australia: 11 novel deletions characterized by Multiplex Ligation-dependent Probe Amplification, Blood Cells, Mol. Dis. 44 (2010) 146–151. https://doi.org/10.1016/j.bcmd.2009.12.011.

[77] A.O.M. Wilkie, J. Lamb, P.C. Harris, R.D. Finney, D.R. Higgs, A truncated human chromosome 16 associated with α thalassaemia is stabilized by addition of telomeric repeat (TTAGGG)n, Nature. 346 (1990) 868–871. https://doi.org/10.1038/346868a0.

[78] M.C. Sollaino, M.E. Paglietti, D. Loi, R. Congiu, R. Podda, R. Galanello, Homozygous deletion of the major alpha-globin regulatory element (MCS-R2) responsible for a severe case of hemoglobin H disease, Blood. 116 (2010) 2193–2194. https://doi.org/10.1182/blood-2010-04-281345.

[79] L. Romao, L. Osorio-Almeida, D. Higgs, J. Lavinha, S. Liebhaber, Alpha-thalassemia resulting from deletion of regulatory sequences far upstream of the alpha-globin structural genes, Blood. 78 (1991) 1589–1595. https://doi.org/10.1182/blood.V78.6.1589.1589.

[80] M.R.S.C. Wenning, C.L. Harteveld, P.C. Giordano, E.M. Kimura, S.T.O. Saad, F.F. Costa, M.F. Sonati, Hemoglobin H disease resulting from the association of the - α 3.7 rightward deletion and the (αα) MM deletion in a Brazilian patient, Eur. J. Haematol. 69 (2002) 179–181. https://doi.org/10.1034/j.1600-0609.2002.02743.x.

[81] N.O. Mota, E.M. Kimura, R.D. Ferreira, G.A. Pedroso, D.M. Albuquerque, D.M. Ribeiro, M.N.N. Santos, C.M. Bittar, F.F. Costa, M. de F. Sonati, Rare α0-thalassemia deletions detected by MLPA in five unrelated Brazilian patients, Genet. Mol. Biol. 40 (2017) 768–773. https://doi.org/10.1590/1678-4685-gmb-2016-

0330.

[82] V. Viprakasit, C.L. Harteveld, H. Ayyub, J.S. Stanley, P.C. Giordano, W.G. Wood, D.R. Higgs, A novel deletion causing α thalassemia clarifies the importance of the major human alpha globin regulatory element, Blood. 107 (2006) 3811–3812. https://doi.org/10.1182/blood-2005-12-4834.

[83] S. Luo, X. Chen, Q. Zhong, Q. Wang, Z. Xu, L. Qin, J. Wang, D. Yuan, T. Yan, N. Tang, Analysis of rare thalassemia caused by HS-40 regulatory site deletion, Hematol. (United Kingdom). 25 (2020) 286–291. https://doi.org/10.1080/16078454.2020.1799587.

[84] G. Van Ros, D. Beale, H. Lehmann, Haemoglobin Stanleyville II (α78 Asparagine → Lysine), Br. Med. J. 4 (1968) 92–93. https://doi.org/10.1136/bmj.4.5623.92.

[85] L. Chen, L. Liang, S. Liao, Y. Huang, L. Liang, Y. Li, Compound heterozygotes of Hb Constant Spring and Hb Stanleyville II in HbE/β 0 -thalassemia, Clin. Chem. Lab. Med. 59 (2021) e374–e376. https://doi.org/10.1515/cclm-2020-1336.

[86] F.F. Costa, M.F. Sonati, M.A. Zago, Hemoglobin Stanleyville II (α78 Asn→Lys) is associated with a 3.7-kb α-globin gene deletion, Hum. Genet. 86 (1991) 319–320. https://doi.org/10.1007/BF00202420.

[87] N.H. Jiang, S. Liang, X.J. Wen, R. Liang, C. Su, Z. Tang, Hb Westmead: An α2-globin gene mutation detected by polymerase chain reaction and STU i cleavage, Hemoglobin. 15 (1991) 291–295. https://doi.org/10.3109/03630269109027881.

[88] W. Wong, A.Y. Chan, S. Yip, E. Shiu-Kwan Ma, Thalassemia Intermedia Due to Co-inheritance of β 0 /β + -Thalassemia and (− − SEA) α- Thalassemia/Hb Westmead [α 122(H5)His > Gln (α 2)] in a Chinese Family, Hemoglobin. 28 (2004) 151–156. https://doi.org/10.1081/HEM-120035917.

[89] P.J. Fleming, W.G. Hughes, R.K. Farmilo, K. Wyatt, W.N. Cooper, Hemoglobin Westmead α 2 122(H5)His←Gln β 2 : A New Hemoglobin Variant with the Substitution in the α 1 β 1 Contact Area, Hemoglobin. 4 (1980) 39–52. https://doi.org/10.3109/03630268009042372.

[90] V. Viprakasit, S. Ekwattanakit, N. Chalaow, S. Riolueang, S. Wijit, P. Tanyut, N. Chat-Uthai, K. Tachavanich, Clinical presentation and molecular identification of four uncommon alpha globin variants in Thailand: Initiation codon mutation of α2-Globin Gene (HBA2:c.1delA), donor splice site mutation of α1-globin gene (IVSI-1, HBA1:c.95 + 1G>A), hemoglobin Queens P, Acta Haematol. 131 (2014) 88–94. https://doi.org/10.1159/000353119.

[91] Y.-C. Gu, L.-H. Gu, J.B. Wilson, B. Cepreganova, M. Ramachandran, E.L.D. Walker, T.H.J. Huisman, P. Potitong, Hb Westmead [α122(H5)HIS→GLN], Hb E [β26(B8)6Ll→LYS], and α-Thalassemia-2 (3.7 KB Deletion) in a Laotian Family, Hemoglobin. 15 (1991) 297–302. https://doi.org/10.3109/03630269109027882.

[92] E. Vichinsky, Hemoglobin E Syndromes, Hematology. 2007 (2007) 79–83. https://doi.org/10.1182/asheducation-2007.1.79.

[93] N.F. Olivieri, Z. Pakbaz, E. Vichinsky, HbE/β-thalassemia: Basis of marked clinical diversity, Hematol. Oncol. Clin. North Am. 24 (2010) 1055–1070. https://doi.org/10.1016/j.hoc.2010.08.008.

[94] S. Fucharoen, D.J. Weatherall, The Hemoglobin E Thalassemias, Cold Spring Harb. Perspect. Med. 2 (2012) 1–16. https://doi.org/10.1101/cshperspect.a011734.

[95] A.M. Soler, B.F. Piellusch, L. da Silveira, G.A. Pedroso, P. López, E. Savio, M.D.F. Sonati, J. da Luz, Alpha thalassemia and alpha-MRE haplotypes in Uruguayan patients with microcytosis and hypochromia without anemia, Genet. Mol. Biol. 44 (2021) 1–6. https://doi.org/10.1590/1678-4685-gmb-2020-0399.

[96] S. Alimohammadi-Bidhendi, S. Azadmehr, M. Razipour, S. Zeinali, M. Eslami, E. Davoudi-Dehaghani, Regulatory Mutation Study in Cases with Unsolved Hypochromic Microcytic Anemia and α-Major Regulatory Element Haplotype Analysis in Iran, Hemoglobin. 45 (2021) 37–40. https://doi.org/10.1080/03630269.2021.1882482.

[97] S.H. Orkin, Regulation of Globin Gene Expression in Erythroid Cells, Eur. J. Biochem. 231 (1995) 271–281. https://doi.org/10.1111/j.1432-1033.1995.tb20697.x.

[98] M.R. Loyd, Y. Okamoto, M.S. Randall, P.A. Ney, Role of AP1/NFE2 binding sites in endogenous α-globin gene transcription, Blood. 102 (2003) 4223–4228. https://doi.org/10.1182/blood-2003-02-0574.

[99] N.C. Andrews, H. Erdjument-Bromage, M.B. Davidson, P. Tempst, S.H. Orkin, Erythroid transcription factor NF-E2 is a haematopoietic-specific basic–leucine zipper protein, Nature. 362 (1993) 722–728. https://doi.org/10.1038/362722a0.

[100] R.A. Shivdasani, S.H. Orkin, Erythropoiesis and globin gene expression in mice lacking the transcription

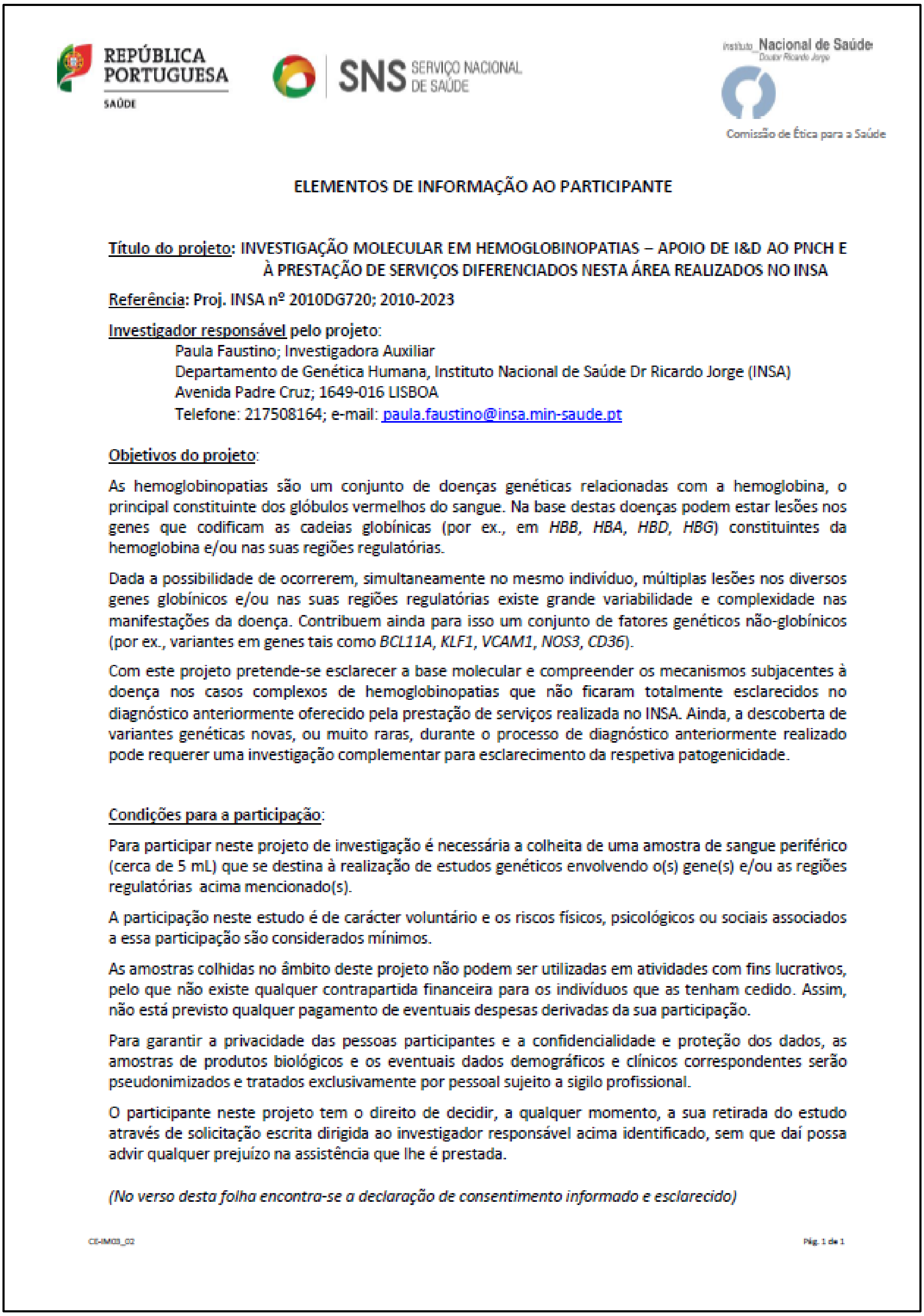
factor NF-E2., Proc. Natl. Acad. Sci. 92 (1995) 8690–8694. https://doi.org/10.1073/pnas.92.19.8690.

[101] D.R. Higgs, W.G. Wood, A.P. Jarman, J. Sharpe, J. Lida, I.M. Pretorius, H. Ayyub, A major positive regulatory region located far upstream of the human alpha-globin gene locus., Genes Dev. 4 (1990) 1588–1601. https://doi.org/10.1101/gad.4.9.1588.

[102] E. Anguita, J.A. Sharpe, J.A. Sloane-Stanley, C. Tufarelli, D.R. Higgs, W.G. Wood, Deletion of the mouse α-globin regulatory element (HS −26) has an unexpectedly mild phenotype, Blood. 100 (2002) 3450–3456. https://doi.org/10.1182/blood-2002-05-1409.

[103] Lesson Explainer: Translation, Nagwa. (2022). https://www.nagwa.com/en/explainers/246104727272/ (accessed May 20, 2022).

7. Supplementary materials



REPÚBLICA PORTUGUESA
SAÚDE

SNS SERVIÇO NACIONAL DE SAÚDE

Instituto Nacional de Saúde Doutor Ricardo Jorge

Comissão de Ética para a Saúde

ELEMENTOS DE INFORMAÇÃO AO PARTICIPANTE

Título do projeto: INVESTIGAÇÃO MOLECULAR EM HEMOGLOBINOPATIAS – APOIO DE I&D AO PNCH E À PRESTAÇÃO DE SERVIÇOS DIFERENCIADOS NESTA ÁREA REALIZADOS NO INSA

Referência: Proj. INSA nº 2010DG720; 2010-2023

Investigador responsável pelo projeto:
Paula Faustino; Investigadora Auxiliar
Departamento de Genética Humana, Instituto Nacional de Saúde Dr Ricardo Jorge (INSA)
Avenida Padre Cruz; 1649-016 LISBOA
Telefone: 217508164; e-mail: paula.faustino@insa.min-saude.pt

Objetivos do projeto:

As hemoglobinopatias são um conjunto de doenças genéticas relacionadas com a hemoglobina, o principal constituinte dos glóbulos vermelhos do sangue. Na base destas doenças podem estar lesões nos genes que codificam as cadeias globínicas (por ex., em *HBB*, *HBA*, *HBD*, *HBG*) constituintes da hemoglobina e/ou nas suas regiões regulatórias.

Dada a possibilidade de ocorrerem, simultaneamente no mesmo indivíduo, múltiplas lesões nos diversos genes globínicos e/ou nas suas regiões regulatórias existe grande variabilidade e complexidade nas manifestações da doença. Contribuem ainda para isso um conjunto de fatores genéticos não-globínicos (por ex., variantes em genes tais como *BCL11A*, *KLF1*, *VCAM1*, *NOS3*, *CD36*).

Com este projeto pretende-se esclarecer a base molecular e compreender os mecanismos subjacentes à doença nos casos complexos de hemoglobinopatias que não ficaram totalmente esclarecidos no diagnóstico anteriormente oferecido pela prestação de serviços realizada no INSA. Ainda, a descoberta de variantes genéticas novas, ou muito raras, durante o processo de diagnóstico anteriormente realizado pode requerer uma investigação complementar para esclarecimento da respetiva patogenicidade.

Condições para a participação:

Para participar neste projeto de investigação é necessária a colheita de uma amostra de sangue periférico (cerca de 5 mL) que se destina à realização de estudos genéticos envolvendo o(s) gene(s) e/ou as regiões regulatórias acima mencionado(s).

A participação neste estudo é de carácter voluntário e os riscos físicos, psicológicos ou sociais associados a essa participação são considerados mínimos.

As amostras colhidas no âmbito deste projeto não podem ser utilizadas em atividades com fins lucrativos, pelo que não existe qualquer contrapartida financeira para os indivíduos que as tenham cedido. Assim, não está previsto qualquer pagamento de eventuais despesas derivadas da sua participação.

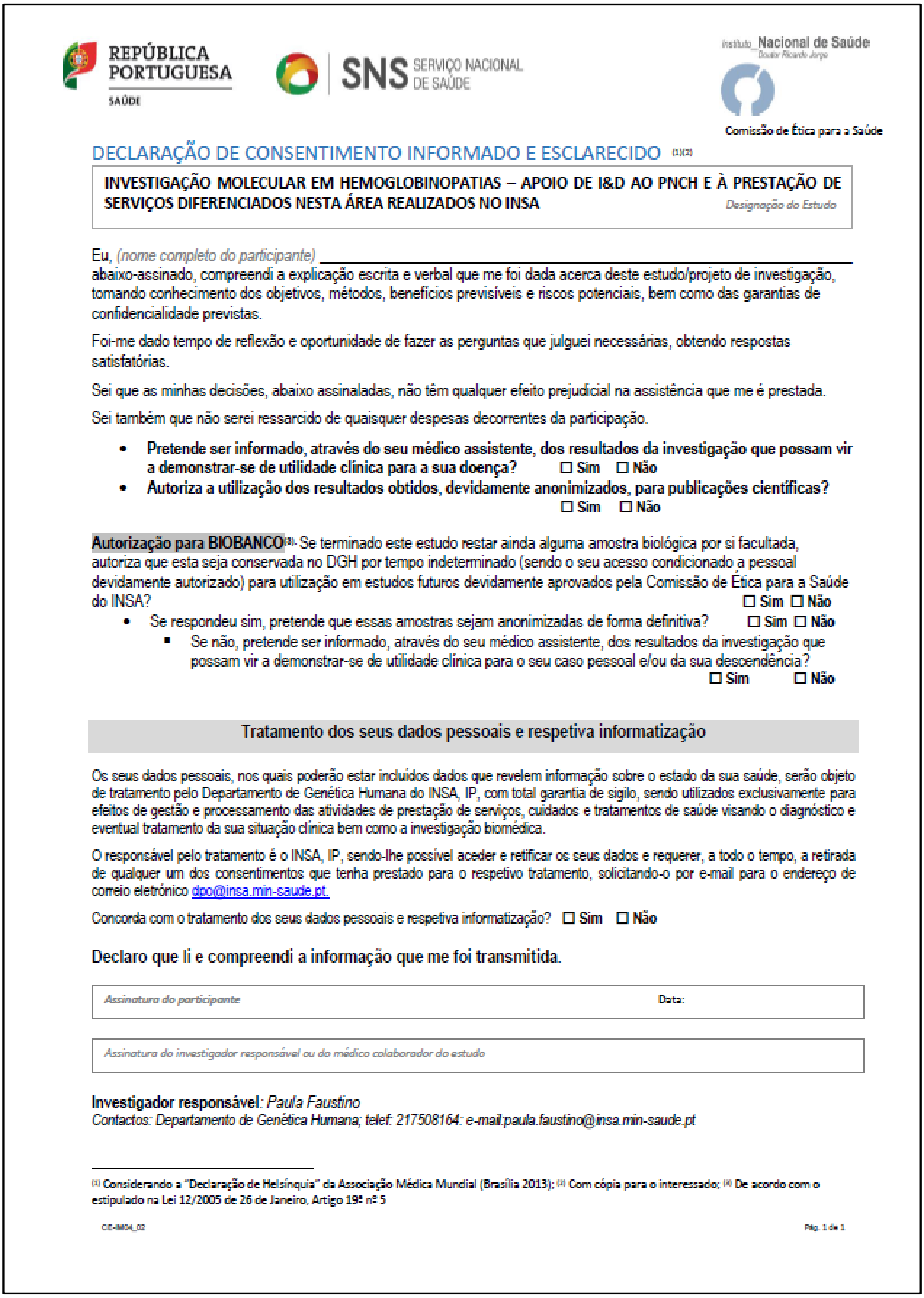
Para garantir a privacidade das pessoas participantes e a confidencialidade e proteção dos dados, as amostras de produtos biológicos e os eventuais dados demográficos e clínicos correspondentes serão pseudonimizados e tratados exclusivamente por pessoal sujeito a sigilo profissional.

O participante neste projeto tem o direito de decidir, a qualquer momento, a sua retirada do estudo através de solicitação escrita dirigida ao investigador responsável acima identificado, sem que daí possa advir qualquer prejuízo na assistência que lhe é prestada.

(No verso desta folha encontra-se a declaração de consentimento informado e esclarecido)

CE-IM03_02 Pág. 1 de 1

Figure 7.1 - Information given to the adult participants for the molecular study of hemoglobinopathies.



REPÚBLICA PORTUGUESA
SAÚDE

SNS SERVIÇO NACIONAL DE SAÚDE

Instituto Nacional de Saúde Doutor Ricardo Jorge

Comissão de Ética para a Saúde

DECLARAÇÃO DE CONSENTIMENTO INFORMADO E ESCLARECIDO [1][2]

INVESTIGAÇÃO MOLECULAR EM HEMOGLOBINOPATIAS – APOIO DE I&D AO PNCH E À PRESTAÇÃO DE SERVIÇOS DIFERENCIADOS NESTA ÁREA REALIZADOS NO INSA

Designação do Estudo

Eu, *(nome completo do participante)* ____
abaixo-assinado, compreendi a explicação escrita e verbal que me foi dada acerca deste estudo/projeto de investigação, tomando conhecimento dos objetivos, métodos, benefícios previsíveis e riscos potenciais, bem como das garantias de confidencialidade previstas.

Foi-me dado tempo de reflexão e oportunidade de fazer as perguntas que julguei necessárias, obtendo respostas satisfatórias.

Sei que as minhas decisões, abaixo assinaladas, não têm qualquer efeito prejudicial na assistência que me é prestada.

Sei também que não serei ressarcido de quaisquer despesas decorrentes da participação.

- **Pretende ser informado, através do seu médico assistente, dos resultados da investigação que possam vir a demonstrar-se de utilidade clínica para a sua doença?** ☐ **Sim** ☐ **Não**
- **Autoriza a utilização dos resultados obtidos, devidamente anonimizados, para publicações científicas?** ☐ **Sim** ☐ **Não**

Autorização para BIOBANCO[3]. Se terminado este estudo restar ainda alguma amostra biológica por si facultada, autoriza que esta seja conservada no DGH por tempo indeterminado (sendo o seu acesso condicionado a pessoal devidamente autorizado) para utilização em estudos futuros devidamente aprovados pela Comissão de Ética para a Saúde do INSA? ☐ **Sim** ☐ **Não**

- Se respondeu sim, pretende que essas amostras sejam anonimizadas de forma definitiva? ☐ **Sim** ☐ **Não**
  - Se não, pretende ser informado, através do seu médico assistente, dos resultados da investigação que possam vir a demonstrar-se de utilidade clínica para o seu caso pessoal e/ou da sua descendência? ☐ **Sim** ☐ **Não**

Tratamento dos seus dados pessoais e respetiva informatização

Os seus dados pessoais, nos quais poderão estar incluídos dados que revelem informação sobre o estado da sua saúde, serão objeto de tratamento pelo Departamento de Genética Humana do INSA, IP, com total garantia de sigilo, sendo utilizados exclusivamente para efeitos de gestão e processamento das atividades de prestação de serviços, cuidados e tratamentos de saúde visando o diagnóstico e eventual tratamento da sua situação clínica bem como a investigação biomédica.

O responsável pelo tratamento é o INSA, IP, sendo-lhe possível aceder e retificar os seus dados e requerer, a todo o tempo, a retirada de qualquer um dos consentimentos que tenha prestado para o respetivo tratamento, solicitando-o por e-mail para o endereço de correio eletrónico dpo@insa.min-saude.pt.

Concorda com o tratamento dos seus dados pessoais e respetiva informatização? ☐ **Sim** ☐ **Não**

Declaro que li e compreendi a informação que me foi transmitida.

Assinatura do participante	**Data:**

Assinatura do investigador responsável ou do médico colaborador do estudo

Investigador responsável: *Paula Faustino*
Contactos: Departamento de Genética Humana; telef. 217508164: e-mail:paula.faustino@insa.min-saude.pt

---

[1] Considerando a "Declaração de Helsínquia" da Associação Médica Mundial (Brasília 2013); [2] Com cópia para o interessado; [3] De acordo com o estipulado na Lei 12/2005 de 26 de Janeiro, Artigo 19º nº 5

CE-IM04_02 Pág. 1 de 1

Figure 7.2 - Informed consent signed by adult participants for the molecular study of hemoglobinopathies.

Table 7.1 - Conventional PCR conditions for the amplification of the genes *HBA2* and *HBA1*

Genes	Primers				PCR Mixture				PCR reaction		
	ID	Sequence	Size (nt)	Tm (°C)		Reagents	Concentration	Volume (µL)	Temperature (°C)	Time	
HBA2 and *HBA1*	A_5'UTR_F1 (Fw)	5' - GGACTCCCCTGCGGTCCAGG - 3'	20	62.0	Mixture I	α Buffer + βME	10x	2.5			
						DMSO	5%	2.5			
						BSA	10 mg/mL	0.4			
						MgCl$_2$	0.1 M	0.5			
						dNTPs	100 mM	0.5			
						Primer A_5'UTR_F1 (Fw)	25 pmol/µL	0.5			
	C3 (Rv)	5' - CTCCATTGTTGGCACATTCCGGG - 3'	23	58.8		Primer C3 (Rv)	25 pmol/µL	0.5			
						H$_2$O	-	16.3	95	10 min	
						AmpliTaq	5 U/µL	0.3	95	1 min	x 32
						DNA	≈ 100 ng	1.0	61	1 min	
					Mixture II	α Buffer + βME	10x	2.5	72	1 min	
						DMSO	5%	2.5	72	10 min	
						BSA	10 mg/mL	0.4	4	10 min	
	C2 (Rv)	5' - CTGCTGTCCACGCCCATGCC - 3'	20	60.0		MgCl$_2$	0.1 M	0.5	-	15 min	
						dNTPs	100 mM	0.5			
						Primer A_5'UTR_F1 (Fw)	25 pmol/µL	0.5			
						Primer C2 (Rv)	25 pmol/µL	0.5			
						H$_2$O	-	16.3			
						AmpliTaq	5 U/µL	0.3			
						DNA	≈ 100 ng	1.0			

ID – Identification; **Fw** – Forward; **Rv** – Reverse; **nt** – Nucleotides; **Tm** – Melting Temperature; **βME** – β-mercaptoethanol; **DMSO** – Dimethyl Sulfoxide; **BSA** – Bovine Serum Albumin; **MgCl₂** – Magnesium Chloride; **dNTPs** – Deoxyribonucleotide Triphosphate; AmpliTaq® DNA Polymerase (*Applied Biosystems*); **min** - Minutes

Table 7.2 - Conventional PCR conditions for the amplification of the regulatory region HS-40

Region	Primers				PCR mixture			PCR reaction		
	ID	Sequence	Size (nt)	Tm (°C)	Reagents	Concentration	Volume (µL)	Temperature (°C)	Time	
αMRE/HS-40	h-M13-αMRE (Fw)	5' - CGACGTTGTAAAACGACGGCCAGTCGGCCGGGAGGCTCTCAGGAACAAGA - 3'	50	76.1	α Buffer + βME	10x	2.5			
					DMSO	5%	2.5			
					BSA	10 mg/mL	0.4	94	5 min	
					$MgCl_2$	0.1 M	0.5	94	1 min	
					dNTPs	100 mM	0.5	61	1 min	x 34
	h-bio-αMRE (Rv)	5' - CGCCCGCCCCGCCCCCGTGCCCCCCGCGCCGCCCGCCGCGGCCGCAGCCTGGCTGTGAACACTTT - 3'	65	88.6	Primer h-M13-αMRE (Fw)	25 pmol/µL	0.5	72	1 min	
					Primer h-bio-αMRE (Rv)	25 pmol/µL	0.5	72	10 min	
					H_2O	-	16.3	4	10 min	
					AmpliTaq	5 U/µL	0.1	-	15 min	
					DNA	≈ 100 ng	1.0			

αMRE – α-Major Regulatory Element; **ID** – Identification; **Fw** – Forward; **Rv** – Reverse; **nt** – Nucleotides; **Tm** – Melting Temperature; **βME** – β-mercaptoethanol; **DMSO** – Dimethyl Sulfoxide; **BSA** – Bovine Serum Albumin; **$MgCl_2$** – Magnesium Chloride; **dNTPs** – Deoxyribonucleotide Triphosphate; AmpliTaq® DNA Polymerase (*Applied Biosystems*); **min** – Minutes

Table 7.3 – Gap-PCR conditions for the amplification of a fragment containing the $(\alpha\alpha)^{ALT}$ deletion

Gene	Primers				PCR mixture			PCR reaction		
	ID	Sequence	Size (nt)	Tm (°C)	Reagents	Concentration	Volume (µL)	Temperature (°C)	Time	
$(\alpha\alpha)^{ALT}$ deletion	C16orf35.2 (Fw)	5' GCACAGGGACACAGCTGGACAC 3'	22	60.4	LA PCR Buffer II Mg^{2+} plus	10 x	2.50			
					BSA	10 mg/mL	0.35	95	5 min	
					dNTPs	25 mM	4.00	95	1 min	
					Primer C16orf35.2 (Fw)	15 pmol/µL	0.50	65	1 min	x 30
	C16orf35.5 (Rv)	5' GATCAGGGAGTGGGGCCAGTGG 3'	22	62.3	Primer C16orf35.5 (Rv)	15 pmol/µL	0.50	72	15 sec	
					H_2O	-	16.9	72	10 min	
					TaKaRa LA Taq Hs	1.25 U/µL	0.25	4	-	
					DNA	≈ 2 ng	1.0			

ID – Identification; **Fw** – Forward; **Rv** – Reverse; **nt** – Nucleotides; **Tm** – Melting Temperature; **BSA** – Bovine Serum Albumin; **dNTPs** – Deoxyribonucleotide Triphosphate; **min** – Minutes; **sec** - Seconds

Table 7.4 - Composition of the buffer solutions used during the study

α Buffer		TBE Buffer (10x)		Bromophenol Blue Solution (10 mg/mL)	
$(NH_4)_2SO_4$	166 mM	Tris HCl	108 g	Glicerol	18.75 µL
Tris HCl	670 mM	Boric acid	55 g	EDTA	75.00 µL
$MgCl_2$	15 mM	EDTA	7.4 g	Bromophenol Blue	0.22 g
EDTA	0.67 mM	H_2O	until 1L	H_2O	8.75 µL
H_2O	-	**TBE Buffer (1x)**		NaOH (until it turns blue)	
β-mercaptoethanol	100 mM	250 mL TBE (10x) + H_2O until 2.5L			

$(NH_4)_2SO_4$ – Ammonium Sulfate; **HCl** – Hydrochloric Acid; **$MgCl_2$** – Magnesium Chloride; **EDTA** – Ethylenediamine Tetraacetic Acid; **TBE** – Tris-Borate-EDTA; **NaOH** – Sodium

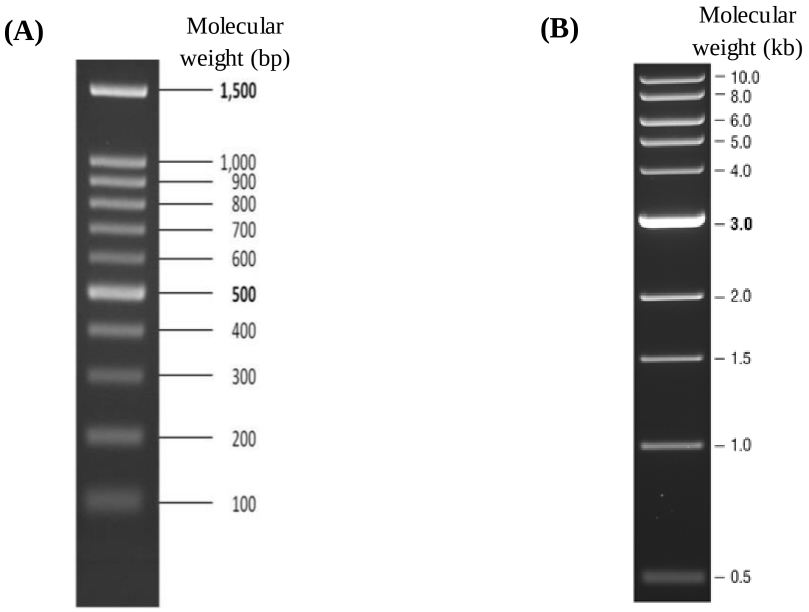


Figure 7.3 - Molecular markers used in electrophoresis performed in an agarose gel. (A) 100 bp plus DNA Ladder (*BIORON*) marker in a 1.5% agarose gel. **(B)** 1 kb DNA Ladder (BioLabs) marker in a 0.8% agarose gel. The molecular weight of each band is presented in base pairs (bp) and kilobases (kb), respectively.

Table 7.5 - Gap-PCR conditions for the detection of the 3.7kb α-thalassemia deletion

Genes	Primers				PCR Mixture				PCR reaction		
	ID	Sequence	Size (nt)	Tm (°C)		Reagents	Concentration	Volume (µL)	Temperature (°C)	Time	
HBA2, *HBA1* and Hybrid Gene	C10 (Fw)	5' - GGGATGCACCCACTGGCACT - 3'	20	57.9	Mixture I	α Buffer + βME	10x	2.5			
						DMSO	5%	2.5			
						BSA	10 mg/mL	0.4			
						$MgCl_2$	0.1 M	0.5			
						dNTPs	100 mM	0.5			
						Primer C10 (Fw)	25 pmol/µL	0.5			
	C3 (Rv)	5' - CTCCATTGTTGGCACATTCCGGG - 3'	23	58.8		Primer C3 (Rv)	25 pmol/µL	0.5			
						H_2O	-	16.3	94	5 min	
						AmpliTaq	5 U/µL	0.3	94	1 min	x 34
						DNA	≈ 100 ng	1.0	64	1 min	
					Mixture II	α Buffer + βME	10x	2.5	72	1 min	
						DMSO	5%	2.5	72	10 min	
						BSA	10 mg/mL	0.4	4	10 min	
	C2 (Rv)	5' - CTGCTGTCCACGCCCATGCC - 3'	20	60.0		$MgCl_2$	0.1 M	0.5	-	15 min	
						dNTPs	100 mM	0.5			
						Primer C10 (Fw)	25 pmol/µL	0.5			
						Primer C2 (Rv)	25 pmol/µL	0.5			
						H_2O	-	16.3			
						AmpliTaq	5 U/µL	0.3			
						DNA	≈ 100 ng	1.0			

ID – Identification; **Fw** – Forward; **Rv** – Reverse; **nt** – Nucleotides; **Tm** – Melting Temperature; **βME** – β-mercaptoethanol; **DMSO** – Dimethyl Sulfoxide; **BSA** – Bovine Serum Albumin; **$MgCl_2$** – Magnesium Chloride; **dNTPs** – Deoxyribonucleotide Triphosphate; AmpliTaq® DNA Polymerase (*Applied Biosystems*); **min** – Minutes

Table 7.6 - Multiplex Gap-PCR conditions for the detection of the 3.7kb and 4.2kb α-thalassemia deletions

Gene	Primers				PCR mixture			PCR reaction		
	ID	Sequence	Size (nt)	Tm (°C)	Reagents	Concentration	Volume (µL)	Temperature (°C)	Time	
HBA2, *HBA1* and Hybrid Gene	α2/3.7-F (Fw)	5' - CCCCTCGCCAAGTCCACCC - 3'	19	59.7	LA PCR Buffer II Mg^{2+} plus	10x	2.50			
					BSA	10 mg/mL	0.35			
	3.7/20.5-R (Rv)	5' - AAAGCACTCTAGGGTCCAGCG - 3'	21	56.3	dNTPs	100 mM	4.00	96	15 min	
					Primer α2/3.7-F (Fw)	50 pmol/µL	1.00	98	45 sec	
	α2-R (Rv)	5' - AGACCAGGAAGGGCCGGTG - 3'	19	57.6	Primer 3.7/20.5-R (Rv)	25 pmol/µL	0.50	64	90 sec	x 30
					Primer α2-R (Rv)	25 pmol/µL	0.50	72	120 sec	
					Primer 4.2-F (Fw)	25 pmol/µL	0.50	72	5 min	
	4.2-F (Fw)	5' - GGTTTACCCATGTGGTGCCTC - 3'	21	56.3	Primer 4.2-R (Rv)	25 pmol/µL	0.50	4	10 min	
					H_2O	-	14.9	-	15 min	
	4.2-R (Rv)	5' - CCCGTTGGATCTTCTCATTTCCC - 3'	23	57.1	TaKaRa LA Taq Hs	1.25 U/µL	0.25			
					DNA	≈ 15 ng	1.00			

ID – Identification; **Fw** – Forward; **Rv** – Reverse; **nt** – Nucleotides; **Tm** – Melting Temperature; **BSA** – Bovine Serum Albumin; **dNTPs** – Deoxyribonucleotide Triphosphate; **min** – Minutes; **sec** - Seconds

Table 7.7 - Conditions for Sanger sequencing

PCR mixture			PCR reaction		
Reagents	Concentration	Volume (µL)	Temperature (°C)	Time	
Big Dye Buffer	5x	1.75	96	4 min	
Big Dye	2x	0.50	96	10 sec	
Primer Fw or Rv	2 pmol/ µL	1.00	55	5 sec	x 25
H_2O	-	5.25	60	4 min	
PCR amplification product	-	1.50	60	8 min	
			4	10 min	
			-	15 min	

Fw – Forward; **Rv** – Reverse; **min** – Minutes; **sec** - Seconds

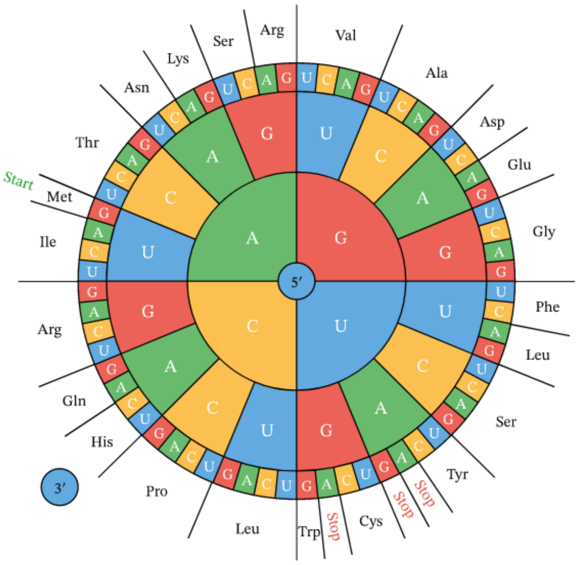


Figure 7.4 - Genetic code. Association between the DNA nucleotide sequence and the respective amino acids in the protein: a set of three nucleotides (codon) corresponds to the initiation codon (start), one amino acid or one termination codon (stop). Taken from [103].

Table 7.8 - Nucleotide code from the International Union of Pure and Applied Chemistry (IUPAC)

Nucleotide Code	Base	Description
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T	T	Thymine
R	A or G	Purine
Y	C or T	Pyrimidine
S	G or C	Amino
W	A or T	Keto
K	G or T	Strong interaction
M	A or C	Weak interaction
B	C or G or T	Not A
D	A or G or T	Not C
H	A or C or T	Not G
V	A or C or G	Not T
N	A or C or T or G	Any base
. or -	None	Gap

Table 7.9 - Amino acids code from the International Union of Pure and Applied Chemistry (IUPAC)

Aminoacid	Nomenclature	Symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table 7.10 - MLPA conditions for the detection of CNV in the α-globin gene cluster

Reaction Mixtures				Reaction Conditions		
Mixture	**Reagents**	**Volume (µL)**	**Volume/tube (µL)**	**Step**	**Temperature (°C)**	**Time**
Mix 0	Rnase A (0.5 mg/mL)	1.0	5.0	RNase Digestion	37	30 min
	DNA (≈75 ng)	1.0		DNA Denaturation	98	10 min
	H_2O	3.0			25	Mix I
Mix I	SALSA Probe Mix	1.5	3.0	Hybridation Reaction	95	1 min
	MLPA Buffer	1.5			60	16 h
Mix II	Ligase-65 Buffer A	3.0	32.0		54	Mix II
	Ligase-65 Buffer B	3.0		Ligation Reaction	54	15 min
	Ligase-65	1.0			98	5 min
	H_2O	25.0			20	Mix III
Mix III	Primer Mix	2.0	10.0	PCR Reaction	95	30 sec
	SALSA Polymerase	0.5			60	30 sec x 35
	H_2O	7.5			72	60 sec
					72	20 min
					15	10 min

min – Minutes; **sec** – Seconds. All the reagents used in Mix I, II and III are from the SALSA® MLPA® probemix P140-C1 HBA kit. Adapted from [40].

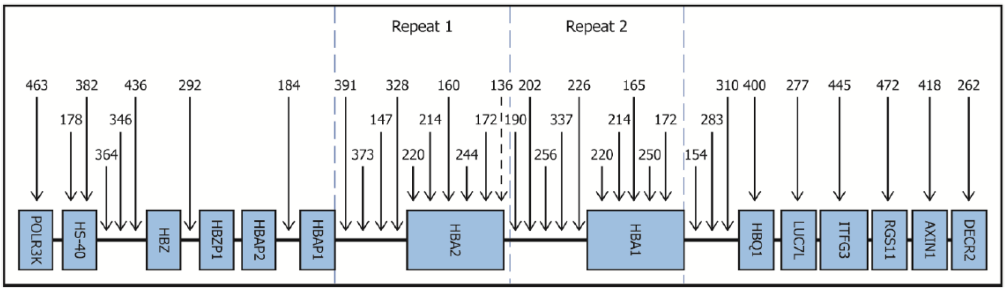


Figure 7.5 - Representation of the location of the probes in the SALSA MLPA probemix P140-C1 HBA. The blue boxes represent the genes in the α-globin gene cluster. The numbers above the arrows represent the amplification size (in nt) of the respective probes. The two regions between the two dashed blue lines present high similarity regarding the probes. This picture is not at scale and it only intends to provide an illustration of the sequential order of the probes. Taken from [40].

Table 7.11 - MLPA probes for the α-globin gene cluster arranged according to their chromosomal location

Length (nt)	**SALSA MLPA probe**	**Gene/Exon**	**Partial sequence (24 nt adjacent to ligation site)**	**Distance to next probe (kb)**
-	-	-	From p-telomere to POLR3K probe	37.2
463	19236-L25316	*POLR3K*	ATATGCTCTGTG-TAAAGTCCTATT	66.4
178	04799-L04797	HS-40	CTGCCCAAGCCA-AGGGTGGAGGCA	0.2
382	04800-L04175	HS-40	GGTACTGCTGAT-TACAACCTCTGG	30.0
364	04926-L23886	9.2 kb before *HBZ*	GGCTGGGGCTCA-AACCAAGGCCCA	5.7
346	04622-L04001	3.5 kb before *HBZ*	CGCAGTGCTAGA-AGGGAGTTCCTG	3.3
436	17214-SP0457-L20489	0.2 kb before *HBZ*	TTCCTCTCCTGT-37 nt spanning oligo-AGTCTAGGAGAG	6.9
292	04624-L04004	Between *HBZ* and *HBZP1*	GTGGAGTAGGCT-TTGTGGGGAACT	7.8
184	04637-L04018	Between *HBAP2* and *HBAP1*	AGTGGCCACAAT-TTGGCAGACAGA	2.5
391	18097-L22521	3.0 kb before *HBA2*	AGGAAGGGGTGA-GAATGAGAGAAA	0.5
373	18090-L08415	2.5 kb before *HBA2*	ATGTCCAGAAGA-AAAGCGGTGACA	1.6
147	18098-L22522	0.9 kb before *HBA2*	GATAAACAAACT-TGGCTCTGGGTA	0.2
328	18092-L22516	0.6 kb before *HBA2*	CCGGGAAGGAAC-AAACACCAGGAC	0.6
220	18099-L22524	*HBA1* e *HBA2*, exon 1	AAGAGTGCCGGG-CCGCGAGCGCGC	0.1
214	18881-L06288	*HBA1* e *HBA2*, exon 1	CGCGCCGACCTT-ACCCCAGGCGGC	0.4
160	08498-L08422	*HBA2*, intron 2	GCGCCTTCCTCT-CAGGGCAGAGGA	0.1
244	04633-L23748	*HBA2*, intron 2	GGGCCTGGGCCG-CACTGACCCTCT	0.1
172	15857-L21812	*HBA1* e *HBA2*, exon 3	GCAGGGGTGAAC-TCGGCGGGGAGG	0.1
136	S0585-SP0043-L09493	Hb Constant Spring mutation	CCAAATACCGTC-28 nt spanning oligo-TGCCCGCTGGGC	0.5
190	18096-L22520	2.5 kb before *HBA1* (0.4 kb after *HBA2*)	TTCTCTCATTCC-CACCCCTTCCTG	0.5
202	18880-L24428	2.0 kb before *HBA1*	GTCACTGCTTTC-CTTCTGGACATG	0.6
256	08494-L08417	1.5 kb before *HBA1*	TTCTCTGCCCAA-GGCAGCTTACCC	0.6
337	14855-L23604	0.9 kb before *HBA1*	CTGGAGCATTCA-ACCTCCTCTGGG	0.3
226	18093-L22517	0.6 kb before *HBA1*	TCCTGGTGTTTA-TTCCTTCCCGGT	0.6
220	18099-L22524	*HBA1* e *HBA2*, exon 1	AAGAGTGCCGGG-CCGCGAGCGCGC	0.1
214	18881-L06288	*HBA1* e *HBA2*, exon 1	CGCGCCGACCTT-ACCCCAGGCGGC	0.4
165	08498-L21607	*HBA1*, intron 2	GCGCCTTCCTCG-CAGGGCAGAGGA	0.1
250	04633-L23600	*HBA1*, intron 2	GGCCCTCGGCCC-CACTGACCCTCT	0.1
172	15857-L21812	*HBA1* e *HBA2*, exon 3	GCAGGGGTGAAC-TCGGCGGGGAGG	0.3
154	08499-L23594	0.1 kb after *HBA1*	TGGGACACACAT-GGCTAGAACCTC	0.3
283	04638-L23602	0.4 kb after *HBA1*	AAGTCCCACTCC-AGCATGGCTGCA	1.9
310	04639-L04020	2.3 kb after *HBA1*	GTTCACTGCCCT-GAAGAAACACCT	1.4
400	19233-L25313	*HBQ1*, exon 3	TGCTCTCTCGAG-GTCAGGACGCGA	25.1
277	15859-L21960	*LUC7L*	ATGTTCCAATGA-AACCAGTGGCAC	33.5
445	17227-L20554	*ITFG3*	GCTGTGATACTT-TTGCCTTTGTCA	31.9
472	18102-L20488	*RGS11*	GAACCTGAGGTT-CCGTGGAATATT	16.3
418	17212-L13393	*AXIN1*	GATCATCGGCAA-AGTGGAGAAGGT	119.4
262	17613-L23601	*DECR2*	CAGAATCGACAT-TCTCATTAACTG	

nt – Nucleotides; **kb** – kilobase. Adapted from [40].

Table 7.12 - Descriptive statistics of the hematological parameters and Shapiro-Wilk's normality test for the population with α-thalassemia

	Mean	SE Mean	Median	Standard Deviation	Minimum	Maximum	Normality Test Shapiro-Wilk (*p*_value)
RBC (x 10^{12}/L)							
Females	5.01	0.07	4.97	0.27	3.62	6.23	**0.188**
Males	5.76	0.10	5.80	0.63	3.78	6.71	0.014
Hb (g/dL)							
Females	11.9	0.2	12.0	1.3	8.5	14.8	**0.733**
Males	13.9	0.2	14.0	1.5	8.6	16.5	0.006
HCT (%)	40.2	0.5	40.2	4.7	29.2	49.3	**0.286**
MCV (fL)	75.8	0.7	75.1	6.8	56.0	98.8	0.010
MCH (pg)	24.0	0.3	23.6	2.7	16.5	34.9	< 0.001
MCHC (g/dL)	31.5	1.0	31.6	1.0	28.8	33.5	**0.173**
RDW (%)	14.9	0.2	14.5	1.8	12.1	23.0	< 0.001

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; **SE** – Standard Error. It is considered a normal distribution when *p*_value ≥ 0.05 (highlighted in bold).

Table 7.13 - Descriptive statistics of the hematological parameters and Shapiro-Wilk's normality test for the population without α-thalassemia

	Mean	SE Mean	Median	Standard Deviation	Minimum	Maximum	Normality Test Shapiro-Wilk (*p*_value)
RBC (x 10^{12}/L)							
Females	4.05	0.06	4.04	0.33	3.53	4.59	**0.339**
Males	4.67	0.08	4.63	0.40	4.07	5.77	**0.923**
Hb (g/dL)							
Females	13.1	0.1	13.0	0.6	12.2	14.1	**0.954**
Males	15.0	0.2	15.2	0.9	13.3	17.3	**0.974**
MCV (fL)	92.6	0.7	93.2	4.6	82.1	102.1	**0.966**
MCH (pg)	32.3	0.3	32.4	1.9	27.9	36.4	**0.962**
RDW (%)	12.9	0.1	12.9	0.7	11.6	14.4	**0.978**

RBC – Red Blood Cells; **Hb** – Hemoglobin; **MCV** – Mean Corpuscular Volume; **MCH** – Mean Corpuscular Hemoglobin; **RDW** – Red Cell Distribution Width; **SE** – Standard Error. It is considered a normal distribution when *p*_value ≥ 0.05 (highlighted in bold).

Table 7.14 - Descriptive statistics of the hematological parameters and Shapiro-Wilk's normality test for the population with the -$\alpha^{3.7}$/αα genotype

	Mean	SE Mean	Median	Standard Deviation	Minimum	Maximum	Normality Test Shapiro-Wilk (*p*_value)
RBC (x 10^{12}/L)							
Females	4.83	0.07	4.80	0.36	3.95	5.77	**0.663**
Males	5.52	0.12	5.64	0.51	4.30	6.36	**0.149**
Hb (g/dL)							
Females	12.7	0.2	12.7	0.8	11.3	14.8	**0.688**
Males	14.5	0.3	14.6	1.4	11.1	16.5	**0.292**
HCT (%)	41.9	0.6	41.5	4,0	34.2	49.3	**0.171**
MCV (fL)	82.0	0.5	82.0	3.7	73.2	98.8	< 0.001
MCH (pg)	26.4	0.3	26.4	2.0	22.4	34.9	< 0.001
MCHC (g/dL)	31.9	0.1	31.8	0.8	30.3	33.5	**0.383**
RDW (%)	14.1	0.2	13.9	1.4	12.1	19.1	< 0.001

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; **SE** – Standard Error. It is considered a normal distribution when *p*_value ≥ 0.05 (highlighted in bold).

Table 7.15 - Descriptive statistics of the hematological parameters and Shapiro-Wilk's normality test for the population with the -α$^{3.7}$/-α$^{3.7}$ genotype

	Mean	SE Mean	Median	Standard Deviation	Minimum	Maximum	Normality Test Shapiro-Wilk (*p*_value)
RBC (x 10^{12}/L)							
Females	5.15	0.10	5.15	0.58	3.62	6.23	**0.601**
Males	5.97	0.14	6.20	0.66	3.78	6.71	0.001
Hb (g/dL)							
Females	11.2	0.2	11.0	1.2	8.5	13.4	**0.720**
Males	13.4	0.3	13.7	1.5	8.6	15.5	0.001
HCT (%)	38.6	0.7	37.7	4.9	29.2	48.9	**0.118**
MCV (fL)	70.7	0.5	71.3	3.9	56,0	77.3	0.012
MCH (pg)	22.1	0.2	22.2	1.4	16.5	24.6	< 0.001
MCHC (g/dL)	31.2	0.2	31.3	1.0	28.8	32.9	**0.163**
RDW (%)	15.6	0.3	15.1	1.8	13.6	23.0	< 0.001

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; **SE** – Standard Error. It is considered a normal distribution when *p*_value ≥ 0.05 (highlighted in bold).

Table 7.16 - Description of the primers used for the sequencing of the *HBA2* and *HBA1* genes

Genes	Primers			
	ID	Sequence	Size (nt)	Tm (°C)
HBA2	A_5'UTR_F1 (Fw)	5' - GGACTCCCCTGCGGTCCAGG - 3'	20	62.0
	C3 (Rv)	5' - CTCCATTGTTGGCACATTCCGGG - 3'	23	58.8
	D71 (Fw)	5' - AGAGGATCACGCGGGTTGC - 3'	19	55.4
HBA1	A_5'UTR_F1 (Fw)	5' - GGACTCCCCTGCGGTCCAGG - 3'	20	62.0
	C113 (Rv)	5' - AAAACTCAGGCACACACAGG - 3'	20	51.8

Fw – Forward; **Rv** – Reverse; **nt** – Nucleotides; **Tm** – Melting Temperature