UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Hereditary Anemia – characterization of the genetic basis and subjacent mechanisms

Bárbara Düemke Coelho Faleiro

Mestrado em Biologia Humana e Ambiente

Dissertação orientada por:
Doutora Paula Faustino, Instituto Nacional de Saúde Dr. Ricardo Jorge
Prof. Doutora Maria Teresa Rebelo, Faculdade de Ciências da Universidade de Lisboa

2020

“The important thing is not to stop questioning. Curiosity has its own reason for existence. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery each day.”

— Albert Einstein

Acknowledgements

First, I would like to thank Dr. Fernando de Almeida, president of the Directive Board of the National Institute of Health Doctor Ricardo Jorge (INSA), Lisboa, Dr. Glória Isidro, coordinator of the Human Genetics Department and Doctor João Lavinha, responsible for the I&D Unit, without their permission this work would not be possible.

To my supervisor, Doctor Paula Faustino that accepted and kindly received me in "Grupo de I&D em Hemoglobinopatias, Metabolismo do Ferro e Patologias Associadas", and to my internal supervisor Prof. Doctor Maria Teresa Rebelo, that although my dissertation project does not fall on her area of expertise, reviewed my work.

This project would not be possible without the subjects with suspicion of IRIDA and some of the IDA cases supplied by medical doctors Dr. Raquel Maia and Dr. Paula Kjöllerstrom, both from the Hematology Unit of Hospital Dona Estefânia, Lisboa, as well as the remaining IDA and hemoglobinopathies cases provided by Dr. Armandina Miranda, from "Unidade Laboratorial de Referência do Departamento de Promoção da Saúde e Prevenção de Doenças não Transmissíveis" of INSA.

I would like to thank the collaboration and technical support of Dr. José Ferrão when carring out the MLPA technique, and all the members of "Unidade de Tecnologia e Inovação" of INSA, with a special thanks to Doctor Luís Vieira, Joana Mendonça and Miguel Machado for all the help during the NGS process.

To the remaining members of the group, Marisa Silva and Joana Ferreira thank you for all the help and company in the laboratory, especially Joana that long after my stay in the lab helped me review the written work, and to Rita Simão that shared with me the workbench and work difficulties during this year. In addition, I would like to thank Sérgio Paulino that, although he was not from my workgroup, was my biggest companion in INSA, thank you for all the help, support and well-needed breaks.

To Prof. Doctor Constança Coelho for the help regarding the statistical methods used in this dissertation.

To Prof. Doctor Francisco Pina Martins, even though he was not my internal supervisor a lot of times acted as one, for that I would like to thank you, for always being willing to help me, regardless of the time of day, and for being tireless when looking for ways to overcome the difficulties I had encountered in this dissertation regarding the bioinformatical and statistical analysis.

To my friends that started this master's degree with me two years ago Beatriz, Mariana, Maria, Donato, and specially Sara, thank you all for your friendship and support.

To Tiago thank you so much for the support and the rides home after a long day of work and the sing a longs and geeky talks that made all the problems disappear for a moment.

To Carolina thank you for always trying to help and shine a bright light in my perspective of how the work was progressing.

To Vítor my other half and personal cheerleader, thank you for your love and unwavering support, you always believe in me and make me believe in myself.

Finally, to my mother and brother for their support, especially you mom you are a force of nature, this thesis is for you. And last but not least, I would like to thank my grandmother “Tó”, although she is not with us anymore, she was the one that ignited my love for science.

This endeavor took more time and energy than I would ever expect, but I finally did it. To all who contributed in any form to my dissertation and helped me along the way, thank you!

This dissertation was conducted in the Department of Human Genetics, I&D Unit, Group of Hemoglobinopathies, Iron Metabolism and Associated Pathologies, at the National Institute of Health Doutor Ricardo Jorge (INSA), IP, Lisbon, Portugal. The work was partially funded by Fundação para a Ciência e a Tecnologia (FCT) ISAMB, and INSA projects 2012DGH720 and 2013DGH910. Additionally, it is a result of the GenomePT project (POCI-01-0145-FEDER-022184), supported by COMPETE 2020 - Operational Programme for Competitiveness and Internationalisation (POCI), Lisboa Portugal Regional Operational Programme (Lisboa2020), Algarve Portugal Regional Operational Programme (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), FCT.

Resumo

A anemia é um problema de saúde pública global que afeta 24,8 % da população mundial, sendo maioritariamente causada por deficiências nutricionais de ferro. Para além disso, quando perturbados os mecanismos moleculares associados à homeostase do ferro é possível que surjam anemias hereditárias de difícil diagnóstico, como é o caso da *Iron-Refractory Iron Deficiency Anemia* (IRIDA). Por outro lado, alterações na síntese ou estrutura das cadeias globínicas originam também anemias hereditárias denominadas hemoglobinopatias.

Primeiramente, de forma a explicar fenótipos atípicos de hemoglobinopatias e entender os mecanismos fisiopatológicos subjacentes aos mesmos, 15 casos clínicos de hemoglobinopatias foram estudados utilizando as metodologias PCR, gap-PCR, *Multiplex Ligation-Dependent Probe Amplification* (MLPA) e sequenciação automática de Sanger para analisar vários genes globínicos (*HBB; HBD; HBG1; HBG2; HBA1*; e *HBA2*). Este estudo permitiu o esclarecimento das causas moleculares associadas aos fenótipos complexos destes indivíduos, nos quais estavam presentes lesões moleculares responsáveis por: variantes estruturais raras de hemoglobina, talassémias do tipo alfa e beta e polimorfismos responsáveis pela modulação da expressão dos genes globínicos

De seguida, o trabalho focou-se no estudo de 10 casos suspeitos de IRIDA, através da amplificação do gene *TMPRSS6*, que codifica a proteína matriptase-2, por PCR-longo e análise por *Next-generation Sequencing*, seguida de análise in silico, de forma a prever o possível efeito patogénico na proteína de variantes genéticas novas/raras. Recorreu-se à ferramenta bioinformática *Human Splicing Finder* de forma a detetar potenciais alterações no processo de *splicing* e às ferramentas *PolyPhen-2*, *SIFT* e *Missense3D*, a fim de analisar o impacto das variantes *missense* ao nível da estrutura e função da proteína codificada. Esta investigação permitiu apenas justificar o fenótipo de um dos casos em estudo, cujo quadro clínico julga-se ser fruto da herança digénica de uma nova mutação patogénica (*TMPRSS6*: c.871G>A; G291S) e 3 polimorfismos modeladores no mesmo gene (K253E, S361S e Y739Y) em conjunto com a co-herança do alelo para a deleção α-talassémica de 3,7 kb, que levam à presença de um fenótipo tipo-IRIDA.

De forma a detetar a presença e entender o papel de polimorfismos comuns em genes relacionados com o metabolismo do ferro, foi realizado um estudo populacional de associação para investigar a influência que os polimorfismos nos genes *TMPRSS6* (V736A e K253E) e *HFE* (H63D e C282Y) têm sobre os parâmetros hematológicos e do status do ferro de 317 indivíduos adultos, portugueses, divididos em 3 grupos: portadores de β-talassémia (BTT), indivíduos com anemia ferropénica (IDA) e controlos normais. Os indivíduos foram genotipados recorrendo às técnicas PCR-(RFLP) *Restriction fragment length polymorphism* e *Amplification-Refractory Mutation System* (ARMS)-PCR. Os resultados sugerem que ambos os genótipos heterozigótico (TC) e homozigótico (CC) da variante V736A exercem um efeito protetor contra a anemia ferropénica e que indivíduos controlo do sexo feminino com o genótipo CC têm os parâmetros volume globular médio (MCV), hemoglobina corpuscular média (MCH) e saturação de transferrina (TSAT) significativamente aumentados quando comparados com o genótipo wild-type (TT). Os resultados da análise de regressão logística detetaram uma associação positiva entre a presença da variante H63D e os níveis de dispersão do volume eritrocitário (RDW) nos controlos saudáveis e verificou-se uma associação entre esta mesma variante e os parâmetros do status do ferro de indivíduos BTT.

Após o estudo de associação, foi testado o desempenho de diagnóstico de 13 índices matemáticos diferenciadores de anemia microcítica em indivíduos BTT ou IDA, com o intuito de determinar quais seriam os mais adequados à população Portuguesa. A análise revelou que para uma população adulta de mulheres portuguesas, os índices com melhor desempenho são RBC, G&K e RDWI. Os resultados foram ao encontro do reportado para outros países Mediterrânicos onde a β-talassémia é endémica. O mesmo verificou-se para o Brasil e França, provavelmente devido, respetivamente, à herança portuguesa e a fluxos migratórios. No geral, os resultados indicam que estes índices podem auxiliar no diagnóstico da causa subjacente à anemia microcítica, reconhecendo indivíduos suspeitos de BTT e encaminhando-os para testes bioquímicos e moleculares adicionais.

Em suma, este trabalho permitiu entender melhor as alterações genéticas responsáveis pelas anemias hereditárias, bem como a forma como estas alteram o metabolismo do ferro, levando a um melhor conhecimento dos fatores que conduzem à variabilidade clínica em pacientes com estas doenças. Concluímos também que o estudo destas patologias tornou possível um correto diagnóstico de algum dos indivíduos estudados, permitindo um adequado acompanhamento clínico e aconselhamento genético, que leva por sua vez a uma melhoria da qualidade de vida dos casos em questão. Ademais, este trabalho reforça a importância do estudo dos marcadores moleculares e do uso de ferramentas não moleculares que complementam as técnicas de diagnóstico já implementadas.

Palavras-chave: Anemias hereditárias; Hemoglobinopatias; IRIDA; *Next-Generation Sequencing*; *TMPRSS6*; *HFE*; Portadores de β-talassémia; Anemia ferropénica; Anemia microcítica; Índices hematológicos.

Abstract

Anemia is a global public health problem affecting 24.8 % of the world population and is mostly caused by nutritional iron deficiencies. Additionally, molecular mechanism´s disturbances associated with the iron homeostasis may cause hereditary anemias of challenging diagnosis, such as Iron-Refractory Iron Deficiency anemia (IRIDA). Alternatively, globin chains structure or synthesis changes may lead to hemoglobinopathies, another type of hereditary anemia.

Firstly, in order to explain atypical hemoglobinopathies' phenotypes and understand underlying pathophysiological mechanisms, 15 hemoglobinopathies clinical cases were studied by PCR, gap-PCR, Multiplex Ligation-Dependent Probe Amplification (MLPA) and automated Sanger sequencing of the various globin genes. These techniques made it possible to clarify the molecular causes behind these complex phenotypes, with molecular lesions being responsible for rare hemoglobin structural variants, β- and α-thalassemias, and the presence of polymorphisms that modulate the expression of globin genes.

Secondly, we focused on ten IRIDA suspects, by amplifying their *TMPRSS6* gene by long-PCR followed by Next-generation Sequencing and *in silico* analysis for predicting the possible pathogenic effect of the detected genetic variants. Through this analysis, only the phenotype of one patient could be justified by the presence of digenic inheritance of the novel damaging mutation (*TMPRSS6*: c.871G>A; G291S) and 3 common modulating SNPs in the same gene (K253E, S361S and Y739Y), in addition to the co-inheritance of the $-\alpha^{3.7kb}$-thalassemia allele, which may add up to an IRIDA-like phenotype.

Thirdly, to detect the presence and understand the roles and of common SNPs in genes related to iron metabolism, a case-control study was performed to investigate the influence of *TMPRSS6* (V736A and K253E) and *HFE* (H63D and C282Y) on the hematological and iron parameters of 317 Portuguese subjects divided into 3 groups: β-thalassemia trait (BTT), iron deficiency anemia (IDA) and controls. Subjects were genotyped using PCR-(RFLP) Restriction fragment length polymorphism and Amplification-Refractory Mutation System (ARMS)-PCR techniques. Results suggests that both heterozygous (TC) and homozygous (CC) genotypes of V736A act as a protective factor against IDA, and that in the female control group, the CC genotype had significantly increased MCV, MCH and TSAT parameters when compared to the wild type (TT). Logistic regression results detected a positive association between H63D variant and RDW in control subjects, and an association between this variant and the iron parameters of BTT subjects.

Finally, we evaluated the diagnostic performance of 13 mathematical indices for differentiating microcytic anemia in BTT and or IDA subjects, to find which ones would better apply to the Portuguese population. When accessing the performance of the indices we found that for a female adult Portuguese population the best performing ones where RBC, G&K and RDWI, similarly to other Mediterranean countries, and also France and Brazil, the latter probably resulting from migratory movements and Portuguese ancestry, respectively. Results imply that these indices can aid diagnosing the condition underlying microcytic anemia, recognizing individuals suspected of BTT and forwarding them to additional biochemical and molecular tests.

The study of hereditary anemia made it possible for us to assign a correct diagnostic to the studied patients, allowing for correct clinical management and genetical counselling, that in turn, improves their quality of life. Additionally, this work has reinforced the importance of the study of molecular markers and use of non-molecular tools complementary to the already implemented diagnostic techniques.

Keywords: Hereditary anemias, Hemoglobinopathies; IRIDA; Next-Generation Sequencing; Genetic variation; *TMPRSS6*; *HFE*; β-thalassemia Trait; Iron deficiency anemia; Microcytic Anemia; Hematological indices.

Initial Statement

This project was carried out in the Human Genetics Department of the National Institute of Health Doctor Ricardo Jorge (INSA), Lisbon.

Some results of this dissertation were presented in scientific meetings in form of scientific posters:

B. D. Faleiro; R. Maia; S. Batalha; J. Mendonça; M. P. Machado; L.Vieira; J. Lavinha and P. Faustino, "The protein Matriptase-2 damaged by a novel missense mutation in the TMPRSS6 gene originates an IRIDA-like phenotype in an African child,", Linderstrøm-Lang Symposium - Protein folding and stability: from molecules to disease, Copenhagen, 2019

B. D. Faleiro; R. Maia; S. Batalha; J. Mendonça; M. P. Machado; L.Vieira; J. Lavinha and P. Faustino, "Novel mutation in addition to functional TMPRSS6 gene polymorphisms originate an IRIDA-like phenotype in an African child,", SPGH 23rd Annual Meeting, Coimbra, 2019

B. D. Faleiro; J. Lavinha and P. Faustino, "Evaluation of mathematical indices as tools for distinguishing β-thalassemia trait from iron deficiency anemia in Portuguese females with microcytic anemia,", SPGH 23rd Annual Meeting, Coimbra, 2019

Abstracts of two posters presented at the Annual Meeting of the Portuguese Society of Human Genetics were published in the scientific journal *Medicine.*

Proceedings of the 23rd Annual Meeting of the Portuguese Society of Human Genetics" in *Medicine,* 2020, vol.99, no. 9, p. 24.

Proceedings of the 23rd Annual Meeting of the Portuguese Society of Human Genetics" in *Medicine,* 2020, vol.99, no. 9, p.30.

The citations and bibliographic references present in this dissertation were done according to the IEEE citation style.

Table of Contents

Acknowledgements I
Resumo III
Abstract V
Initial Statement VI
Table of Contents VII
List of Figures IX
List of Tables XII
Chapter 1 – Introduction 1
1.1 Hemoglobin 1
1.1.1 Erythropoiesis 1
1.1.2 Structure and function 2
1.1.3 Globin gene clusters 3
1.1.4 Ontogeny of hemoglobin 4
1.2 Iron 5
1.2.1 Iron metabolism 5
1.2.2 Iron homeostasis 8
1.3 Anemia 10
1.4 Hereditary anemia by iron deficiency 11
1.4.1 Iron deficiency anemia 11
1.4.2 IRIDA – Iron-Refractory Iron Deficiency Anemia 11
1.4.3 Diagnosis 13
1.4.4 Treatment 15
1.5 Hereditary anemia by changes in the globin chains – Hemoglobinopathies 15
1.5.1 Structural hemoglobinopathies 16
1.5.2 Beta-thalassemia 18
1.5.3 Alfa-thalassemia 21
1.5.4 Failure to switch globin chain synthesis 22
1.5.5 Diagnosis 23
1.5.6 Treatment 25
1.6 Objectives 26
Chapter 2 – Materials and Methods 27
2.1 Population sample 27
2.2 Biological sample 28
2.3 DNA extraction, quantification and quality 29

2.4 Genomic DNA amplification through Polymerase Chain Reaction....29
2.5 Restriction fragment length polymorphism....32
2.6 Multiplex Ligation-dependent Probe Amplification....33
2.7 Automated Sanger Sequencing....34
2.8 Next-Generation Sequencing....35
2.8.1 Library preparation, cluster generation and sequencing....35
2.8.2 Data analysis – variant calling....36
2.9 Variant effect prediction....37
2.9.1 Prediction of splice-affecting variants....37
2.9.2 Prediction of protein changes....37
2.10 Statistical analysis....38
2.10.1 Genetic association study....38
2.10.2 Microcytic anemia differentiation....38
Chapter 3 – Results and Discussion....40
3.1 The basis of hemoglobinopathies....40
3.1.1 Cases 1 and 2: Heterozygous for Hb Leiden....40
3.1.2 Case 3: Heterozygous for Hb Pôrto-Alegre....42
3.1.3 Case 4: Compound heterozygous for Hb S/N-Baltimore....44
3.1.4 Case 5: A sickle cell anemia patient with coinheritance of alpha-thalassemia....45
3.1.5 Case 6: Compound heterozygous for Hb Nigeria + deletion -$\alpha^{3.7kb}$....48
3.1.6 Case 7: Heterozygous for the Cape Verde deletion....49
3.1.7 Case 8: Hb F variability....51
3.1.8 Case 9: Heterozygous for Hb Strasbourg....53
3.1.9 Case 10: Heterozygous for Hb J-Iran....54
3.1.10 Case 11: Heterozygous for Hb E....56
3.1.11 Cases 12, 13, 14 and 15: Segregation in a family of three hemoglobin defects....57
3.2 Cases with suspicion of IRIDA....60
3.2.1 Patients' genetic variants identification and validation....60
3.2.2 *In silico* studies of pathogenicity of found variants....62
3.2.3 Case 6 to 8 and 10 – Unexplained phenotypes....68
3.2.4 Case 9 – IRIDA-like phenotype....68
3.3 Genetic association study....71
3.4 Differentiating microcytic anemia....77
Chapter 4 – Conclusions....82
Bibliographic references....84
Supplementary material....99

List of Figures

Chapter 1

Figure 1.1 – Hemoglobin's structure.. 2
Figure 1.2 – Map of the α-like and β-like globin gene clusters. 3
Figure 1.3 – Normal developmental switches in globin expression.....4
Figure 1.4 – Body iron supply and storage.....6
Figure 1.5 – Absorption of dietary iron by the intestinal enterocyte.....7
Figure 1.6 – Acquisition and use of iron by erythroid precursors.....7
Figure 1.7 – Recycling of the erythrocyte iron by macrophages. 8
Figure 1.8 – Transcriptional regulation of hepcidin expression in hepatocytes......9
Figure 1.9 – Matriptase-2 regulation hepcidin and subsequent iron export.....12
Figure 1.10 – *TMPRSS6* gene structure and matriptase-2 schematic representation.....13
Figure 1.11 – Algorithm for evaluation and treatment of iron deficiency anemia by therapeutic trial. 14
Figure 1.12 – Oxygen dissociation curve of hemoglobin.....17
Figure 1.13 – Spectrum of β-thalassemias according to disease severity and transfusion requirement. ...18
Figure 1.14 – Pathophysiology of β-thalassemia and main genetic mechanisms that contribute to the phenotypic diversity. 19
Figure 1.15 – The global distribution of the β-thalassemia mutations.. 20
Figure 1.16 – HFE protein model.. 20
Figure 1.17 – The different types of α-thalassemia.....21
Figure 1.18 – Diagrammatic representation of common deletions that can lead to α-thalassemia trait. ...22
Figure 1.19 – The global distribution of the α-thalassemias.....22
Figure 1.20 – Ratio of α:β chain synthesis in the different α- and β-thalassemias.....23
Figure 1.21 – Algorithm for diagnosis of hemoglobinopathies in context of carrier screening.....24

Chapter 2

Figure 2.1 – Gap-PCR analysis for diagnosis of $-\alpha^{3.7}$ kb deletion (α-thalassemia). 30
Figure 2.2 – Interpretation of ARMS-PCRs results according to the subject genotype for the TMPRSS6 common variant V736A.....31
Figure 2.3 – Interpretation of multiplex ARMS-PCR results according to the subject genotype for the HFE common variants C282Y and H63D.....32
Figure 2.4 – Interpretation of PCR-RFLP for genotyping the TMPRSS6 common variant K253E. 33

Chapter 3

Figure 3.1 – Biochemical characterization and molecular identification of a hemoglobin variant, Hb Leiden.. 41
Figure 3.2 – Biochemical characterization and molecular identification of hemoglobin variant, Hb Pôrto-Alegre. 43

Figure 3.3 – Hemoglobin Pôrto-Alegre oligomers (T_4). 44
Figure 3.4 – Biochemical characterization and molecular identification of two hemoglobin variants present in compound heterozygoty, Hb S/N-Baltimore. 45
Figure 3.5 – Biochemical characterization of the hemoglobin variant Hb SS. 46
Figure 3.6 – Molecular identification of a sickle cell anemia patient homozygous for the alpha thalassemia -$\alpha^{3.7}$ deletion. 47
Figure 3.7 – Pathophysiology of sickle cell disease 47
Figure 3.8 – Biochemical characterization and molecular identification of a compound heterozygous for Hb Nigeria and alpha thalassemia -$\alpha^{3.7}$ deletion. 48
Figure 3.9 – Biochemical characterization of unknown hemoglobin variant. 49
Figure 3.10 – Molecular identification of β-globin gene cluster Cape Verde deletion. 50
Figure 3.11 – Biochemical characterization of Hb F polymorphisms resulting in anormal chromatographic profiles. 51
Figure 3.12 – Molecular identification of Hb F associated polymorphisms. 52
Figure 3.13 – Biochemical characterization and molecular identification of hemoglobin variant, Hb Strasbourg. 53
Figure 3.14 – Biochemical characterization and molecular identification of hemoglobin variant, Hb J-Iran.. 55
Figure 3.15 – Biochemical characterization and molecular identification of hemoglobin variant, Hb E. .56
Figure 3.16 – Biochemical characterization and molecular identification of three hemoglobin defects in the proposita, HbA2' (HbB2) and α- and β-thalassemia trait simultaneously. 58
Figure 3.17 – Pedigree analysis and genotypes of the family studied. 59
Figure 3.18 – NGS total coverage analysis. 61
Figure 3.19 – Protein and gene and locations of found variants in the coding regions of the *TMPRSS6* gene. 62
Figure 3.20 – Phyre2 protein model 63
Figure 3.21 – PolyPhen-2 pathogenicity prediction of variant V736A on matriptase-2. 64
Figure 3.22 – Conservation profile of matriptase-2 at the position of the variant V736A 64
Figure 3.23 – PolyPhen-2 pathogenicity prediction of variant K253E on matriptase-2. 65
Figure 3.24 – Conservation profile of matriptase-2 at the position of the variant K253E. 65
Figure 3.25 – PolyPhen-2 pathogenicity prediction of variant G291S on matriptase-2. 66
Figure 3.26 – Conservation profile of matriptase-2 at the position of the variant G291S. 66
Figure 3.27 – Structural analysis: comparison between wild-type and mutant GS91S. 67
Figure 3.28 – Smoothed receiver operating characteristics (ROC) curves of the three best performant discriminant indices for discrimination of beta-thalassemia trait and iron deficiency anemia in a female Portuguese population. 80

Supplementary material

Figure S.1 – Informed consent for molecular study of hemoglobinopathies (Part I). 103
Figure S.2 – Informed consent for molecular study of hemoglobinopathies (Part II). 104

Figure S.3 – Informed consent for molecular study of hemochromatosis and other rare genetic diseases associated with disturbances in the iron homeostasis (Part I)..... 105
Figure S.4 – Informed consent for molecular study of hemochromatosis and other rare genetic diseases associated with disturbances in the iron homeostasis (Part II). 106
Figure S.5 – DNA ladders used in agarose gel electrophoresis**.** 117
Figure S.6 – Standard genetic code..... 122

List of Tables

Chapter 1

Table 1.1 – Hemoglobin levels to diagnose anemia at sea level. 13
Table 1.2 – Relative extent of iron stores based on serum ferritin concentration. 14

Chapter 2

Table 2.1 – Common genetic variants in genes *TMPRSS6* and *HFE* researched in case and control subjects of the population study. 28
Table 2.2 – Molecular methodologies applied according to the category of the study approach and gene under study. 29
Table 2.3 – RefSeq sequences used for comparative analysis of Sanger sequencing results. 34
Table 2.4 – Amplicons designed for the NGS study of the *TMPRSS6* gene. 35
Table 2.5 – Discriminant indices used for distinguishing patients with microcytic anemia. 39

Chapter 3

Table 3.1 – Summary of the molecular lesions in the origin of hemoglobinopathies in the studied individuals 40
Table 3.2 – Hematological features for Cases 1 and 2. 40
Table 3.3 – Hematological features for Case 3. 42
Table 3.4 – Hematological features for Case 4. 44
Table 3.5 – Hematological features for Case 5. 46
Table 3.6 – Hematological features for Case 6. 48
Table 3.7 – Hematological and biochemical features for Case7. 49
Table 3.8 – Hematological and biochemical features for Case 8. 51
Table 3.9 – Hematological features for Case 9. 53
Table 3.10 – Hematological and biochemical features for Case 10. 54
Table 3.11 – Hematological features for Case 11. 56
Table 3.12 – Hematological and biochemical features from individuals in the family study (Cases 12 to 15). 57
Table 3.13 – Hematological and iron status features in subjects tested for IRIDA. 60
Table 3.14 – Identification of variants present in the *TMPRSS6* gene of positive control subjects. 61
Table 3.15 – Identification of variants present in the *TMPRSS6* gene of unexplained cases. 68
Table 3.16 – Hematological and iron status features for Case 9 before and after intravenous iron treatment. 69
Table 3.17 – Identification of variants present in the TMPRSS6 gene of Case 9. 69
Table 3.18 – Hematological and biochemical features in subjects tested for some IRIDA causing mutations. 69
Table 3.19 – Demographic, hematological and iron status parameters in control and case populations. .. 71

Table 3.20 – Frequency distribution of TMPRSS6 V736A and K253E and HFE H63D and C282Y gene variants. 73
Table 3.21 – Distribution of V736A, K253E, H63D and C282Y genotype frequencies in IDA patients and controls 74
Table 3.22 – Hematological and biochemical data and comparison of studied groups as well as sex groups. 77
Table 3.23 – Diagnostic accuracy findings for all considered indices used in microcytic anemia female subjects. 79
Table 3.24 – Accuracy findings for calculated new best thresholds of the best performing indices for female microcytic anemia subjects. 80
Table 3.25 – Difference between parameters of β^{+}- and β^{0}-thalassemia trait female subjects. 81

Supplementary material

Table S.1 – Identified mutations in TMPRSS6 gene associated with IRIDA phenotype in literature and ClinVar database. 99
Table S.2 – Hematological and iron status parameters range for a normal adult, and indicators for Iron Deficiency Anemia, IRIDA, α-Thalassemia and β-Thalassemia Trait..... 107
Table S.3 – Hematological and iron status parameters normal range for pediatric subjects. 108
Table S.4 – Conventional PCR conditions for the amplification of the fragments of interest in the globin genes. 110
Table S.5 – Gap-PCR conditions for detection of the alpha-thalassemia deletion -$\alpha^{3.7kb}$. 113
Table S.6 – Long-PCR conditions for the amplification of the gene TMPRSS6. 114
Table S.7 – ARMS-PCR conditions for detection of the variant rs855791 (c.2207T>C - V736A) in the *TMPRSS6* gene. 115
Table S.8 – ARMS-PCR conditions for detection of the variants rs1799945 (c.187C>G - H63D) and rs1800562 (c.845G>A - C282Y) in the *HFE* gene. 116
Table S.9 – Composition of buffer solutions used throughout the project. 117
Table S.10 – Conventional PCR conditions for amplification of *TMPRSS6* gene's Exon 7..... 118
Table S.11 – Reaction mixture for RFLP detection of the variant rs2235324 (c.757A>G - K253E) in gene *TMPRSS6*. 118
Table S.12 – β-globin gene cluster MLPA probes arranged according to chromosomal location. 119
Table S.13 – MLPA assay conditions for detection copy number variations in β-globin gene cluster. .. 120
Table S.14 – Conventional PCR conditions for confirmation of Cape Verde deletion (~7.7 kb) in β-globin gene cluster. 120
Table S.15 – Automated Sanger sequencing reagents and conditions..... 121
Table S.16 – IUPAC code for nucleotides..... 121
Table S.17 – IUPAC code for amino acid. 121
Table S.18 – Conventional PCR conditions for amplification of *TMPRSS6* gene for variant confirmation. 122
Table S.19 – Population study database variable description. 124

Table S.20 – Detected variants in the *TMPRSS6* gene by NGS analysis of 10 subjects with suspicion of IRIDA. 125
Table S.21 – Hematological and biochemical data and comparison of case and control groups as well as sex groups. 127
Table S.22 – Hematological and biochemical data in control subjects, clustered for gender and V736A status. 128
Table S.23 – Hematological and biochemical data in beta-thalassemia trait subjects, clustered for gender and V736A status. 129
Table S.24 – Hematological and biochemical data in iron deficiency anemia subjects, clustered for gender and V736A status. 130
Table S.25 – Hematological and biochemical data in control subjects, clustered for gender and K253E status. 131
Table S.26 – Hematological and biochemical data in beta-thalassemia trait subjects, clustered for gender and K253E status. 132
Table S.27 – Hematological and biochemical data in iron deficiency anemia subjects, clustered for gender and K253E status. 133
Table S.28 – Hematological and biochemical data in control subjects, clustered for gender and H63D status. 134
Table S.29 – Hematological and biochemical data in beta-thalassemia trait subjects, clustered for gender and H63D status. 135
Table S.30 – Hematological and biochemical data in iron deficiency anemia subjects, clustered for gender and H63D status. 136
Table S.31 – Hematological and biochemical data in control subjects, clustered for gender and C282Y status. 137
Table S.32 – Hematological and biochemical data in beta-thalassemia trait subjects, clustered for gender and C282Y status. 138
Table S.33 – Hematological and biochemical data in iron deficiency anemia subjects, clustered for gender and C282Y status. 138
Table S.34 – Association of genetic variants with hematological and iron parameters in control group. 139
Table S.35 – Association of genetic variants with hematological and iron parameters in beta-thalassemia group 140
Table S.36 – Diagnostic accuracy findings for all considered indices used in microcytic anemia subjects. 141
Table S.37 – Diagnostic accuracy findings for all considered indices used in microcytic anemia male subjects. 142
Table S.38 – Accuracy findings for all calculated new best thresholds. 146
Table S.39 – Accuracy findings for all calculated new best thresholds for male microcytic anemia subjects. 147
Table S.40 – Accuracy findings for all calculated new best thresholds for female microcytic anemia subjects. 147

List of abbreviations and symbols

A	Adenine
aa	Amino acids
AD	Allelic depth reads
ARMS	Amplification refractory mutation system
AUC	Area Under the Curve
bp	Base pair
BM	Bone marrow
BMP	Bone morphogenetic protein
BMPR	Bone morphogenetic protein receptor
BPG	2,3-Bisphosphoglyceric acid
BT	Beta-thalassemia
BTT	Beta-thalassemia trait
BTI	Beta-thalassemia intermedia
BTM	Beta-thalassemia major
C	Cytosine
CBC	Complete blood count
CE	Capillary electrophoresis
CI	Confidence interval
CNV	Copy number variation
CO	Carbon monoxide
CO_2	Carbon dioxide
CP	Ceruloplasmin
CUB	C1r/C1s, urchin embryonic growth factor and BMP1domain
Da	Dalton
DCYTB	Duodenal cytochrome B
ddNTPs	2',3'-dideoxynucleotides
dL	Deciliter
DMT1	Divalent metal transporter 1
DNA	Deoxyribonucleic acid
dNTPs	2'-deoxynucleotide triphosphates
DP	Read depth
DPG	2,3-diphosphoglycerate
EDTA	Ethylenediamine tetra-acetic acid
EPO	Erythropoietin
ERFE	Erythroferrone
EtBr	Ethidium bromide
FDR	False discovery rate
Fe	Iron
Fe^{2+}	Ferrous iron
Fe^{3+}	Ferric iron
Fe_2Tf	Diferric transferin
fL	Femtolitre
FPN	Ferroportin
Ft	Ferritin
G	Guanine
g	gram
GAS	Genetic association studies

Gb	Gigabyte
gDNA	Genomic DNA
GIT	Gastrointestinal track
GQ	Genotype quality
GVCF	Genomic variant call format
H	Heavy
HAMP	Gene encoding hepcidin
Hb	Hemoglobin
Hb A	Adult hemoglobin
Hb A1c	Glycated hemoglobin
Hb A2	Adult hemoglobin normal variant
HBA1	Gene encoding α-globin
HBA2	Gene encoding α-globin
HBB	Gene encoding β-globin
HBD	Gene encoding δ-globin
HBE1	Gene encoding ε-globin
Hb F	Fetal hemoglobin
HBG1	Gene encoding ^{A}γ-globin
HBG2	Gene encoding ^{G}γ-globin
HBZ	Gene encoding ζ-globin
HEPH	Hephaestin
HFE	Gene encoding homeostatic iron regulator
HH	Hereditary hemochromatosis
HIF	Hypoxia inducible factor
HLPC	High performance liquid chromatography
H_2O	Water
H_2O_2	Hydrogen peroxide
HO-1	Inducible heme oxygenase 1
HP	Hemoglobin-haptoglobin complex
HPX	Heme-hemopexin complex
HS	DNase I hypersensitive site
Ht	Hematocrit
HWE	Hardy–Weinberg equilibrium
IAC	Internal amplification control
ID	Iron deficiency
IDA	Iron deficiency anemia
IL-6	Interleukin-6
INDEL	Insertions or deletions
IRIDA	Iron-refractory iron deficiency anemia
IRP	Iron regulatory protein
ISC	Iron-sulfur cluster
IUPAC	International Union of Pure and Applied Chemistry
IVS	Intervening sequence
JAK-STAT	Janus-associated kinase–signal transducers and activators of transcription
k	Kilo
kb	Kilo base
kDa	Kilo Dalton
L	Light
LCR	Locus control region

LDLR	Low-density lipoproteins receptor domain
LIC	Liver iron concentration
MCV	Mean corpuscular volume
MCH	Mean cell hemoglobin
MCHC	Mean corpuscular hemoglobin concentration
MED	Mediterranean
Mfrn1	Mitoferrin 1
mg	Milligram
MLPA	Multiplex ligation-dependent probe amplification
mRNA	Messenger ribonucleic acid
MT2	Matriptase-2
ng	Nanogram
NGS	Next-generation Sequencing
NO	Nitric oxide
NPV	Negative predictive values
NTDT	Non-transfusion dependent
O_2	Oxygen
$O_2^{-\bullet}$	Superoxide radical
$OH^{\bullet}$	Hydroxyl radical
OMIM	Online Mendelian Inheritance in Man
OR	Odds ratio
p	*p*-value
PCBP1	Poly (rC)-binding protein 1
PCR	Polymerase chain reaction
pH	Potential for hydrogen
pCO_2	Partial pressure of carbon dioxide
pO_2	Partial pressure of oxygen
PPV	Positive predictive values
QC	Quality control
QoL	Quality of life
QUAL	Variant quality
RBC	Red blood cells
RDW	Red cell distribution width
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
ROC	Receiver operating characteristic
ROS	Reactive oxygen species
RP-HPLC	Reverse phase high performance liquid chromatography
RR	Risk ratio
SCD	Sickle cell disease
SCT	Sickle cell trait
SEA	Sperm protein, enterokinase and agrin domain
SENS	Sensitivity
-SH	Thiol group
SI	Serum iron
SNP	Single-nucleotide polymorphism
SP	Serine protease
SPEC	Specificity
S-S	Disulfide bridges

STEAP3	Six-transmembrane epithelial antigen of the prostate 3
T	Thymine
TBE	Tris-Borate-EDTA
TBIC	Total binding iron capacity
TDT	Transfusion dependent
Tf	Transferrin
TfR1	Transferrin receptor 1
TfR2	Transferrin receptor 2
TM	Transmembrane region
TMPRSS6	Transmembrane serine protease 6 or matriptase-2
TMPRSS6	Gene encoding matriptase-2
TTSP	Type II transmembrane serine protease family
TSAT	Transferrin saturation
U	Uracil
UV	Ultra-violet
WHO	World Health Organization
WT	Wild type
YI	Youden´s index
α	Alpha
β	Beta
γ	Gamma
δ	Delta
ε	Epsilon
ζ	Zeta
θ	Theta
μ	Mu
μg	Micro gram
χ^2	Qui-square
Ψ	Psi
3D	Three-dimensional

Chapter 1 – Introduction

The field of medical genetics has been revolutionized by the fast identification of genes associated with human disease. Together with an improved understanding of the molecular etiology of genetic disorders, this has led to more accurate diagnostic, prognostic and potential therapeutic tools. Although many genetic diseases are multifactorial, classically the model used for the discovery of single gene disorders is based on the assumption that the spread of a trait in families is based on the transmission of a single molecular defect [1].

Single gene disorders are remarkably diverse affecting structure and/or function, and consequentially generating a wide spectrum of disease severity [2], [3]. The identification of faulty genes underlying inherited diseases has determined that patients with the same genotype can have different patterns of clinical expression [3]. Despite their clinical diversity, these disorders have a common biological basis, potentially being passed on to offspring, requiring basic genetic and management services including accurate diagnosis; risk assessment; information for the patient and their family; and access to options for managing risk and services for affected children. Monogenic disorders are divided into 'rare single gene disorders' and 'common single gene disorders', the first having in its majority diseases with a prevalence of <1/1 000 000, with an European birth prevalence definition of 5 per 10 000 [2], [4].

Following the discovery of a disease gene, a natural first step is to investigate the mutational spectrum in large patient cohorts. These results are needed to correlate specific mutations with various phenotypic aspects, like severity and age of onset. One is often unable to draw conclusions from such studies, leading to the expansion of the monogenic model of disease transmission to account for other factors [1]. Sometimes associating phenotype to genotype is complicated due to the complex interaction of the environment with the different allelic variants and interaction with other genetic risk factors involved at the secondary and tertiary levels. Evidence for these modifier genes comes from the range of phenotypes within families sharing the same genotype [3].

In this context, both classical and novel genomic techniques have a large impact on the study of monogenic diseases, as is the case of the hereditary anemia investigated in this dissertation. Most importantly, the correct diagnosis of these disorders and its consequent understanding of the molecular and cellular pathophysiology and genetic epidemiology will help clinicians manage these patients improving their quality of life (QoL) [5], [6].

1.1 Hemoglobin

1.1.1 Erythropoiesis

Erythropoiesis is the regulated process from which derive mature erythrocytes, also called red blood cells (RBC). In this process, when provided the proper nutrients such as folate, cobalamin (vitamin B12) and iron (Fe), stem cells present in the bone marrow (BM) are stimulated by a hormone called erythropoietin (EPO), produced by kidney cells. Stimulus from this hormone and other factors result in the proliferation and differentiation of these cells, in a pathway that will ultimately end with the extrusion of the cells' nucleus and the accumulation of high concentrations of hemoglobin and some residual ribonucleic acid (RNA) in their cytoplasm. [7]. These immature cells called reticulocytes enter circulation and suffer a 24-hour remodeling process, becoming mature erythrocytes once the remaining globin chain mRNA is translated. A mature RCB will circulate for 100-120 days until oxidative damage leads to its removal from circulation, followed by senescence and destruction by macrophages in the spleen [8].

Humans complete the synthesis of approximately 2 million erythrocytes per second, with each mature cell containing about 280 million molecules of hemoglobin [9]. Mainly because of hemoglobin's sheer quantitative predominance in the erythrocyte's cytoplasm, it dominates the pathophysiology of several RBC disorders and modulates others for which is not directly responsible [8].

1.1.2 Structure and function

Hemoglobin (Hb) is a metalloprotein contained within the RBCs. This molecule is an essential part of the vertebrate's circulatory system, with its main function being the transport of oxygen (O_2) from the lungs to the tissues [10]. Other functions of this protein include: the transport of three other gases, carbon dioxide (CO_2), carbon monoxide (CO), and nitric oxide (NO), which also have significant biological roles, and buffering action inside the erythrocytes. Hemoglobin is able to have these functional properties due to its structure (Figure 1.1), in particular the characteristic folds of the amino acid chains [11], [12].

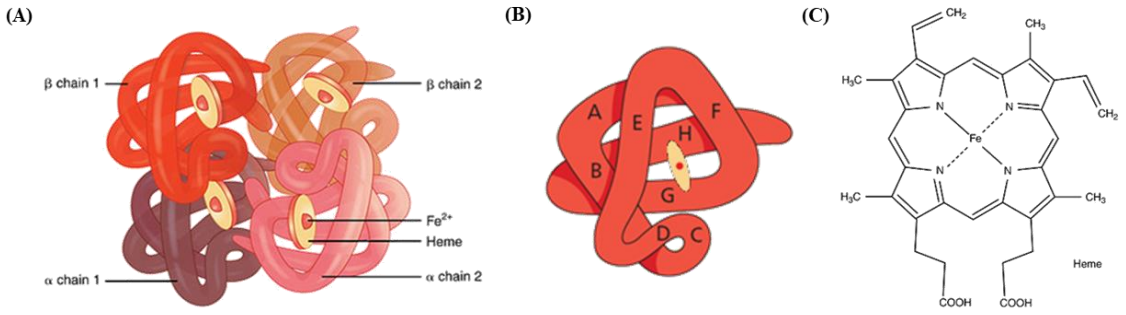


Figure 1.1 – Hemoglobin's structure. (A) Hemoglobin tetramer composed of two identical α-like chains and two identical β-like chains with each chain possessing one heme group. **(B)** Representation of the tertiary structure of hemoglobin monomer (globin chain), containing eight helical regions (A to H). **(C)** Heme structure composed of a protoporphyrin IX molecule complexed with a ferrous iron cation Fe^{2+}. Adapted from Betts *et al.* [13] and Bain [12].

The normal adult hemoglobin (Hb A) is a 64.5 kDa tetrameric protein comprised of two symmetric dimers of alpha (α) and beta (β) globin polypeptide chains (α2β2), of 141 and 146 amino acids (aa) respectively, each accommodating a heme group (Figure 1.1 A) [10]. The globin chain units adopt a three-dimensional (3D) structure called globin fold containing eight helical regions A to H (Figure 1.1 B), this produces an internal hydrophobic environment that protects the iron of heme from oxidation, and an external hydrophilic environment that renders solubility to the molecule [12]. The heme group is composed of a protoporphyrin IX molecule complexed with a ferrous iron cation (Fe^{2+}) (Figure 1.1 C), this hemoglobin component is essential for oxygen transport since it is the one that reversibly binds O_2 via the iron cation [14]. This group biosynthesis occurs in all metabolically active cells that contain mitochondria, having a bigger role in erythropoietic tissue, where it is required for hemoglobin synthesis, and in hepatic tissue, where it forms the basis of various heme-containing enzymes [15].

Hemoglobin is an allosteric protein, this means that when one heme group in hemoglobin binds oxygen, the conformational change in the protein entails cooperation from the three remaining heme sites, which will successively bind to oxygen molecules as well [16]. Areas where the globin chains contact with each other or to heme are highly conserved due to their functional importance [17].

1.1.3 Globin gene clusters

Hemoglobin synthesis occurs during erythropoiesis and is controlled by two multigene clusters on chromosomes 16 (16p13.3) and 11 (11p15.5) containing the α-like and β-like globin genes, respectively (Figure 1.2) [6]. In these clusters, the genes and non-expressed pseudogenes are arranged along the chromosomes presenting an expression pattern that shifts from genes more to 5' to the ones more to 3', during different stages of the development [11].

The α-globin gene cluster is of approximately 30 kilobases (kb) of deoxyribonucleic acid (DNA) and is composed of: the *HBZ* gene, zeta (ζ)-globin; the pseudogene *HBZP* (Ψζ); the *HBM* gene, mu (μ)-globin, originally considered a pseudogene but currently thought to be a protein-coding gene; two pseudogenes alpha *HBAP2* and *HBAP1* (Ψα2 and Ψα1), two identical alpha genes *HBA2* and *HBA1* (α2 and α1), which will ultimately result in a single α-globin peptide; and the *HBQ1* gene, theta (θ), of unknown function since is unclear whether the translation of this gene is able to participate in the formation of a functional hemoglobin tetramer [8].

The β-globin gene cluster is of approximately 70 kb of DNA and is composed by: the *HBE1* genes, epsilon (ε)-globin, a gene only expressed in the embryo; the duplicated gamma-globin genes *HBG2* (^{G}γ) and *HBG1* (^{A}γ), that code for a dominant β-like globin during the fetal life, these two globin chains differ from glycine to alanine at position 136 of the γ chain; the pseudogene *HBBP1*, beta (Ψβ1); and the *HBD*, delta (δ), and *HBB*, beta, responsible for globin chains present in the adult [8], [12].

The expression of globin genes is controlled by upstream regulatory regions. In the β-globin cluster, there are five DNase I hypersensitive sites (HS), HS1 to HS5, that together are referred to as the β-globin locus control region (β-LCR). Similarly, there are four regulatory elements in the α cluster, with HS-40 appearing to have a dominant role. Studies suggest that the molecular switches between globin genes are a consequence of the competition between the globin promoters for access to their activating upstream regulatory elements [6].

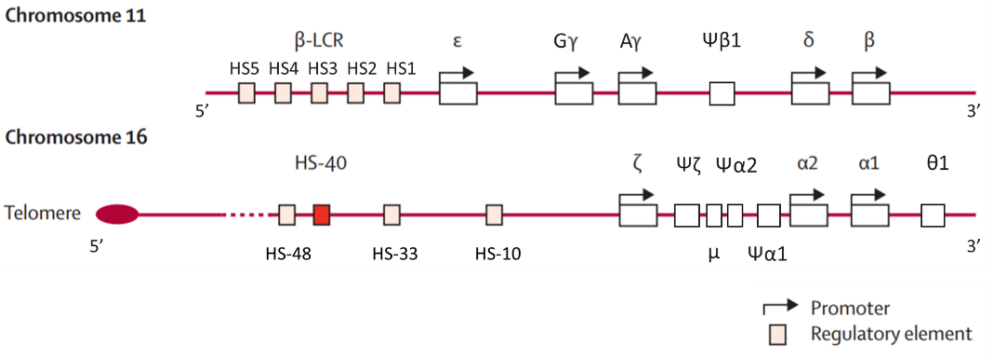


Figure 1.2 – Map of the α-like and β-like globin gene clusters. The structure of the α-globin gene cluster (5'-ζ-Ψζ-μ-Ψα2-Ψα1-α2-α1-θ1-3') and β-globin gene cluster (5'-ε-^{G}γ-^{A}γ-Ψβ-δ-β-3') are shown, as well as the promoter regions that bind transcription factors. Enhancers and related regulatory elements also bind transcription factors and interact with promoters to increase the levels of RNA transcribed. **HS** - Major upstream regulatory element (DNase I hypersensitive site). **β-LCR** - β-locus control region. Adapted from Higgs *et al.* [6].

Globin genes share the same basic organizational features being all composed of three exons separated by two introns, also named intervening sequence (IVS), and have 5' cap structure and 3' poly(A) tail. Their organization differs only in the relative size of the introns since both α-genes introns are small with 100-300 base pairs (bp) and β-genes have one small and one large intron with 1000-1200 bp. In these genes, the second exon is responsible for encoding the protein sites responsible for the heme-binding and the third for the globin chain contact points [8].

1.1.4 Ontogeny of hemoglobin

As previously stated, the hemoglobin composition of the erythrocyte changes during the development according to the sites of erythropoiesis that differ in gestation and postnatal development (Figure 1.3 A). The switching, sequential activation and inactivation, of α- and β-like globin genes produce different hemoglobin tetramers: three distinct embryonic (Hb Gower-I [ζ2ε2], Hb Gower-II [α2ε2], and Hb Portland [ζ2γ2]), expressed in the yolk-sac erythroblasts; fetal (Hb F – α2γ2); and two adults (Hb A – α2β2 and Hb A2 – α2δ2), (Figure 1.3 B) [6]. Unlike the other globin chains, α-globin is present throughout all development stages, with its gene expression starting early in the first trimester and being sustained for life [8].

Hb F is the main hemoglobin present during the intrauterine life, with γ-globin genes expression beginning early in embryogenesis. This hemoglobin synthesis starts a rapid decline just before birth, persisting until 9 months of age when it reaches a level of 1 % of the blood's total hemoglobin (Figure 1.3 B) [8], [17]. During intrauterine life and at birth, Hb F presents a $^{G}\gamma:^{A}\gamma$ ratio of 2:1 to 3:1, a few months after birth this ratio changes to approximately 2:3 [12]. When compared to the adult hemoglobin, Hb F is characterized by a higher oxygen affinity due to a reduced sensitivity to inhibitory effects of the organic phosphate, 2,3-diphosphoglycerate (DPG), a metabolite of red cell glycolysis [18].

In the postnatal life, the predominant hemoglobin is Hb A, with normal adult RBC containing approximately 97.0 %. The remaining hemoglobin present in the erythrocytes is Hb A2, produced in small amounts representing about 2.0 % of the total hemoglobin (Figure 1.3 B) [19]. For this reason defects in the δ-chains that compose Hb A2 are normally of no clinical consequence [8].

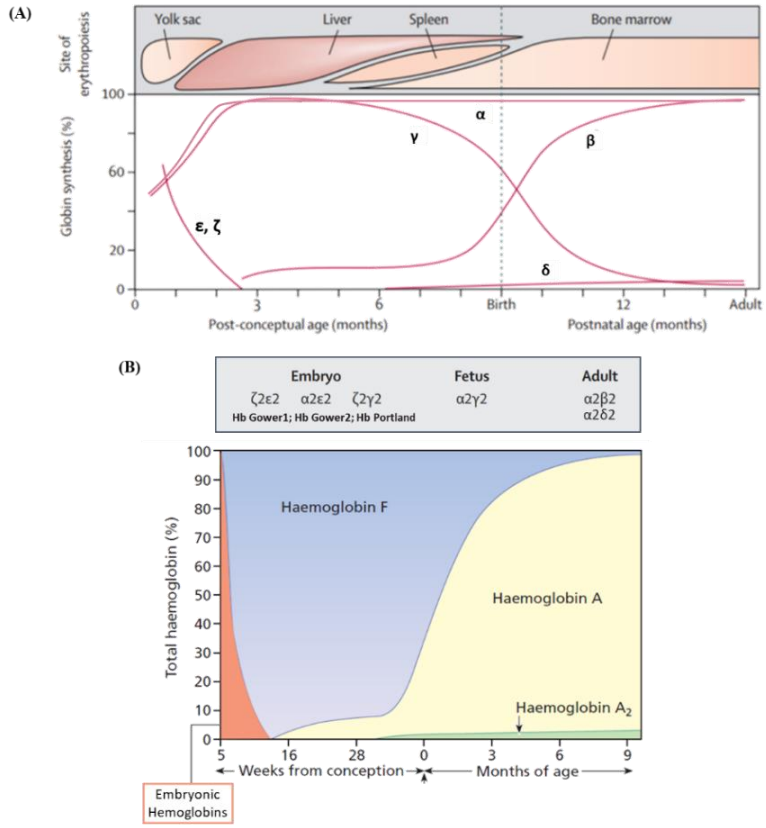


Figure 1.3 – Normal developmental switches in globin expression. (A) The sites of erythropoiesis at different stages of development and the levels of expression of the embryonic, fetal, and adult globin chains at various gestational ages are shown. **(B)** Types and composition of hemoglobin produced at each developmental stage as well as the average percentages of various hemoglobins present during the embryonic and fetal periods and infancy. Adapted from Bain [12] and Higgs *et al.* [6].

The transcription rates of the α- and non–α-globin genes are not equal, there is an excess of α-globin messenger (m)RNA present in the RBC. However, this is compensated by a somewhat more efficient β-globin mRNA translation, leading to counterbalancing forces that result in identical syntheses of α- and β-globin chains [8]. The pattern of hemoglobin synthesis during development clarifies why changes in the α-chains cause clinical problems from early fetal life and why changes in the β-chains are of difficult diagnosis in the neonatal period [19].

1.2 Iron

Iron is an essential element required by most organisms, due to its ability to act both as an electron donor and acceptor by readily alternating between ferric (Fe^{3+}) and ferrous (Fe^{2+}) forms [20]. For this reason, in proteins, iron serves as a prosthetic group, such as in heme and iron-sulfur (Fe-S) clusters (ISCs), which are involved in crucial cellular processes, including respiration, oxygen transport (hemoglobin) and storage (myoglobin), DNA synthesis, energy production, and host defense [21].

While essential for life, the ease with which iron can gain and lose electrons renders it the ability, when improperly sequestered, to catalyze the formation of highly reactive oxygen species (ROS). This happens through Fenton and Haber-Weiss-type reactions (Equations 1.1 to 1.3) [22], creating, for example, hydroxyl ($OH^{\bullet}$) free radicals yielded from superoxide ($O_2^{-\bullet}$) and hydrogen peroxide (H_2O_2), which in excess can damage lipids, protein and DNA, resulting in cellular dysfunction, apoptosis, and necrosis [20], [23].

(Fenton) $$Fe^{3+} + O_2^{\bullet -} \rightarrow Fe^{2+} + O_2 \quad (1.1)$$

(Fenton) $$Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^- + OH^{\bullet} \quad (1.2)$$

(Haber-Weiss) $$O_2^- + H_2O_2 \rightarrow O_2 + OH^{\bullet} + OH^- \quad (1.3)$$

Iron's bioavailability is limited since under aerobic conditions Fe^{2+} is readily oxidized in solution to Fe^{3+}, which is theoretically insoluble at the physiological potential for hydrogen (pH) of 7.4 [24]. Therefore, both the total body iron and the quantity within each cell must be elaborately controlled to ensure adequate iron availability while avoiding excess iron toxicity and limiting iron access to invading pathogens [20], [25]. On one hand, within systemic circulation cellular requirements are met by, the monomeric transport glycoprotein, transferrin (Tf), the physiologic carrier of iron through the plasma and extracellular fluid, which keeps iron soluble and nontoxic, unable to engage in Fenton/Haber–Weiss reactions [20], [24]. On the other, the iron within the cell is either used or sequestered with cytosolic ferritin (Ft), an iron storage protein, comprising of both heavy (H) chains, with ferroxidase activity, and light (L) chains, storing up to 4,500 iron atoms in a shell-like structure, whilst ready for mobilization [20], [21].

Humans are unable to excrete iron and daily losses are minor, representing less than 0.05 % of the total body iron. Therefore, iron metabolism must be tightly regulated, with iron homeostasis firmly controlling iron absorption and performing efficient iron recycling [20]. Disorders in iron homeostasis span from iron excess to iron deficiency, as well as unequal distribution, when some tissues are iron-loaded and others iron-deficient [21].

1.2.1 Iron metabolism

The iron traffic in the body works like a circular economy (Figure 1.4) [21]. Out of the 20-25 mg/daily required by an adult, just 1-2 mg come from dietary sources and are absorbed by the duodenal enterocytes, compensating for obligatory daily losses. These losses occur predominantly

through desquamation of epithelial cells in the intestine and the skin, and through minor bleeding, with women suffering from additionally iron losses through menstruation [26]. The remaining needs are satisfied by iron recycling at the macrophage level upon phagocytosis of senescent erythrocytes, moreover, iron from damaged muscles is locally recycled to regenerating fibers. The excess iron is stored in ferritin the liver hepatocytes and macrophages as a reserve [21].

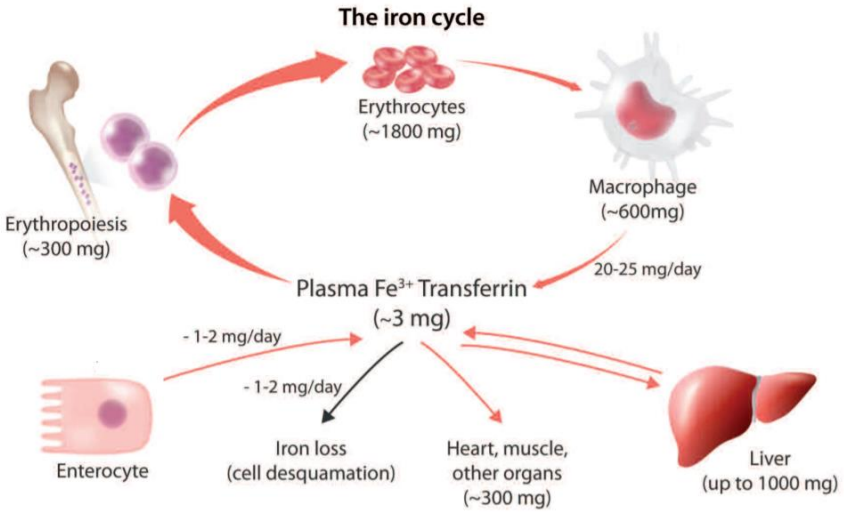


Figure 1.4 – Body iron supply and storage. Iron circulates bound to transferrin to be released to all organs/tissues. Most iron (20-25 mg) recycled by macrophages, which phagocytize senescent red blood cells, is supplied to the bone marrow for erythropoiesis. The daily uptake of dietary iron by duodenal enterocytes is 1-2 mg; the same amount is lost through cell desquamation and blood loss. Excess iron is stored in the liver hepatocytes and macrophages as a reserve. Arrows indicate directions. Numbers (in mg) are a mean estimate. Adapted from Camaschella *et al.* [21].

Intestinal iron absorption: Dietary iron comes in two forms, heme and non-heme iron, and is absorbed by duodenal enterocytes (Figure 1.5). Most of the non-heme dietary iron presented to the organism on its ferric form. Before absorption, Fe^{3+} is converted to Fe^{2+} by the reducing action of other dietary constituents, such as ascorbic acid (vitamin C) at the stomach's low pH, or by the action of membrane-associated ferrireductases, at the enterocyte's apical membrane, facing the gut lumen, such as duodenal cytochrome B (DCYTB). The ferrous iron is then absorbed through the proton-coupled divalent metal transporter 1 (DMT1), also responsible for the intake of other metals ionic form, are examples cobalt, zinc and cadmium [20], [27].

At the same time, heme iron, which originates from meat, as a component of the hemoproteins, hemoglobin and myoglobin, is absorbed by an independent mechanism through receptor-mediated endocytosis that is still not understood [27]. Once heme iron is released intracellularly it is broken down by the inducible heme oxygenase 1 (HO-1) and iron is released in its ferric state. Afterwards, cytosolic iron can either then be retained for cellular requirements or stored in cytosolic Ft, mediated by the cytosolic iron chaperone poly (rC)-binding protein 1 (PCBP1), or be exported to circulating Tf by the basolateral iron exporter ferroportin (FPN), which requires oxidation by membrane-bound hephaestin (HEPH), [20], [28].

Iron absorption is modulated by a liver expressed peptide, hepcidin, or locally modulated through transcriptional (hypoxia inducible factor [HIF]-2α) and posttranscriptional (iron regulatory protein/ iron-responsive element system [IRP1/2]) mechanisms [20].

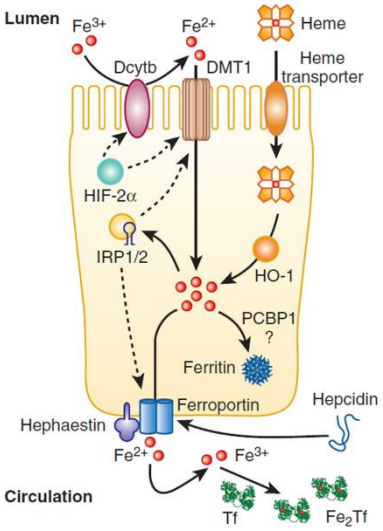


Figure 1.5 – Absorption of dietary iron by the intestinal enterocyte. The metal transporter (DMT1) takes up ferrous iron, reduced by duodenal cytochrome B (DCYTB), on the luminal side of the enterocyte. Unused iron inside the cell is either stored in ferritin (Ft) or exported to circulating transferrin (Tf) by ferroportin (FPN) after ferrous iron is oxidized to ferric iron by hephaestin (HEPH). Hypoxia inducible factor (HIF)-2α stimulates the expression of the apical (DMT1) and basolateral (FPN) transporters Heme, after entering the cell through an unknown mechanism, is converted to iron by heme oxygenase (HO-1). Adapted from Brittenham [20] and Anderson *et al.* [29].

Iron utilization for erythropoiesis: Once passed the digestive barrier, iron reaches the blood while being transferrin-bound, moving in its majority towards the bone marrow for erythropoiesis (Figure 1.6) [30]. There, diferric transferrin (Fe_2Tf) enters the erythroblasts via transferrin receptors on the cell surface beginning the Tf cycle. Two different isoforms of the transferrin receptor exist, that bind to mono and diferric Tf, and are encoded by two separate genes. One is transferrin receptor 1 (*TfR1*) expressed on all iron-requiring cells that function as the physiologic transferrin iron importer, the other is transferrin receptor 2 (*TfR2*), a homolog of *TfR1*, which is expressed only in hepatocytes [20]. Fe_2Tf binding to TfR1 is followed by internalization of the complex by clathrin-mediated endocytosis. On the endosome, ferric iron is released from Tf and reduced by the metalloredutase six-transmembrane epithelial antigen of the prostate 3 (STEAP3), so it can be exported into the cytosol by DMT1. Cytosolic iron is used either for the formation of iron-containing proteins or biosynthesis of ISCs and heme in the mitochondria, where Fe^{2+} is imported by the inner membrane protein mitoferrin 1 (Mfrn1) [28], [31].

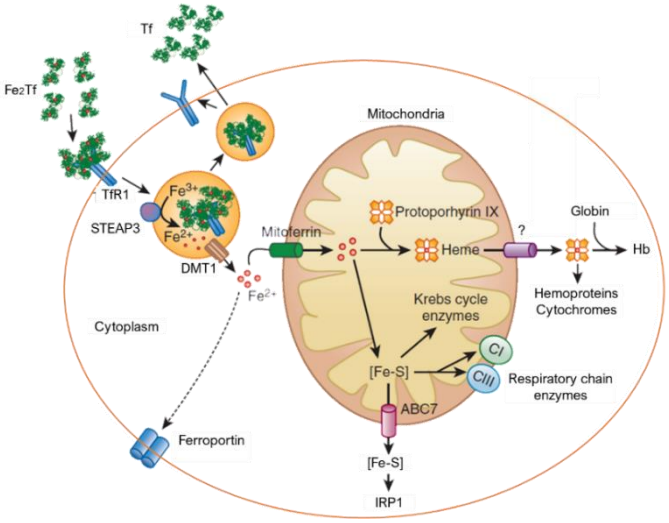


Figure 1.6 – Acquisition and use of iron by erythroid precursors. Iron is imported in the transferrin (Tf) cycle and principally used for the synthesis of heme. **DMT1** – divalent metal transporter 1; **Fe2Tf** – diferric transferrin; **Hb** – hemoglobin; **IRP1** – iron regulatory protein 1; **STEAP3** – six-transmembrane epithelial antigen of the prostate 3; **TFR1** – transferrin receptor 1. Adapted from Brittenham [20].

Iron recycling: Splenic and hepatic macrophages have to bear most of the burden of maintaining adequate levels of plasma iron, by scavenging for senescent RBC, so that the iron present in the hemoglobin becomes available again. Additionally, they also retrieve small amounts of iron from hemoglobin-haptoglobin (HP) and heme-hemopexin (HPX) complexes, by binding them to the CD163 and CD91 scavenging receptors, respectively (Figure 1.7) [20], [27]. These complexes are too large to be filtered by the kidneys, an aspect that helps limit the renal loss of iron. Inside the macrophage, the heme group is catabolized by HO-1, and the subsequent iron is exported through FPN and oxidized by ceruloplasmin (CP). In the absence of iron deficiency, some of the excess iron is retained as Ft [20].

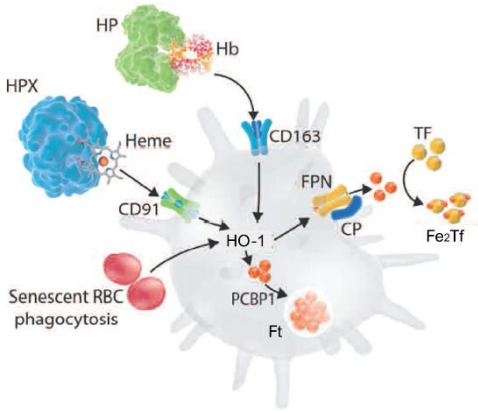


Figure 1.7 – Recycling of the erythrocyte iron by macrophages. Macrophages recover iron from phagocytized red blood cells (RBC) after heme is degraded by heme oxygenase. They also recover heme from hemoglobin (Hb), haptoglobin (HP) or heme-hemopexin (HPX) complexes. Iron not used inside the cells is either stored in ferritin (Ft) or exported to the circulation by ferroportin (FPN) with the cooperation of ceruloplasmin (CP). Adapted from Camaschella *et al.* [21].

Iron storage: By keeping iron storage, the body forestalls the presence of free iron whenever iron levels increase while securing its immediate availability during iron deficiency [27]. The liver is both the biggest iron storage organ and the central site for control of systemic iron homeostasis, by regulating the molecular pathways necessary for iron balance [20], [26]. Hepatocytes can acquire iron from plasma Tf via the transferrin cycle, from HP and HPX complexes via endocytosis, like macrophages. Notably, when the rate of iron influx into plasma exceeds the rate of iron acquisition by transferrin, non-transferrin-bound plasma iron can enter hepatocytes and other specific cells, independently of the Tf mechanism [20].

1.2.2 Iron homeostasis

The liver controls systemic iron homeostasis by producing the hepatic hormone hepcidin. Hepcidin is a 25-amino acid peptide that is predominantly, but not exclusively, secreted by the liver that circulates in the bloodstream and is excreted by the kidneys [32]. This hormone controls iron export to the plasma by binding to FPN, inducing its internalization and lysosomal degradation in enterocytes, macrophages and hepatocytes [21]. Loss of FPN prevents intestinal iron uptake and release of iron from intracellular stores of iron-recycling macrophages, resulting in decreased serum iron [33]. Also, at the same time, hepcidin was identified for its role as an antimicrobial [34].

Hepcidin is encoded by the *HAMP* gene located at chromosome 19 (19q13.1) and results from the proteolytic processing of an 84-amino acid prepropeptide, comprised of a signal peptide, a pro-region and the fully-active 25 aa sequence, by furin mediated maturation [27]. Hepcidin expression and consequent modulation of the hepcidin/FPN axis is controlled and coordinated by several regulators that (1) modulate the iron supply and demand (iron status regulator); (2) provide

the proper iron for erythropoiesis (erythroid regulator); (3) limit iron availability to pathogenic microorganisms (inflammatory regulator); and (4) respond to oxygen tension (Figure 1.8) [32].

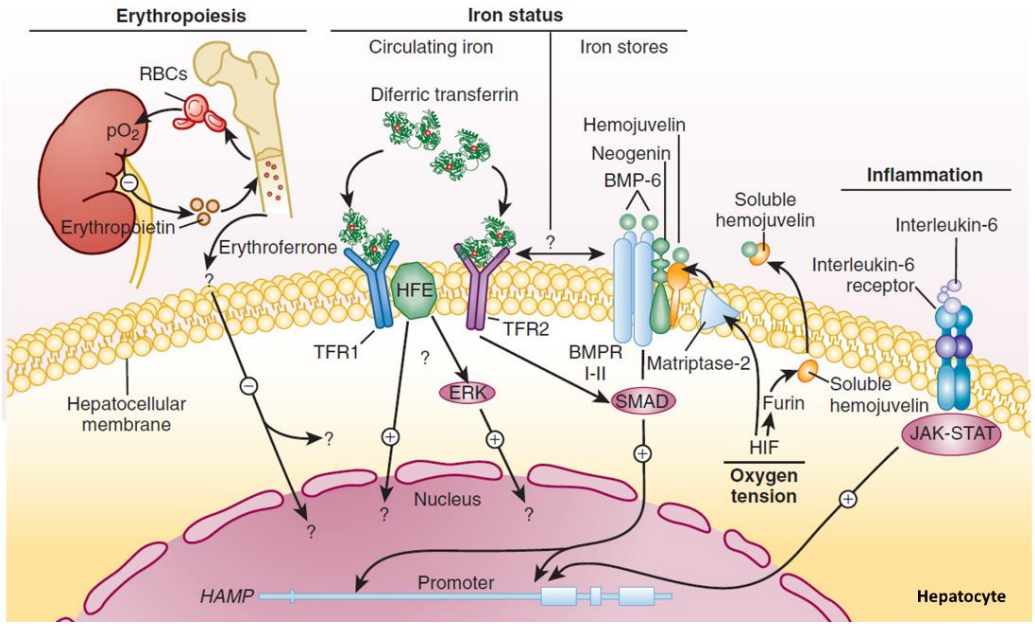


Figure 1.8 – Transcriptional regulation of hepcidin expression in hepatocytes. Erythropoiesis, iron status, oxygen tension, and inflammation regulate hepcidin. Increased erythropoiesis decreases hepcidin expression by mechanisms that remain to be defined. Increased body iron status increases hepcidin expression through two mechanisms: a circulating-iron signal provided by diferric transferrin (Fe_2Tf) and a cellular-iron-stores signal provided by bone morphogenetic protein 6 (BMP6). The Fe_2Tf acts through transferrin receptors 1 and 2 (TfR1/2) and is modulated by the hemochromatosis protein (HFE). The BMP-6 signal acts through its receptor and is modulated by the BMP coreceptor hemojuvelin and neogenin. Decreased oxygen tension leads to decreased hepcidin expression by increasing the transcription of two genes, matriptase-2 and furin, that are responsive to hypoxia-inducible factor (HIF). Matriptase-2 cleaves hemojuvelin from the cell surface, preventing its function as a coreceptor. Furin cleaves hemojuvelin during processing to produce a soluble form that serves as a BMP-6 decoy. Infections and other forms of inflammation increase hepcidin expression by the cytokine interleukin-6. **BMPR** – Bone morphogenetic protein receptor; ***HAMP*** – Hepcidin gene; **JAK-STAT** – Janus-associated kinase–signal transducers and activators of transcription; and **pO_2** – Partial pressure of oxygen. Adapted from Brittenham [20] and Fleming and Ponka [35].

Iron status regulates hepcidin transcription in hepatocytes by circulating iron and tissue iron stores. Regarding the liver iron stores, the hepatic expression of this hormone is accomplished through crosstalk with liver sinusoidal endothelial cells, which produce bone morphogenetic protein (BMP) 6 or 2 (BMP6; BMP2), these consequently initiate intracellular signal transduction through SMAD proteins, activating the hepatocyte BMP-SMAD pathway, that upregulates hepcidin transcription [21], [35]. BMP6 is manly produced by liver nonparenchymal cells in response to hepatocyte iron stores, and somewhat by duodenal enterocytes in response to dietary iron, its activity is enhanced by cell-surface expression of the BMP coreceptor hemojuvelin [20], [35]. BMP2 is less iron sensitive and is highly expressed in basal conditions.

The stronger inhibitor of hepcidin is the liver transmembrane serine protease, matriptase-2 (MT2), which cleaves the BMP co-receptor hemojuvelin, thereby weakens the BMP-SMAD signaling [21]. Furthermore, under hypoxic conditions, HIF transcription factors upregulate expression of MT2, which cleaves hemojuvelin from the hepatocellular surface and attenuates BMP6 mediated signaling to hepcidin [35].

The circulatory iron signal is provided by Tf, which, when iron-bound, serves as a ligand for receptors TfR1 and TfR2. The ferri-transferrin signal seemingly modulates the physical interaction of these two receptors with the hemochromatosis protein (HFE), positively modulating

hepcidin transcription. HFE is a major histocompatibility complex class I–like molecule without iron-transport properties [35], [36]. The HFE/TfR mechanism is not yet completely understood but seems to act through the same SMAD complex activated by BMP6/2 [36].

Increased erythropoiesis reduces hepatic hepcidin synthesis. In this context, hemolysis, hemorrhage, and administration of erythropoietin can also lower circulating hepcidin concentrations. The mechanism by which erythropoiesis modulates hepcidin expression is still undefined, with the erythroferrone (ERFE) hormone being a candidate mediator. The erythropoietic activity can override competing influences, having a greater influence on hepcidin expression than does body iron status [20], [35].

Inflammatory cytokines such as interleukin (IL)-6 upregulate hepcidin expression by activating the Janus-associated kinase–signal transducers and activators of transcription (JAK-STAT) pathway. High hepcidin levels induce iron retention in macrophages, high serum ferritin levels and iron-restricted erythropoiesis. For full hepcidin activation, the IL-6 pathway requires functional BMP-SMAD signaling [21].

Disorders of iron metabolism thought deregulation of the hepcidin/FPN axis are associated with numerous diseases including anemias and iron overload diseases, such as hereditary hemochromatosis and β-thalassemia [37].

1.3 Anemia

Anemia is a public health problem that affects 24.8 % of the world´s population. For Portugal, the World Health Organization (WHO) estimated, with regression models, a prevalence of 15.0 %, but a recent nationwide cross-sectional study (EMPIRE study) has estimated a prevalence of 19.9 % [38], [39]. This prevalence value goes up to 20.6 % when accounting for hematinic factors, attributing to anemia a significance of ‘moderate public health problem’ (prevalence from 20.0 % to 39.9 %), even worse for pregnant women (53.8 %), for whom it is considered a ‘severe public health problem’ (prevalence ≥ 40.0 %) [39], [40].

This condition, which has significant consequences in human health and socio-economic development, happening in both industrialized and non-industrialized countries [41], occurs when the quantity of hemoglobin, and therefore, its oxygen-carrying capacity, cannot respond to the body’s physiologic needs. These vary with demographics and behavior, as well as in different stages of pregnancy, thus the anemia designation relies on the comparison of patient’s sex- and age-appropriate normal range values [7], [40].

There are a variety of causes that result in anemia, those can be isolated or often coexist, such as hemorrhage, chronic inflammation, parasitic infection and nutritional deficiencies, including folate, vitamin B12 and vitamin A, with iron deficiency being regularly the culprit, as it is the most important contributing factor to the global burden of the disease. The pathology can occur at all stages of life; however, it is more prevalent in women, especially pregnant ones, and young children [38], [40].

Although anemia can be acquired by nutritional iron deficiency, when disturbed the molecular processes hereditary pathologies may arise that prevent iron absorption at the duodenal level. Nonetheless, changes in the globin synthesis process also lead to a set of pathologies named hemoglobinopathies, which are responsible for a substantial portion of all hereditary anemia [6].

1.4 Hereditary anemia by iron deficiency

1.4.1 Iron deficiency anemia

Iron deficiency (ID) can be isolated or in association with anemia, representing one of the five major causes of disability burden worldwide [42]. The development of ID results from the interaction of 3 distinct risk factors: increased host requirements, limited supply, and increased blood loss [43].

Iron deficiency anemia (IDA) usually builds up slowly. Firstly, iron stores begin to decline (iron deficiency), consequently lack of iron limits the production of hemoglobin and other metabolically active compounds that require iron (iron restricted erythropoiesis), therefore, it will ultimately entail insufficient iron for the full hemoglobinization of mature RBCs (iron deficiency anemia). At this point, hepcidin's transcription is suppressed to help preserve the supply of iron for vital functions in other tissues [44], [45]. In the same way, hepcidin levels are substantially lower in menstruating women, when compared to males and post-menopausal women, as an effort to maintain normal levels of transferrin and increase iron absorption [21], [44]. In some cases, we can be in the presence of functional iron deficiency, this is a disorder in which total body iron stores are normal or increased, but the iron supply to the bone marrow is inadequate [46].

Clinical features of IDA depend on the severity of the anemia, age, comorbidities, chronicity and speed of onset [46]. This type of anemia can be chronic and frequently asymptomatic, and that is why it may often go undiagnosed. Low delivery of oxygen to body tissues and decreased activity of iron-containing enzymes leads to symptoms such as: weakness; fatigue; difficulty in concentrating; and reduced work capacity. In children and adolescents, IDA can impact motor and cognitive development, and in pregnant women is associated with a risk of preterm labour, low neonatal weight, and increased newborn and maternal mortality. Additionally, iron deficiency subjects are more likely to acquire infections, develop heart failure and restless leg syndrome [47].

Iron deficiency anemia is generally acquired. However, there can be a genetic predisposition for this condition, for instance, iron-refractory iron deficiency anemia [48].

1.4.2 IRIDA – Iron-Refractory Iron Deficiency Anemia

Novel forms of anemia have been identified on account of the recent progress made in the understanding of molecular mechanisms involved in iron homeostasis [43]. Amongst these is iron-refractory iron deficiency anemia (IRIDA; Online Mendelian Inheritance in Man [OMIM] #206200) a rare autosomal-recessive disorder caused by homozygous or compound heterozygous mutations on the matriptase-2 or transmembrane serine protease 6 gene (*TMPRSS6*) located at chromosome 22 (22q12.3) [49]. There is a lack of epidemiological data regarding this disease, the available reports evidence that this rare condition is spread all over the world, occasionally found in different countries and different ethnic groups [48], [50].

Even though IRIDA's prevalence is still unknown, this disease should not be overlooked, especially in children and young adults, when other known causes of refractory IDA are excluded. Identification of affected siblings can be suggestive of an inherited condition, but sporadic cases may occur due to recessive transmission. For this condition, refractoriness to oral iron treatment is clinically defined as a hemoglobin increment of < 1 g/dL after 4-6 weeks of therapy at a daily dose of at least 100 mg of elemental iron [43].

Produced primarily by the liver, MT2 downregulates *HAMP* expression of hepcidin, the systemic iron regulatory hormone, through the BMP/SMAD pathway, cleaving the coreceptor hemojuvelin from the hepatocyte plasma membrane (Figure 1.9). In IRIDA patients, mutations in *TMPRSS6* result in high hepcidin levels that block iron duodenal absorption and iron release from macrophages, leading to low levels of circulating iron, which is insufficient to the erythropoiesis needs. [48], [51]. Mice without functional MT2 (both *mask* mice with truncated *Tmprss6* lacking the protease domain and *Tmprss6* knockout mice) presented hypochromic microcytic, low mean cell hemoglobin (MCH < 27 pg) and low mean corpuscular volume (MCV < 80 fL), anemia and alopecia. These phenotypes, similar in humans, are a result of improper high levels of *Hamp* mRNA expression [51].

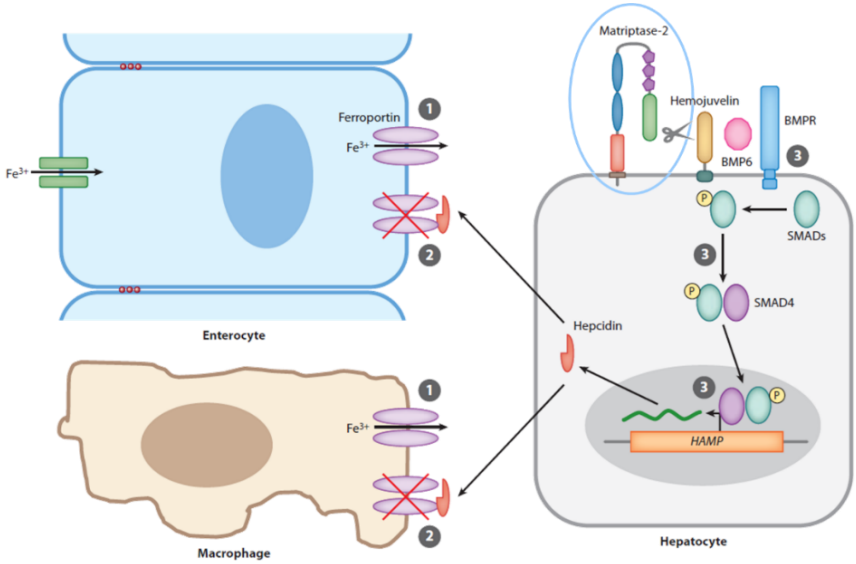


Figure 1.9 – Matriptase-2 regulation hepcidin and subsequent iron export. (1) Dietary iron and iron recycled from erythrocytes are stored in enterocytes and splenic macrophages and are released to the circulation through ferroportin (FPN)**. (2)** Hepcidin produced by hepatocytes binds FPN and targets the channel for degradation. **(3)** Hepcidin gene (*HAMP*) expression is positively regulated by bone morphogenic protein (BMP)-6, which signals through the BMP receptor (BMPR)-SMAD pathway in a hemojuvelin-dependent manner. Matriptase-2 (circled in blue) increases cellular iron export by degrading hemojuvelin to reduce hepcidin production and thus increasing FPN levels. Adapted from Szabo & Bugge [33].

Proteolytic events at the cell surface regulate many important cellular processes requiring transduction of signals across the cell surface [52]. In 2002, Velasco *et al.* identified TMPRSS6 as a member of the type II transmembrane serine protease family (TTSP) [53]. The gene encoding for this protein, in humans, spans for 18 exons and is highly conserved across mammalian species, such as macaque, dog, cow, mouse and rat [49], [52]. MT2 is an 811 aa protein that exhibits a structural organization, similar to other members of the TTSP family, containing several domains: a short cytoplasmic N-terminal domain (exon 1); a transmembrane region (TM; exons 1-2); a sperm protein, enterokinase and agrin (SEA) domain (exons 3-6); two C1r/C1s, urchin embryonic growth factor and BMP1 (CUB) domains (exons 7-8 and 9-11, respectively); three low-density lipoproteins receptor (LDLR) domains, classes A1, A2 and A3 (exons 12-14); a trypsin-like serine protease domain (exons 15–18), with the highly conserved catalytic triad of histidine (H617), aspartic acid (D668), and serine (S762) residues; and finally a C-terminal end (Figure 1.10) [49], [54]. In Table S.1 are summarized the mutations in the *TMPRSS6* gene, and corresponding protein changes in MT2, associated with IRIDA phenotype as described so far in the literature and the ClinVar database.

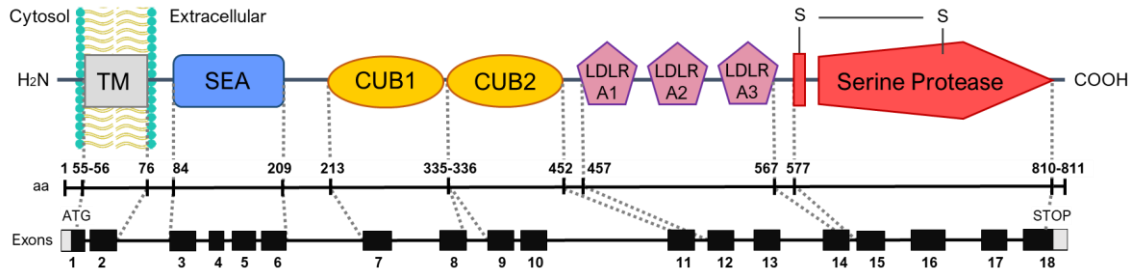


Figure 1.10 – *TMPRSS6* gene structure and matriptase-2 schematic representation. Matriptase-2 (MT2) contains a short N-terminal cytoplasmic domain, a transmembrane region (TM), a sperm protein, enterokinase and agrin (SEA) domain, two C1r/C1s, urchin embryonic growth factor and BMP1 (CUB) domains, three low-density lipoproteins receptor (LDLR) domains, classes A1, A2 and A3, a trypsin-like serine protease domain, and finally a C-terminal end. Original figure.

Matriptase is synthesized as an inactive, single-chain, membrane-bound polypeptide (zymogen) that undergoes activation into an active protease by autocatalytic cleavage [33], [52]. Zymogen activation is irreversible, and proteolytic activity typically is terminated either by proteolytic shedding of the protease from the cell surface or by its internalization and lysosomal degradation [33].

In contrast to the current understanding of autosomal recessive disorders, haploinsufficiency is observed in some *TMPRSS6* mutations, both in humans and murine models [51]. Thereby, gene sequence variants lead to a spectrum of protein dysfunction, going from severe loss of function mutations, causing classic IRIDA, to hypomorphic mutations that result in a mild reduction of MT2's activity, such as the common single nucleotide polymorphism (SNP) rs855791 - V736A in the catalytic domain [55], which is associated with variation in iron status within the general population, resulting in genetic susceptibility to IDA [56], [57].

1.4.3 Diagnosis

Anemia cannot be reliably diagnosed by the clinical presentation of symptoms like fatigue. For this reason, in the daily clinical routine, anemia is defined by a hemoglobin concentration, as described in the WHO guidelines, which vary by sex and age (Table 1.1) [58].

Table 1.1 – Hemoglobin levels to diagnose anemia at sea level.

Population	Non-anemia Hb (g/dL)	Anemia Hb (g/dL)		
		Mild	Moderate	Severe
Children 6 - 59 months of age	≥ 11.0	10.0-10.9	7.0-9.9	≤ 7.0
Children 5 - 11 years of age	≥ 11.5	11.0-11.4	8.0-10.9	≤ 8.0
Children 12 - 14 years of age	≥ 12.0	11.0-11.9	8.0-10.9	≤ 8.0
Non-pregnant women (15 years of age and above)	≥ 12.0	11.0-11.9	8.0-10.9	≤ 8.0
Pregnant women	≥ 11.0	10.0-10.9	7.0-9.9	≤ 7.0
Men (15 years of age and above)	≥ 13.0	11.0-12.9	8.0-10.9	≤ 8.0

Adapted from WHO [40]; **Hb** – Hemoglobin.

IDA is the only microcytic anemia disorder in which mobilizable iron stores are absent, whereas, in all others, iron stores are normal or increased, for this reason, the diagnosis depends on indirect indicators of iron status. Both transferrin saturation (TSAT) and serum ferritin are consensually decreased in iron deficiency anemia, with serum Ft being the most sensitive and specific test to identify isolated iron deficiency since it reflects low stores. Albeit, inflammation must be taken into consideration since ferritin is a positive acute phase response protein, thereby will no longer reflect the size of the iron store (Table 1.2). As ID progresses, because of low iron

and increased transferrin synthesis, TSAT that is the proportion of the available iron-binding sites on transferrin that are occupied by iron atoms decreases < 16% [44]. Out of the iron status parameters, the least reliable one for diagnosis of ID is serum iron (SI), because it could be detected as an artefact of contamination of laboratory equipment, has a nocturnal rhythm and it may normalize hours after iron ingestion [59].

Table 1.2 – Relative extent of iron stores based on serum ferritin concentration.

Status	Serum ferritin (µg/L)		
	Less than 5 years of age	5 years of age or older	
		Male	Female
Depleted iron stores	< 12	< 15	< 15
Depleted iron stores in the presence of infection	< 30	-	-
Severe risk of iron overload (adults)	-	> 200	>150

Adapted from WHO [60].

Red cell distribution width (RDW) is a measure of the variation of red cell volume, in IDA is normally increased, particularly in the initial phases of iron deficiency, when newly produced microcytic and hypochromic RBC coexist with normal ones. The same happens after iron therapy when abnormal erythrocytes are beginning to be substituted by normal ones [44].

The diagnosis of IDA often is confirmed by the outcome of a therapeutic trial of iron (Figure 1.11). A specific sequential response to treatment with iron constitutes the final definitive proof that a shortage of iron is the true cause of anemia. The diagnostic response consists of reticulocytosis, (3 to 5 days after adequate iron therapy), followed by a significant increase in hemoglobin concentration (3-4 weeks) that will persist until normal values are restored (2-3 months). Results from this test must be assessed for possible confounding factors, such as poor compliance with iron therapy; malabsorption of therapeutic iron, for example, due to interaction with other medication; continuing blood loss; the effects of coexisting conditions, in particular infectious (*Helicobacter pylori* or parasitic infections), chronic inflammatory diseases, or malignant disorders; and genetic conditions such as IRIDA. In the end, the therapeutic trial only stablishes the presence of iron deficiency, therefore, the search for underlying causes IDA must continue despite a positive response to iron therapy [45].

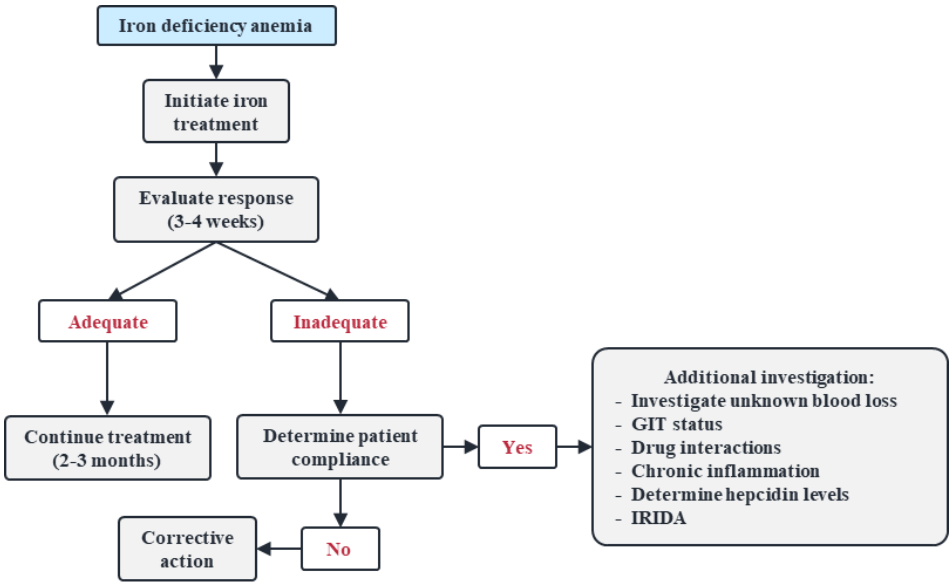


Figure 1.11 – Algorithm for evaluation and treatment of iron deficiency anemia by therapeutic trial. GIT – Gastrointestinal track; **IRIDA** – Iron deficiency iron-refractory anemia. Adapted from Harmse [61].

In IRIDA, we often observe in patients extremely low transferrin saturation levels and normal/borderline low serum ferritin levels, differentiating IRIDA from other microcytic anemias. Measuring serum hepcidin comes in handy for this diagnostic, because the hormone levels are considerably elevated when compared to IDA. While the hepcidin assay can be reliable, at present it is only used in research, since it is not yet widely available for diagnostic purposes [44]. Final IRIDA diagnosis requires identification of genetic lesions by sequencing the exons and exon-intron boundaries of the *TMPRSS6* gene [43].

1.4.4 Treatment

There are several approaches when treating IDA that are dependent on the underlying cause [62]. Before iron therapy prescription, if dietary history suggests a deficiency, patients should be steered into boost their diet with foods rich in heme iron (red meat or liver) and non-heme iron (cereals, egg yolk and green leafy vegetables). Citrus fruits are of added value since vitamin C is known to increase iron absorption, by creating an acidic gastric environment. Although dietary review and counselling are important, diet alone is usually insufficient when treating most IDA patients [63].

Overall, the need for iron treatment is high in children and modest in adults who have limited iron requirements [43]. Oral iron preparations commonly contain one of three iron salts: iron sulphate, iron gluconate, and iron fumarate, which may be prescribed as tablets or elixirs. On occasion, oral supplementation treatment does not result in increased hemoglobin [63]. Notwithstanding the refractory response to iron of IRIDA patients, not all subjects are completely refractory to oral iron treatment, with long-term administration of oral iron being able to induce partial correction of anemia in patients from unrelated families [43]. Parenteral iron therapy is reserved for iron intolerant patients, regardless of regime modifications, and those with severe chronic anemia caused by advanced inflammatory diseases, uncontrollable blood loss and gastrectomy or small bowel resection [45].

1.5 Hereditary anemia by changes in the globin chains – Hemoglobinopathies

Inherited hemoglobin disorders are the most common human monogenic diseases, with approximately 7 % of the world's population being a carrier and around 300 000 to 400 000 babies being born each year with severe forms of these diseases, accounting for about 3.4 % of deaths in children less than 5 years of age, and so resulting in a heavy burden for health care [64]. The global geographical distribution of the hemoglobinopathies is nowadays properly documented, being mostly endogenous to the population of South Europe, the Mediterranean countries, Middle East, South East Asia, and other malaria-endemic countries in Africa, since these disorders render heterozygote advantage against *Plasmodium falciparum* [65], [66]. Population movements have had a major impact on disease epidemiology and public health, with migration from these parts of the world leading to the introduction of hemoglobinopathies into new populations, hence the importance of geographical ancestry when studying these diseases [65].

Hemoglobinopathies generally fall into three main groups: (1) structural hemoglobin variants, these are qualitative changes to the globin chains normally resulting from single amino-acid substitutions that may alter the stability or functional properties of the hemoglobin molecule, therefore possibly leading to a clinical disorder [17], [64]; (2) thalassemias, these are a heterogeneous group of inherited anemias characterized by the production of structurally normal, but decreased quantities or absence of one or more globin chains subunits, gaining their name according to the particular globin chains that are ineffectively synthesized (e.g. α- and β-

thalassemia). The resulting syndromes arise from the combined consequences of inadequate hemoglobin production and imbalanced accumulation of globin subunits, the former causes hypochromia and microcytosis, the latter leads to ineffective erythropoiesis and hemolytic anemia [64], [67]; and (3) failure to switch globin chain synthesis, this is the result of continued synthesis of high levels of Hb F in adult life leading to hereditary persistence of fetal hemoglobin, a condition that influences the severity of clinical manifestations of other hemoglobinopathies [17], [67].

Normally, allelic interaction in hemoglobinopathies can present themselves in three forms: one involving one of the two β or four α allelic genes in 'carrier' individuals, a heterozygous condition sometimes referred to as a trait, an autosomal codominant character; another when two similar allelic variants are found in the same individual forming a homozygote; and finally simultaneous presence of two different allelic variants constituting a compound heterozygote. Within these disorders, some might be endemic, occurring in very high frequency, with variable clinical presentation and for this reason presenting a great impact in the populations, while others are rare and/or sporadic representing a unique opportunity to experiment and better understand hemoglobin function [68].

1.5.1 Structural hemoglobinopathies

To date, there have been described nearly 1000 mutations that alter the structure of hemoglobin, with most having no clinical consequence [69]. The functional abnormal effect is dependent on where the mutation occurs, and if clinically relevant, it results in either an increased tendency to aggregate due to abnormal polymerization, to the instability of the hemoglobin molecule causing hemolytic anemia or altered O_2 affinity [17]. Certain structural hemoglobin variants are characterized by the presence of an abnormal structure as well as a biosynthetic defect, and thus can be responsible for unusual forms of thalassemic hemoglobinopathies, of this are examples Hb E and Hb Knossos [8], [69]. In thalassemic hemoglobinopathies globin chain synthesis is reduced because the corresponding nucleotide substitution leads to abnormal splicing or causes instability and prompt degradation of the corresponding mRNA or proteolysis of the variant chain [68].

Abnormal polymerization: The best known examples of the malaria parasite exerting selective pressure on the human population are the most common structural hemoglobinopathies S (sickle cell trait; Hb S) and C (Hb C), in which the sixth position glutamate of the *HBB* gene is replaced by valine and lysine, respectively. By reducing the cytoadhesive capacity of parasitized erythrocytes, HbS and Hb C seem to mitigate the life-threatening complications that derive from the infection [70]. In patients homozygous for the disorder, called sickle cell anemia or Hb SS, occurs polymerization of the variant upon deoxygenation through the extensive formation of strong hydrophobic bonds resulting in deformation of the RBC, which acquire the sickle cell shape and hinder the microcirculation of the patient. [68]. As for Hb C, this variant induces erythrocyte dehydration and intracellular crystal formation, when in homozygous state results only in mild hemolytic anemia [71].

Instability: Unstable hemoglobins are those exhibiting reduced solubility or higher susceptibility to oxidation of aa residues within globin chains, however, most present only mild instability in *in vitro* laboratory tests, therefore, are associated with minimal clinical manifestations. Impairment of hemoglobin's solubility usually arises from mutations that disturb hydrogen bonding or the hydrophobic interactions, which either retain the heme group within the heme-binding pockets or hold the tetramer together [69]. The resulting instability promotes the

separation of the heme group from the globin chain with the slightest oxidative stress. Once freed from its cleft, heme will bind nonspecifically to other regions of the globin chains forming precipitated hemichromes leading to further denaturation and aggregation of the globin subunits. To these denatured hemoglobin forms, containing α- and β-globin chains, globin fragments, and heme is given the name Heinz body, these interact with the delicate erythrocyte membrane components, thereby reducing cell deformability, rigid cells are consequentially detained in the splenic for removal causing hemolytic anemia [17], [69].

Changes in O_2 affinity: Efficient oxygen delivery by hemoglobin is dependent on the sigmoid shape of the hemoglobin-oxygen affinity curve (Figure 1.12). During the transition from the completely deoxygenated to oxygenated state, the first oxygenation steps take place with difficulty. After that, allosteric cooperation and subsequent oxygen-binding events bring about the precise arrangement of hydrogen bonds, hydrophobic interactions, and salt bridges are broken, rearranging the hemoglobin molecule, thus creating the sigmoid shape of the curve [16], [69].

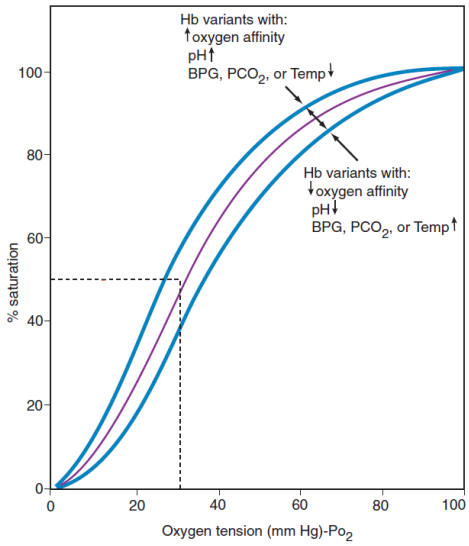


Figure 1.12 – Oxygen dissociation curve of hemoglobin. On the abscissa, the partial pressure of oxygen (pO_2) is indicated in millimeters of mercury. On the left ordinate, the saturation of hemoglobin with oxygen is indicated as a percentage. The percent saturation of hemoglobin with oxygen at different oxygen tensions is depicted by the red sigmoidal curve. The P50 (i.e., oxygen tension at which the hemoglobin molecule is one-half saturated) is approximately 27 mm Hg in normal erythrocytes (dotted lines). Heterotopic modifiers of hemoglobin including pH, temperature and small organic phosphate molecules function can shift the curve leftward by increasing or rightward by decreasing its oxygen affinity. **pCO_2** – Partial pressure of carbon dioxide; **BPG** – 2,3-Bisphosphoglyceric acid. Adapted from Steinberg *et al.* [8].

Hemoglobin variants exhibiting altered oxygen affinity arise from aa substitutions at the interface between globin chains or in regions affecting the hydrogen bonds, hydrophobic interactions, or salt bridges, which influence the interaction of heme with oxygen. Thereby, high-affinity variants feature a tight binding of O_2, delivering less oxygen to the tissues. Inefficient tissue oxygenation is followed by tissue anoxia, increased erythropoietin secretion, and erythrocytosis [17], [69]. On the contrary, most low-affinity variants possess enough oxygen affinity to become fully saturated in the normal lung, at the low capillary pO_2 in the tissues they deliver higher than normal amounts of oxygen. Consequences from the high level of oxygen delivery include a state of "pseudoanemia", in which the low hematocrit level is deceiving because both oxygen delivery and the patients are normal since oxygen requirements can be met by lower-than-normal hematocrit levels, and cyanosis resulting from the amount desaturated hemoglobin circulating in capillaries and veins [69].

1.5.2 Beta-thalassemia

Beta-thalassemia (BT) can be defined as a syndrome of inherited hemoglobin disorders marked by a quantitative deficiency of functional β-globin chains [3]. These syndromes can be the result of point mutations or deletions either restricted to the *HBB* gene or involving β-LCR [72]. Worldwide, WHO has estimated that about 5 % of the population might be carriers of β-thalassemia (β/β^T) and that about 60 000 severely affected infants are born every year [6].

According to severity, there is a wide phenotypic spectrum in β-thalassemia (Figure 1.13) ranging from thalassemia minor, or β-thalassemia trait (BTT), which consists of mild hypochromic microcytic anemia without evident clinical manifestations, to β-thalassemia major (BTM), which is characterized by severe anemia since the first years of life and is transfusion-dependent (TDT). In the middle lies β-thalassemia intermedia (BTI), a term developed to describe patients with manifestations too mild to be considered BTM and too severe to be called BTT since these patients may be transfusion-dependent or require occasional or intermittent transfusion [73], [74].

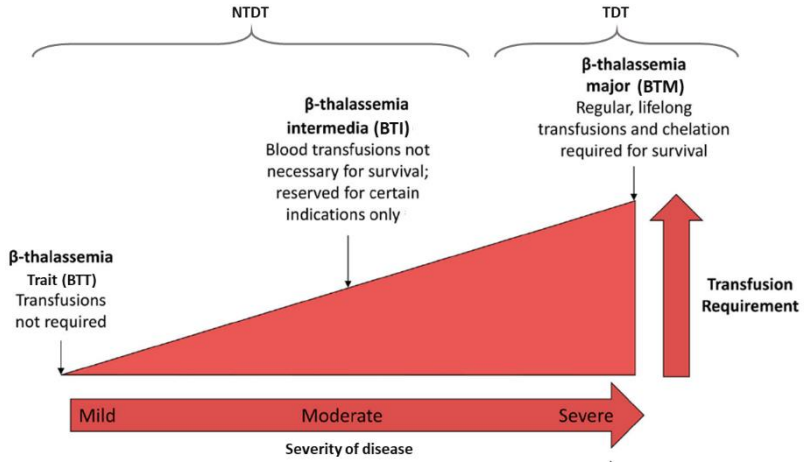


Figure 1.13 – Spectrum of β-thalassemias according to disease severity and transfusion requirement. **NTDT** – Non transfusion-dependent thalassemia; **TDT** - Transfusion dependent thalassemia. Adapted from Salah *et al.* [73].

The underlying pathophysiology responsible for β-thalassemia stems from the quantifiable deficiency of functional β-globin chains, which leads to an imbalanced globin chain production and an excess of α-globin chains (Figure 1.14). Alpha chains aggregate in red cell precursors, forming inclusion bodies, which cause mechanical damage, leading to premature destruction in the BM, causing impaired erythropoiesis, surviving RBCs that reach the peripheral circulation are also prematurely destroyed in the spleen. Thus, anemia in β-thalassemia results from a combination of ineffective erythropoiesis, peripheral hemolysis and an overall reduction in hemoglobin synthesis [3].

The clinical severity variability of this disease reflects both genetic and environmental factors, the former involves many loci, some of which are directly involved with the basic defect in globin synthesis, whereas others, which modify the variable complications of the disease, have nothing to do with globin chains [66]. One powerful factor, with ameliorating effect on the disease, is the high levels of Hb F expression that persist to various degrees in β-thalassemia. This happens because γ-globin can substitute for β-globin, and at the same time generating more functional hemoglobins and reducing the α-globin inclusion burden. Theoretically, patients may also vary in their ability to solubilize unpaired globin chains by proteolysis [67].

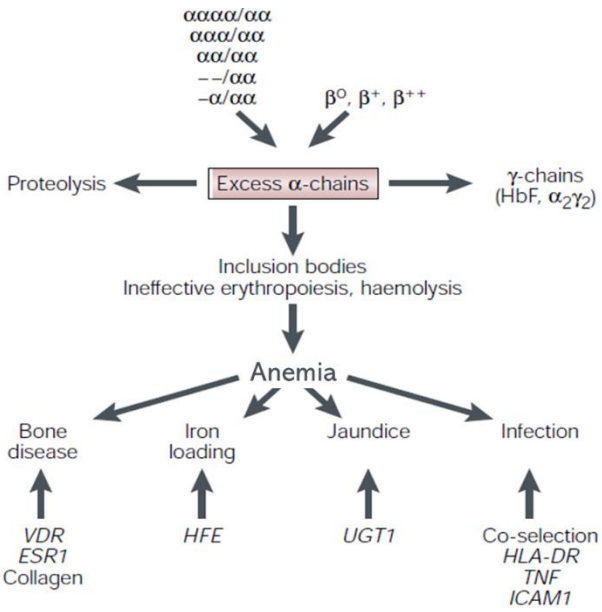


Figure 1.14 – Pathophysiology of β-thalassemia and main genetic mechanisms that contribute to the phenotypic diversity. Factors that modify the β-thalassemia phenotype act at three levels: (1) primary – generally these refer to the nature of the mutation affecting the β-globin gene itself, (2) secondary – loci that affect the α/β globin chain imbalance and include the α- and γ-globin genes. However, loci that affect the stability and amount of the α chain and loci that affect the expression of the γ globin should also be included; and (3) tertiary – loci that are not involved in globin imbalance but modify the complications of the disease in different ways, including bone disease (*VDR, ESR1*, Collagen), jaundice (*UGT1*), iron loading (*HFE*) and infection (*HLA-DR, TNF, ICAM1*). Adapted from Thein [3] and Weatherall [66].

The primary factor that modifies the disease phenotype is the individual alleles that vary according to the severity of the biosynthetic lesion [67]. In this context, thalassemia alleles are remarkably diverse in their effect on the output of β-globin chains, with β^0-thalassaemias being those resulting in no β-globin gene product, and β^+- or β^{++}-thalassemia being those in which exists a marked or mild reduction in the output of β-chains, respectively [72]. In general, carriers of β^+-thalassemia mutations have milder hematological abnormalities than those with β^0-thalassemia mutations [75]. Albeit, patients who are compound heterozygous for a silent mutation and either a β^+- or β^0-thalassemia mutation can present with BTI, with significant anemia. For this reason, correct diagnosis of carriers of silent β^+-thalassemia mutations is important for family counseling, with alleles being most easily described by their phenotypic effects in heterozygotes [72], [75].

The geographical distribution of the different BT alleles (Figure 1.15) shows us that mild β-thalassemia alleles, are relatively uncommon in many high-frequency populations [66]. Particularly in Portugal, there are four common mutations: codon 39 (C>T), IVS-I-1 (G>A), IVS-I-6 (T>C) and IVS-I-110 (G>A) [76], [77]. Variability quite noticeable, when two siblings inheriting identical thalassemia mutations sometimes exhibit markedly different degrees of anemia and erythroid hyperplasia [67]. Inheritance of only one β-thalassemia allele usually results in a mild hypochromic microcytic anemia. These subjects' levels are in average 1 or 2 g/dL lower than the ones seen in normal persons of the same age and gender. Additionally, Hb F levels decline more slowly than usual in the first year of life, and the diagnostic elevated Hb A2 levels are established by approximately 6 months of age [67].

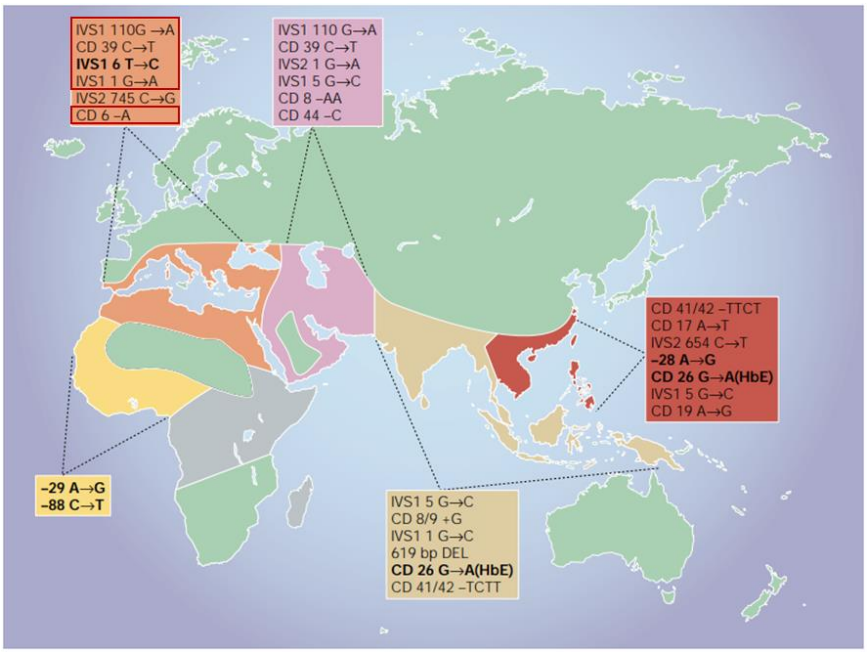


Figure 1.15 – The global distribution of the β-thalassemia mutations. The common mild mutations are shown in bold. In a red box are the common mutation in Portugal. Adapted from Weatherall [66] and Faustino *et al.* [76], [77].

In BT the unbound α-globin bind free heme molecules in the absence of β-globin chains, forming hemichromes, which are toxic insoluble aggregates that precipitate and damage RBC membranes and form ROS, leading to oxidative stress [36]. The subsequent impaired erythropoiesis and erythroid hyperplasia that could induce excess iron absorption and ultimately lead to iron overload, since hepcidin production is suppressed by the increased production of ERFE in the absence of blood transfusion and during the early stages of the disease. Over time iron accumulates in the organs, hepcidin synthesis increases likely as a result of increasing TSAT and increasing iron concentrations in the liver [36], [78]. However, only a small percent of patients with the BTT develop iron overload, suggesting that other factors are involved in these cases. Some studies indicate that iron overload in these patients is the result of a synergistic effect of the β-thalassemia trait and heterozygosity for common hemochromatosis (HH) mutation on the *HFE* gene (Figure 1.16). In the context of HH, the most common mutations are rs1799945 – H63D and rs1800562 – C282Y, being that the latter leads to a greater loss of function of the HFE protein than the former [66], [78].

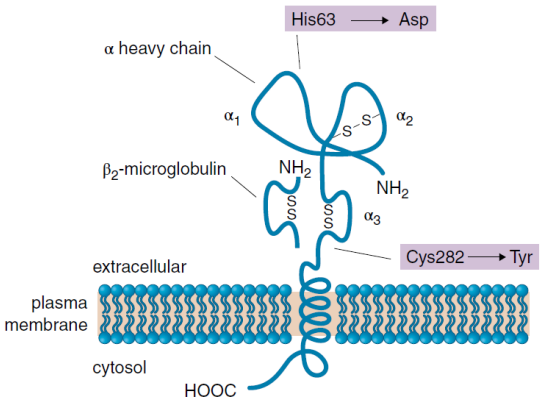


Figure 1.16 – HFE protein model. Hypothetical model of the protein derived from the *HFE* gene based on its homology to the HLA class 1 molecule. The approximate locations of the C282Y and H63D mutations found in primary hemochromatosis are shown. Adapted from Hoffbrand *et al.* [79].

The frequency of the HFE mutations in control subjects has been reported in several European countries, with Northern European countries presenting the highest allelic frequencies for the C282Y [80]. In Portugal, on one hand the allelic frequency of the C282Y mutation decreases from the North (5.9 %) to the South (9.9 %) and is less prevalent in the Madeira islands (3.3 %), on the other the allelic frequency of the H63D is independent of the region ranging 17.0-19.0 %, with similar prevalence in the islands [80], [81].

1.5.3 Alfa-thalassemia

Alpha-thalassemia is due to one of the most common hemoglobin genetic abnormalities and is caused by the reduced or absent production of the alpha globin chains. Contrary to β-thalassemias, point mutations are uncommon in α-globin genes. Most defects come from deletional mutations and a few are described in *cis* forms with point mutations ($-\alpha^T$), normally in the *HBA2* gene [82], [83].

The disease can present four clinical conditions of increased severity: (1) α^+-thalassemia trait (α-/αα);, usually caused by the deletion or dysfunction of one of the four normal α-globin genes (2) Homozygous α^+-thalassemia trait (α-/α-) or α^0-trait (--/αα), both carrier states resulting from deletion or dysfunction of two alpha genes in cis; (3) Hb H disease (α-/--) caused by the presence of only one functioning alpha gene; and (4) Hb Bart hydrops fetalis syndrome (--/--) where there are no functioning alpha genes (Figure 1.17) [82], [83]. The last two represent the clinically relevant forms of the disease, in this disorders the underproduction of α-globin chains gives rise to excess β-like globin chains which form γ^4 tetramers, called Hb Bart's (in fetal life) and β^4 tetramers, called Hb H (in adult life) [84].

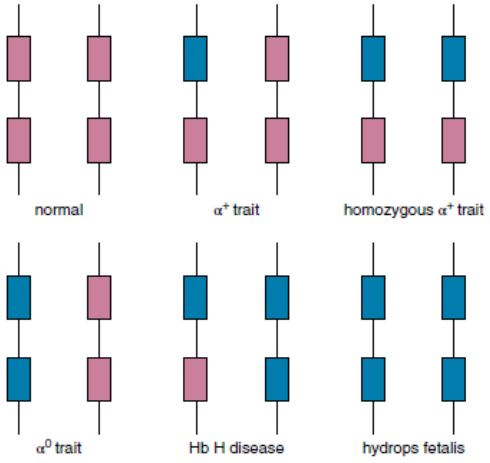


Figure 1.17 – The different types of α-thalassemia. The purple boxes represent normal genes, and the gray boxes represent gene deletions or partially or completely inactivated genes. Adapted from Hoffbrand *et al.* [79].

Clinical characteristics of thalassemia trait are mild, microcytosis, hypochromia, and mild anemia, compared to Hb H disease that presents as moderate to severe microcytic, hypochromic, hemolytic anemia, mild jaundice, moderate hepatosplenomegaly. Hb Bart hydrops fetalis syndrome, the most severe form, is characterized by very severe anemia, marked by hepatosplenomegaly, hydrops fetalis, cardiac failure and abnormalities in the skeletal and urogenital systems. For this reason, this condition is usually not compatible with postnatal life and affected fetuses are either stillborn or die soon after birth [82]. These thalassemias are more difficult to diagnose than β-thalassemias because characteristic elevations in Hb A2 or Hb F do

not occur. However, the gene deletions responsible for the most common varieties (Figure 1.18) are readily detectable by molecular biology methods [67].

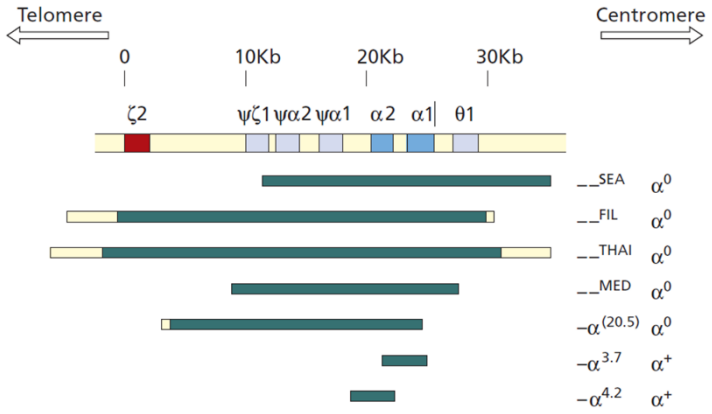


Figure 1.18 – Diagrammatic representation of common deletions that can lead to α-thalassemia trait. The shaded blocks indicate the length of the deletion. Adapted from Bain [85].

Overall, α-thalassemias follow a similar distribution to the β-thalassemias (Figure 1.19). The α^+ deletional forms are in high frequencies throughout the tropical belt and α^0 deletions are highest in South East Asia. Common α^0 deletions are --MED in Mediterraneans and --SEA in South East Asians [83].

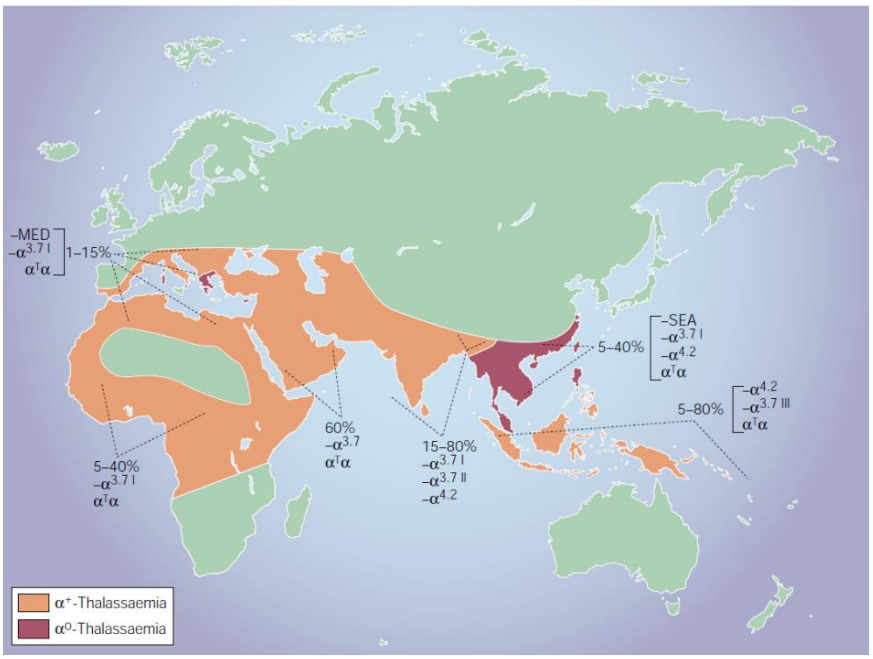


Figure 1.19 – The global distribution of the α-thalassemias. The α^+-thalassemias result from deletions of 3.7 kb or 4.2 kb, which remove a single α-globin gene. There are three sub-varieties of $-\alpha^{3.7}$, designated $-\alpha^{3.7\ I}$, $-\alpha^{3.7\ II}$ and $-\alpha^{3.7\ III}$, depending on the site of the crossover event that underlies the deletion. **MED** – Mediterranean; **SEA** – Southeast Asia. Adapted from Weatherall [66].

1.5.4 Failure to switch globin chain synthesis

Sometimes, expression of γ-globin persists in adult RBC forming a state known as hereditary persistence of fetal hemoglobin (HPFH), mostly results of a common genetic variant [3]. This state is largely asymptomatic since no deleterious effects on patients are observed even when total of the hemoglobin produced in HPFH homozygotes is Hb F. [11], [67]. Although this state is not

given much clinical importance, the co-inheritance of some forms of HPFH is able to modify the phenotypes associated with the structural hemoglobin variants or thalassemias [66]. Hence, these patients demonstrate that reversal of the Hb F to Hb A switch would provide efficacious therapy for β-thalassemia and Sickle Cell Disease [67].

1.5.5 Diagnosis

Molecular diagnosis of hemoglobinopathies may be required for several reasons: to confirm a hematological/biochemical provisional diagnosis; to explain anemia or microcytosis; to identify an abnormality in the pre-symptomatic phase, as in neonatal screening; to identify fetuses at risk of significant hemoglobinopathies and offer the parents an informed choice and genetic counselling, keeping in mind that for this prenatal or preimplantation a prerequisite is timely identification of carrier couples; and finally to identify the presence of sickle cell hemoglobin preoperatively [19], [86]. It is important to know the difference between screening and diagnosis. National carrier screening programs are not designed to detect 100 % of all possible hemoglobinopathies in a carrier, due to the costs and resources required of following up [87]. In Portugal is currently in place a geographically directed screening model. Nevertheless, there are a big number of unscreened individuals since not all the population is eligible, the model does not account for migration flows within the country and there are implementation difficulties within the clinic [88].

Laboratory diagnosis of hemoglobinopathies is carried out based on an analytic march, the diagnosis relies on cut-off indices widely used, however appropriate reference values should be independently defined for each population as there may be slight differences according to the types of thalassemia alleles present [87]. The analysis starts with presumptive identification of hemoglobinopathy, which requires a minimum of two techniques based on different principles, for example complete blood count (CBC) followed by high performance liquid chromatography (HPLC) or capillary electrophoresis (CE). After that, definitive identification is performed through DNA analysis, mass spectrometry or protein sequencing, if possible adding family studies help to elucidate the nature of disorders [19], [88]. Previously α-thalassemia was confirmed by globin chain biosynthesis, when the α/β globin chain biosynthesis ratio was reduced to less than ~0.8 (Figure 1.20) [84]. The algorithm currently for diagnosis of hemoglobinopathies in context of carrier screening is shown in Figure 1.21.

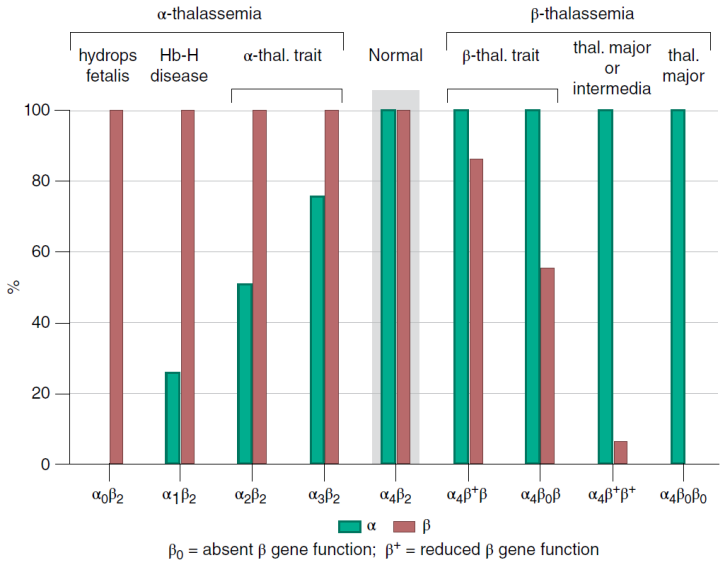


Figure 1.20 – Ratio of α:β chain synthesis in the different α- and β-thalassemias. Adapted from Hoffbrand *et al.* [79].

Once identified a hemoglobin variant, molecular confirmation usually starts by sequencing of the *HBB* gene due to its relatively small size and because it is home of the most common point mutations. When it comes to the α-globin genes, *HBA1* and *HBA2*, having a sequence homology bigger than 96 % makes them difficult to sequence, but many of the common α-thalassemia mutations are large deletions which cannot be found by sequencing, however, they can be found using gap polymerase chain reaction or Multiplex Ligation-Dependent Probe Amplification [89].

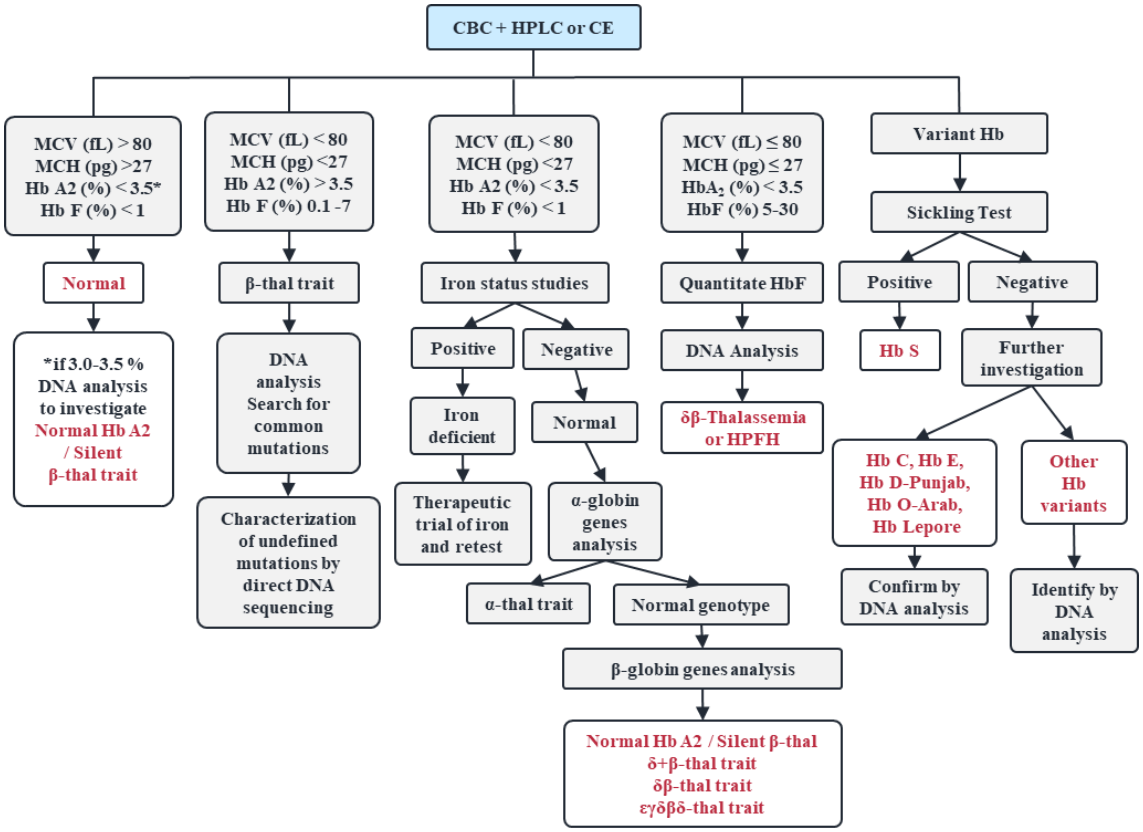


Figure 1.21 – Algorithm for diagnosis of hemoglobinopathies in context of carrier screening. CBC – Complete blood count; **HPLC** – High performance liquid chromatography; **CE** – Capillary electrophoresis; **MCV** – Mean corpuscular volume; **MCH** – Mean corpuscular hemoglobin; **Hb** – Hemoglobin. Adapted from Old [87].

The carrier diagnosis flow chart can arrive to five different results using the well-established and generally accepted cut off values. Although these are designed to detect one of the main types of hemoglobinopathy, there are diagnostic problems such as microcytosis due to iron deficiency which complicate the diagnostic algorithms of the flow chart [87]. Both BTT and IDA are microcytic and hypochromic anemia, making it difficult to differentiate one from another only by hematologic parameters obtained from automated blood cell analyzers [90]. Testing to distinguish these two disorders is important for genetic counseling of carriers of thalassemia trait, as well as to avoid unnecessary investigations in patients incorrectly assumed to have iron deficiency and to prevent unjustified supplementation with iron [67]. In case of BTT presenting with iron overloading, as well as for the moderate and severe forms of thalassemia, high liver iron concentration (LIC) should be assessed with magnetic resonance imaging, which has replaced liver biopsy as the gold standard for the quantification of LIC given its safety and reliability [91].

In case the investigation of microcytosis is non-urgent, there have been published various mathematical formulas for discriminating between iron deficiency and thalassemia trait that can be used to indicate which diagnosis is more likely and which test should be carried out first. However, it should be noted that each of these indices has some degree of inaccuracy and are

subjected to variation in accord with factors such as age, sex and geographical ancestry. Therefore, these should not be used in decision making with reference to women who are already pregnant so that thalassemia trait can be correctly excluded [92].

Progress in the techniques of clinical practice and molecular analysis of hemoglobinopathies has made possible for couples that are at risk of transmitting a serious disease to their offspring and to have a healthy baby through prenatal or preimplantation diagnosis [86].

1.5.6 Treatment

The life expectancy and QoL of patients with severe hemoglobin disorders can be significantly improved when using advanced treatment methods. Following diagnosis, patients should be referred to genetic counseling and hematology appointments for counseling and decision of treatment [93].

Regarding the treatment for hemoglobin variants, it differs according with the presented clinical complications. Typically, patients with sickle cell syndrome are anemic but lead a relatively normal life marked by painful episodes. However, it is important to realize that chronic organ damage and decreased survival occur even in patients without painful events. Therefore, treatment includes clinic management directed at initiating measures to prevent pain crisis, prevent organ complications, and improve survival, by the means of analgesics, antibiotics, hydroxyurea and eventually transfusion therapy [94]. When it comes to unstable hemoglobins the severity of clinical complications varies greatly. While many patients can be managed just by observation and education on how to avoid hemolysis provoking agents, others require transfusions during severe acute hemolytic anemia phases, and in worst case scenario, in the presence of repeated episodes, these patients should be considered candidates for splenectomy. Finally, concerning variants with altered affinity to O_2, on one hand patients with low-affinity hemoglobins are usually asymptomatic and require no treatment, on the other hand patients with high-affinity hemoglobins have mild erythrocytosis and may, on occasion, if blood viscosity is sufficiently elevated require therapeutic phlebotomy. Additionally, in the event of carbon monoxide poisoning they can be treated with supplemental oxygen [69]

The treatment for BT is given according to the severity of the disease: BTT individuals should be given folic acid supplements for anemia, since iron supplements are contraindicated unless there is simultaneous iron deficiency; for BTI transfusion therapy is indicated for patients with complications as a result of major increases in erythropoiesis, as well as for those with anemia and an inability to maintain stable hemoglobin levels of more than 8 g/dL; finally BTM supportive treatment includes lifelong regular transfusions combined with effective iron removal. Treatment for α-thalassemias trait is normally not require, regular anemia should be supplemented with folic acid, similarly to β-thalassemia. For Hb H disease, the treatment is dependent on the severity of the clinical symptoms and transfusions are rarely indicated. Contrarily, Hb Bart's syndrome, requires transfusions in utero and continuously after birth, furthermore stem-cell transplantation can be performed [93].

1.6 Objectives

When it comes to the study of hereditary anemia this dissertation focusses on four objectives:

1. The study of clinical cases of hemoglobinopathies by PCR, gap-PCR, MLPA and automated Sanger sequencing of the various globin genes in order to explain atypical phenotypes and understand underlying pathophysiological mechanisms.

2. The study of clinical cases with suspicion of IRIDA trough amplification of the gene *TMPRSS6* by long-PCR and Next-generation Sequencing (NGS) analysis followed by *in silico* study of possible pathogenic effect of the detected new/rare genetic variants.

3. Investigate the presence and role of common polymorphisms in the genes *TMPRSS6* (V736A and K253E) and *HFE* (H63D and C282Y) on the hematological parameters and the iron status of 317 Portuguese subjects divided into groups: one of β-thalassemia trait subjects, other of patients with iron deficiency anemia and controls.

4. Evaluate the diagnostic performance of mathematical indices when differentiating β-thalassemic trait and non-thalassemic (iron deficiency anemia) microcytosis, to find which ones would better apply to the Portuguese population.

The first two objectives give us the opportunity to integrate genomic approaches into routine clinical care, that despite comprehensive clinical evaluation, sometimes is unable to arrive at a definitive diagnosis, thus restricting clinical management. Therefore, with these objectives we aim to reach an accurate genetic diagnosis, which will benefit patients by providing a better understanding of their prognosis and allowing clinicians to personalize treatment and custom management and surveillance.

For the third objective we undertook a disease-association study of risk alleles, so that we could find the impact that the studied polymorphisms have in our control and case populations, while verifying, sometimes contradictory, associations found over the years in literature. At first sight subjects in our case populations, both microcytic anemias, can sometimes be difficult to distinguish when lacking additional iron status parameters and quantification of hemoglobin A2. For this reason, we reach our final objective that was not only carried out in order to facilitate the screening of microcytic anemia patients in our study, but also in an attempt to elucidate for the first time which of the commonly used discriminant mathematical indices are best suited for a Portuguese population.

Chapter 2 – Materials and Methods

2.1 Population sample

The subjects of this study were divided according to the previously set aims. In order to carry out this project, written informed consent was obtained by clinicians directly from all adult patients and in case of underaged patients' consent, it was obtained from their legal representatives (Figures S.1 to S.4). Prior to the molecular investigation, the blood samples were anonymized, or pseudo-anonymized when there was an intent of giving feedback of any clinically relevant result to the clinician.

For the analysis of individual clinical cases the patients were divided into two groups: i) one for the determination of hemoglobinopathy with a total of 15 patients with atypical phenotypes arriving from "Unidade Laboratorial de Referência do Departamento de Promoção da Saúde e Prevenção de Doenças não Transmissíveis" of INSA; and ii) another group of patients with suspicion of IRIDA with samples from one adult and 10 pediatric patients arriving from "Unidade de Hematologia" of Hospital Dona Estefânia, Lisboa". The second group was admitted for molecular investigation of IRIDA after previous treatment with oral and intravenous iron and exclusion of other causes for microcytic anemia such as hemoglobinopathies, inflammatory disease, gastrointestinal infection, autoimmune disease, and renal or metabolic diseases.

For the population study, the sample was composed of male and non-pregnant female adults (≥ 16 years old) of Portuguese origin and was divided into three groups, according to different inclusion criteria:

- Beta-thalassemia trait (BTT): included 116 subjects (63 males, 53 females) with mean age of 45±18 years (range 18-90) presenting low hemoglobin [Hb < 12g/dL (female) / Hb < 13g/dL (male)], low mean corpuscular volume (MCV < 80 fL), low mean corpuscular hemoglobin (MCH < 27 pg), and increased Hb A2 (> 3.5 %), with molecular confirmation of common pathogenic Portuguese variants in the *HBB* gene: c.92+1G>A; c.92+6T>C; c.92+110G>A or c.1188C>T. The DNA samples of BTT subjects were obtained from previous studies [95], with all the subjects being of southern Portugal origin, due to β-thalassemia trait being more frequent in the southern part of the country [96].
- Iron deficiency anemia (IDA): included 67 subjects (10 males, 57 females) with mean age of 44±16 years (range 16-84) presenting low hemoglobin [Hb <12g/dL (female)/Hb <13g/dL (male)], low mean corpuscular volume (MCV < 80 fL) and high red cell distribution width (RDW > 15 %). Additionally, regarding the iron status, the subjects exhibited low ferritin (Ft < 15μg/L) and/or low transferrin saturation (TSAT < 15 %), which mean respectively depletion of iron stores and impaired iron transport to target tissues. When defining anemia and iron store status recommendations from WHO were taken into consideration [40], [60]. The samples of these individuals also came from "Unidade Laboratorial de Referência do Departamento de Promoção da Saúde e Prevenção de Doenças não Transmissíveis do INSA".
- Controls: included 134 healthy volunteers (42 males, 92 females) with mean age of 39±16 years (range 19-86) presenting hematological and iron status parameters within the normal range.

When defining the case groups, it was made sure that no individual had simultaneously both conditions, with a total of seven cases being excluded. Studied individuals were well characterized for hematological parameters, biochemical parameters regarding the iron status, and partially studied for the genotype of common variants in genes *TMPRSS6* and *HFE* associated with the iron metabolism (Table 2.1).

Table 2.1 – Common genetic variants in genes *TMPRSS6* and *HFE* researched in case and control subjects of the population study.

Genomic context					Alleles		MAF		Aminoacid change
SNP	Chr.	Position	Locus	Strand	WT	Risk	All	EUR	
rs855791	22q12.3	35792882	*TMPRSS6*	Rv	C	T	0.40	0.39	V736A
rs2235324		35799537	*TMPRSS6*	Rv	A	G	0.39	0.43	K253E
rs1799945	6p22.2	26090951	*HFE*	Fw	C	G	0.07	0.17	H63D
rs1800562		26092913	*HFE*	Fw	G	A	0.01	0.04	C282Y

SNP – Single Nucleotide Polymorphism; **Chr. –** Chromosome; **Rv –** Reverse; **Fw –** Forward; **WT** – Wild Type; **MAF** – Minor allele frequency; **EUR** – European. Minor allele frequencies. From Ensembl [97].

Afterwards, when evaluating the performance of mathematical indices for distinguishing microcytic anemias, we investigated from the previously established case groups 111 BTT subjects (60 males, 51 females) and 61 IDA subjects (10 males, 51 females) presenting anemia (Hb < 12g/dL) and microcytosis (MCV < 80fL). Patients with missing values regarding the necessary hematological parameters for the calculation of all the used indices were excluded from this part of the analysis.

Normal range values for hematological and biochemical parameters for adult and pediatric patients are listed in Tables S.2 and S.3, respectively.

2.2 Biological sample

In the "Unidade Laboratorial de Referência do Departamento de Promoção da Saúde e Prevenção de Doenças não Transmissíveis" of INSA, ethylenediamine tetra-acetic acid (EDTA)-anticoagulated blood samples were collected from patients and healthy controls. Then, complete blood count (CBC) was performed by a Coulter LH 750 hematology analyzer (Beckman Coulter, USA) in order to obtain the hematological parameters. Iron status parameters were determined in a Cobas Integra analyser (Roche, Germany). Transferrin saturation was estimated as described in the literature (Equation 2.1) [98].

$$\text{Transferrin Saturation (TSAT) \%} = \left[\frac{\text{Serum iron (Fe) } \mu\text{g/dL}}{\text{Total Iron Binding Capacity (TIBC) } \mu\text{g/dL}}\right] \times 100 \qquad (2.1)$$

In the presence of abnormal parameters in the CBC, as presumptive methods for identification of hemoglobin variants, high performance liquid chromatography (HPLC), in a VARIANT II: β thalassemia method (Bio-Rad, USA), and capillary electrophoresis (CE) at an alkaline pH, performed in Capillarys 2 (Sebia, France) equipment, were carried out. While in HPLC a molecule mixture with a net positive charge is separated into its components by their adsorption onto a negatively charged stationary phase in a chromatography column, followed by their elution by a mobile phase [99], in CE charged molecules are separated at alkaline pH by their electrophoretic mobility, electrolyte pH and electroosmotic flow [100]. In the end, both methods detect molecules based on their retention time. These are complementary methods performed routinely in laboratories and can provide reliable identification of hemoglobin A, A2 and F, as well as many of the common hemoglobin variant [101]. If there was a suspicion of Hb S results were confirmed by the Sickle Solubility Test (ST).

Additionally, if necessary, globin chain analysis was conducted through reversed phase high performance liquid chromatography (RP-HPLC) in an Agilent 1110 Serie Liquid Chromatograph (Agilent Technologies, Germany). For this process, globin chains are denatured and separated

from the heme groups for purposes such as the determination of Hb F composition ($^{G}\gamma$:$^{A}\gamma$ ratio) [102].

2.3 DNA extraction, quantification and quality

At "Unidade de Genética Molecular" of INSA, genomic DNA (gDNA) was automatically extracted from peripheral blood leukocytes using MagNA Pure LC (Roche, Germany) and stored at 4°C, in the Hemoglobinopathies' research laboratory.

DNA quantity and quality were assessed using a NanoDrop One (Thermo Fisher Scientific, USA) spectrophotometer, determining the concentration of DNA (ng/µL) at 260 nm, looking for a 260/280 and a 260/230 absorbance ratio of 1.8-2.0, to rule out protein, salts and residual phenol from nucleic acid extraction. If the A260/280 and A260/230 ratios were off a phenol-chloroform purification was performed [103]. Elution of purified DNA was in 50 µL of MagNA Pure LC DNA Isolation Kit I Elution buffer. Additionally, in some cases, a 0.5 % agarose gel electrophoresis was carried out to look for evidence of substantial band shearing and dilutions of DNA samples were performed according to the criteria of some molecular methods (Table 2.2).

Table 2.2 – Molecular methodologies applied according to the category of the study approach and gene under study.

Category	Molecular Methodology	Gene
Clinical Sequencing Exploratory Research	PCR + Sanger sequencing	*HBB*
		HBD
		HBG1
		HBG2
		HBA2
	Long-PCR + NGS	*TMPRSS6*
Direct detection of known SNPs	ARMS-PCR	*TMPRSS6*
		HFE
	RFLP-PCR	*TMPRSS6*
Duplication/deletion detection	Gap-PCR	*HBA1/HBA2*
	MLPA	β-globin gene cluster

SNP – Single Nucleotide Polymorphism; **PCR** – Polymerase Chain Reaction; **NGS –** Next Generation Sequencing; **ARMS** – Amplification Refractory Mutation System; **RFLP** – Restriction Fragment Length Polymorphism; **MLPA** – Multiplex Ligation-dependent Probe Amplification

2.4 Genomic DNA amplification through Polymerase Chain Reaction

In the 1980s, Kary Mullis created a technique called Polymerase Chain Reaction (PCR) that allowed the user to draw from a complex DNA template a specific sequence and amplify it [104]. The method consists of putting through repetitive cycles of denaturation (94°C-98°C), hybridization (50°C-65°C), and polymerase extension (72°C) an adjusted mixture of the DNA sample with target sequence; a set of oligonucleotide primers (usually 18-25 bp) that will hybridize to opposite template DNA strands and are complementary to the beginning and end of the DNA fragment to be amplified; a *Taq* DNA polymerase, which catalyzes 5′ to 3′ synthesis of DNA; the four 2'-deoxynucleotide triphosphates (dNTPs) required for DNA synthesis – Adenosine (A), Thymine (T), Guanine (G) and Cytosine (C), and Buffer to provide a fitting chemical environment for the polymerase [104], [105].

The procedure occurs in a thermal cycler that heats and cools the reaction tubes to obtain the necessary temperatures for each step of the reaction, resulting in the duplication of the amount of

the target fragment with each cycle, therefore increasing its concentration exponentially, this product can be identified by its size using agarose gel electrophoresis [104], [105].

Agarose gel electrophoresis is a three-stage process: (1) preparation of a gel with the appropriate agarose concentration for the size of DNA fragments to be separate, electrophoresis buffer and ethidium bromide (EtBr); (2) loading the DNA samples with a loading buffer into the sample wells and run the gel at a voltage and for a period that will achieve optimal separation; and (3) visualization of the gel directly upon a ultra-violet (UV) light. This process can separate DNA fragments from approximately 0.5 to 25 kb [106], the observed mobility within the gel's pores is dependent on the molecular mass of the molecules and the electric field applied to the gel since DNA is a highly negatively charged molecule [107].

The ability of PCR to amplify small quantities of DNA enables it to be visualized and subjected to further genetic analysis. This has made the method one of the most important diagnostic techniques in the modern molecular laboratory, being by itself in some cases enough to make the diagnosis of insertions, deletions and point mutations associated with genetic diseases [108], [109].

To increase the sensitivity of the conventional PCR, there have been throughout time several modifications of this technique that accommodate it to different aims [109]. Such modifications include gap-PCR, long-PCR and amplification refractory mutation system (ARMS)-PCR that were performed in this project, besides the use of conventional PCR, which was used for amplification of globin genes.

Gap-PCR: This PCR detects deletion that might be missed by DNA sequencing. The system uses a specific primer set that flanks a known deletion and is based on their inability to generate a PCR product unless a deletion joins the flanking sequences together [110]. When detecting alpha-thalassemia silent carriers, this method presents itself as a better screening alternative for screening than hematological methods [111]. In this work, gap-PCR was used for detection of the $-\alpha^{3.7kb}$ alpha-thalassemia deletion, as described in the literature [112] and schematized and interpreted in Figure 2.1.

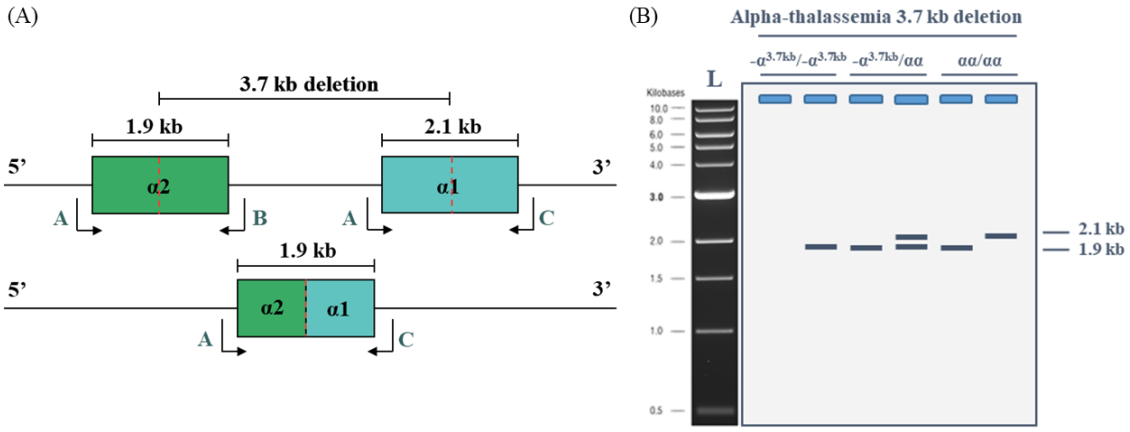


Figure 2.1 – Gap-PCR analysis for diagnosis of $-\alpha^{3.7}$ kb deletion (α-thalassemia). (A) Schematic representation of the breakpoints in genes *HBA2* (α2) and *HBA1* (α1) as dashed lines; the resulting hybrid gene; the location of the used primers (A, B and C); and the sizes of the three resulting fragments**. (B)** Interpretation of gap-PCR results according to the subject genotype. For estimating the size of the expected fragments, a 1Kb DNA Ladder (New England BioLabs, USA). Original figure.

Long-PCR: Conventional PCR has constraints regarding the maximum size of the amplified fragments. This happens because of the rate of errors of misincorporation of *Taq* DNA polymerase. With longer PCR products the chances of misincorporation increase, therefore leading to the delay of strand extension and *Taq* DNA polymerase dissociation from the template strand [113]. To go around this issue, the long-PCR or long-range PCR was created, which can amplify long DNA fragments (3-15 kb). This technique combines the action of two polymerases, a nonproofreading polymerase at high concentration and a proofreading polymerase at a lower concentration. This allows for the study of repetitive and large intronic regions, that are of difficult bioinformatic assemble as well as allowing these fragments to be easily sequenced by Next-Generation Sequencing [114]. The *TMPRSS6* gene was divided into three fragments and amplified with this method, after optimization, using the primers described by Kloss-Brandstätter *et al.* 2012 [115].

ARMS-PCR: The amplification refractory mutation system PCR, also named allele-specific PCR, is a rapid and reliable method, which allows not only SNP detection but also rapid genotyping. In this method, SNP detection is done with two complementary reactions, one uses a pair of primers that are specific to the mutant sequence and the other a pair specific to the wild type. The wild-type set is unable to amplify the mutant sequence and the same happens the other way around, this way extension will only occur if the allele-specific primer is bound to the complementary target sequences. Later on, the genotype of SNPs is determined by analysis of PCR products with gel electrophoresis [110], [116]. Normally, an internal amplification control (IAC) is added. This nontarget DNA sequence present in the sample is coamplified simultaneously with the target sequence, and if there is a negative response (no band in the gel) it will reveal a failure of the PCR [117]. In this dissertation, ARMS-PCR was used to detect the common genetic variant rs855791 [c.2207T>C - p.Val736Ala (V736A)], in gene *TMPRSS6* [118], using exon 2 of the same gene as IAC [115]. As for the common genetic variants of *HFE*, rs1799945 [c.187C>G - p.His63Asp (H63D)], and rs1800562 [c.845G>A - p.Cys282Tyr (C282Y)], a multiplex ARMS-PCR was used, this is a modification that can simultaneously detect both mutations [119]. Due to difficulties in optimizing this process and because this PCR was constructed in a way that, regarding the genotype of the subjects, it will never be expected as a result no band present in the gel, no IAC was introduced. For this reason, exceptionally, in case of uncertainty as to what was the subjects' genotype, Sanger sequencing would be performed. Interpretation of gel electrophoresis for the tested variants is schematized in Figure 2.2 and 2.3.

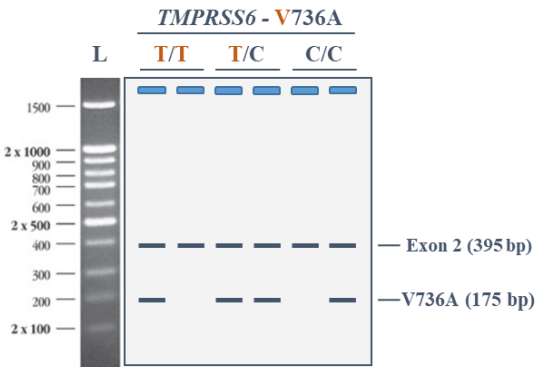


Figure 2.2 – Interpretation of ARMS-PCRs results according to the subject genotype for the *TMPRSS6* common variant V736A. Risk alleles and respective risk amino acid represented in color (orange). For estimating the size of the expected fragments, a DNA Ladder (L) 100 bp plus (AppliChem, USA) was used. Original figure.

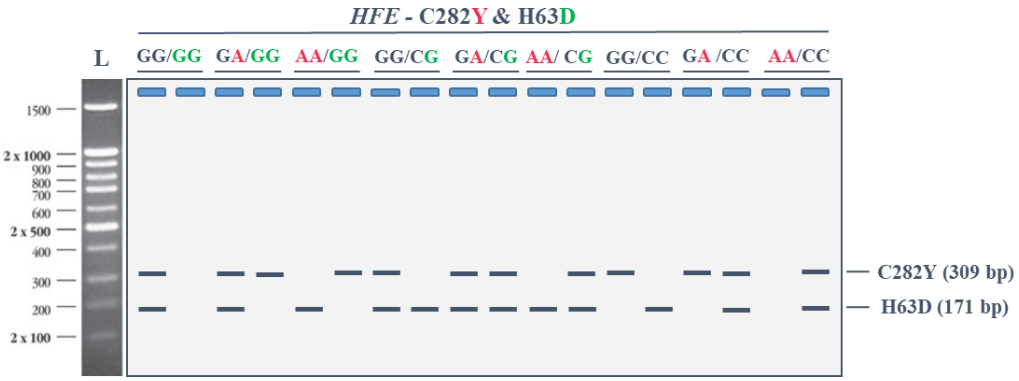


Figure 2.3 – **Interpretation of multiplex ARMS-PCR results according to the subject genotype for the *HFE* common variants C282Y and H63D.** Risk alleles and respective risk amino acid represented in colour (red-C282Y and green-H63D). For estimating the size of the expected fragments, a DNA Ladder (L) 100 bp plus (AppliChem, USA) was used. Original figure.

For every gene and respective variant under study, gDNA was properly amplified in a Thermocycler instrument, T1 or T Gradient (Biometra, Germany), using a specific set of primers and PCR conditions detailed in Supplementary Material, Tables S.4 to S.8. In all PCR, an external negative control was added to account for possible contamination.

PCR results control was performed by agarose gel electrophoresis of the obtained PCR products. The gel was prepared with the appropriate amount of SeaKem® LE Agarose (Lonza, Switzerland) – 0.5, 1, 2, or 3 % (w/v) according to the size of the fragments, in buffer Tris-Borate-EDTA (TBE) 1x containing EtBr (Sigma, USA). PCR products and bromophenol blue loading buffer were loaded into sample wells and standard electrophoresis was carried out in a PowerPac300 or PowerPac Basic (Bio-Rad, USA) electrophoresis system at 70 V for 50-120 min according to the PCR products. Gel revelation was performed in a UV chamber (Uvitec, Cambridge). Conditions were subject to modification if necessary.

Concentration and composition of PCR and loading buffers are described in Supplementary Material's Table S.9 and DNA ladders used for measuring the size of the PCR fragments are in Figure S.5.

2.5 Restriction fragment length polymorphism

When it comes to genetic association studies, one of the most used methods for SNP genotyping is PCR followed by restriction fragment length polymorphism (RFLP). First, PCR is used to amplify a sequence containing the target position, and then the amplification product is cleaved by a restriction endonuclease, which are enzymes that cleave DNA molecules when specific nucleotide sequences (usually 4-6 bp) are recognized. Different nucleotides in the same SNP site will make differ the distance between sites of cleavage of a certain restriction endonuclease, this way SNPs may be differentiated by analysis of patterns derived from the digestion of their amplified DNA when separated by gel electrophoresis [120].

In search of a quick way to genotype the common variant rs2235324 [c.757A>G - p.Lys253Glu (K253E)] of the *TMPRSS6* gene and because we could easily amplify the gene's exon 7, where the SNP is located, we used the online tool NEBcutter version 2.0 (http://tools.neb.com/NEBcutter) [121] to locate restriction enzyme sites within the exon amplified fragment. We looked for a single cutter restriction enzyme that would only digest the fragment into cleaved fragments of different sizes in the presence of the wild-type sequence or the mutated one and found that the enzyme *Ava* I would not digest the DNA fragment unless the

variant was present. This resulted on three different electrophoresis patterns, according to the subject's genotype – AA (1 band pattern, only the 7 exon present), AG (3 band pattern heterozygotes would present the exon 7 fragment plus two other fragments that resulted from the enzyme's activity in the mutated sequence), and GG (2 band pattern composed only by cleaved fragments).

Therefore, we proceed with the use of the PCR-RFLP technique to genotype the sample population for this SNP. The conditions of the PCR and consequent enzymatic restriction are described in Tables S.10 and S.11, its results and their interpretation are schematized in Figure 2.4. Because there was no PCR product purification between amplification and the enzymatic restriction, we optimize the process without using any of the restriction enzyme's accompanying buffer. Before passing to the enzymatic restriction part, the PCR amplification was confirmed with gel electrophoresis.

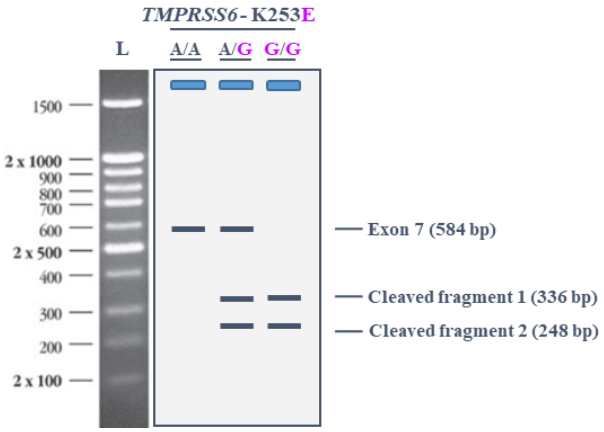


Figure 2.4 – **Interpretation of PCR-RFLP for genotyping the *TMPRSS6* common variant K253E.** Risk alleles and respective risk amino acid represented in color (purple). For estimating the size of the expected fragments, a DNA Ladder (L) 100 bp plus (AppliChem, USA) was used. Original figure.

2.6 Multiplex Ligation-dependent Probe Amplification

Gene deletions or duplications represent about 5 % of all disease-causing mutations, the correct characterization of these mutations is crucial to identify the genotype-phenotype correlation. One approach to detect copy number variation (CNV), defined as a DNA segment, longer than 1 kb, showing a variable copy number compared with a reference genome, is Multiplex Ligation-dependent Probe Amplification (MLPA) [122].

MLPA's principle is based on the amplification of up to 60 probes that each detect a specific DNA sequence (~60 nt), resulting in a set of unique PCR amplicons (64-500 nt) that are separated by CE. After initial denaturation of the sample DNA, a mixture of MLPA probes is added, each probe consists of two oligonucleotides that need to hybridize to the directly adjacent target sequences to be ligated into a single probe. During the following PCR reaction, all ligated probes are amplified at the same time using the same PCR primer pair, resulting in a unique set of PCR amplicons. One PCR primer has a fluorescent label, allowing the amplification products to be visualized during fragment separation on a CE instrument. Fragment separation produces a sample-specific electropherogram. MLPA can only detect relative differences by comparing the MLPA peak patterns of DNA samples, for this purpose inclusion of reference samples in the same run is essential [123].

To detect CNVs in the β-globin gene cluster we used the SALSA® MLPA® probemix P102-C1 HBB kit (MCR Holland, Netherlands) composed of 49 MLPA probes – 39 probes for the β-globin gene cluster and 10 reference probes for detecting autosomal chromosomes (Table S.12). The assay conditions are described in Table S.13 and were performed according to the manufacture instructions, in a Thermocycler instrument T1 (Biometra, Germany) followed by CE in an ABI 3500 genetic analyzer (Applied Biossystems, USA). Data analysis carried out in the software Coffalyser.Net (MCR Holland, Netherlands) taking as reference GenBank chromosomal sequence NG_000007.3. Detected results were confirmed by conventional PCR [124] (Table S.14).

2.7 Automated Sanger Sequencing

DNA sequencing is a post-PCR analysis and a confirmation method for mutations, although automated Sanger sequencing is a first-generation sequencing method it is still used today as the standard for mutation detection [125]. This method synthesizes a complementary DNA template using natural 2'-deoxynucleotides (dNTPs) and termination of synthesis using 2',3'-dideoxynucleotides (ddNTPs) by DNA polymerase. The terminating ddNTPs (ddATP, ddTTP, ddCTP, ddGTP) are labelled with a specific fluorescent dye, with each color representing a different base call. When detected in a CE device the order of the fluorescent fragments reveals the DNA sequence [126].

Automated Sanger sequencing was performed after purification of PCR products with illustra™ ExoProStar™ 1-Step (GE Healthcare, USA) or by column with a JetQuick® DNA Purification kit (GENOMED GmbH, Germany) as instructed by the respective manufacturers. Purified products were sequenced using the BigDye® Terminator v1.1. Cycle Sequencing Kit (Applied Biossystems, USA) as described in Table S.15, in a Biometra® thermocycler. The sequencing reaction was followed by detection of fluorescent signals and strand sizes by CE in an ABI 3500 genetic analyzer (Applied Biossystems, USA).

Results were analyzed using FinchTV 1.5.0 (Geospiza, USA; http://www.geospiza.com) and compared with the RefSeq sequences of the studied genes (Table 2.3), taking into consideration the International Union of Pure and Applied Chemistry (IUPAC) nomenclature (Tables S.16 and S.17) and the genetic code (Figure S.6). Genetic variations spotted in globin genes were cross-referenced with the information present in the HbVar: A Database of Human Hemoglobin Variants and Thalassemias (http://globin.cse.psu.edu/hbvar/menu.html) and ITHANET (https://www.ithanet.eu/) databases.

Table 2.3 – RefSeq sequences used for comparative analysis of Sanger sequencing results.

Gene	Chr.	NG number	NM number	Transcript Size
HBG2	11p15.4	NG_000007.3 (β-globin gene cluster)	NM_000184.2	583 nt
HBG1			NM_000559.2	584 nt
HBD			NM_000519.3	774 nt
HBB			NM_000518.4	626 nt
HBA2	16p13.3	NG_000006.1 (α-globin gene cluster)	NM_000517.4	627 nt
HBA1			NM_000558.4	622 nt

Chr. – Chromosome; **NG** –Reference sequence based on a genomic region; **NM** – Reference sequence based on a protein coding RNA (mRNA); **nt –** nucleotide.

2.8 Next-Generation Sequencing

Recent advances in sequencing technology have enabled faster, cheaper, and more comprehensive screening of entire genomic regions, one of these advancements is Next-Generation sequencing (NGS) [127]. NGS methodology accounts for 4 different stages: (1) sequencing library preparation by "tagmentation", which is the random gDNA fragmentation, followed by 5′and 3′adapter ligation; (2) cluster generation by loading the library into a flow cell where fragments hybridize to its surface and are amplified into a clonal cluster through bridge amplification; (3) Illumina sequencing technology uses a proprietary reversible terminator–based method that detects fluorescent labeled single bases as they are incorporated into DNA template strands; and (4) alignments and data analysis where reads are aligned to a reference sequence with bioinformatics software [128].

For this process to happen, the amplicons from the 3 long-PCR (Table 2.4), which corresponded to the *TMPRSS6* gene sequence (without introns 6 and 10) were mixed into one tube per individual, amounting for a total of 23,518 bp per subject. To ensure that all the fragments were present in the same concentration, the mixture was run through agarose gel electrophoresis. Before this study, five out of the ten studied subjects had the *TMPRSS6* gene analyzed by Sanger sequencing, having been found variants suitable for NGS data validation, this way those individuals acted as positive controls.

Table 2.4 – Amplicons designed for the NGS study of the *TMPRSS6* gene.

Gene	Chr.	F	Localization NC_000022.11	Amplicon NM_153609.3	Region
TMPRSS6	22q12.3	1	g.37095447 - g.37103712	c.-173 (5'UTR) - c.658+104 (Intron 6)	Exon 1 - Exon 6 (8266 bp)
		2	g.37084192 - g.37090022	c.659-240 (Intron 6) - c.1223+103 (Intron 10)	Exon 7 - Exon 10 (5831 bp)
		3	g.37065998 - g.37075418	c.1224-138 (Intron 10) - c.*83 (3'UTR)	Exon 11 - Exon 18 (9421 bp)

Genomic coordinates according to GRCh38.p12. **Chr.** – Chromosome; **F** – Fragment; **NC** – Reference sequence based on a chromosome; **NM** – Reference sequence based on a protein coding RNA (mRNA); **5' UTR** – Five prime untranslated region; **3' UTR –** Three prime untranslated region; **bp** – base pair.

Next-Generation Sequencing analysis was performed in collaboration with "Unidade de Tecnologias e Inovação" of INSA. This process was preceded of amplicon purification with the Agencourt AMPure XP PCR Purification kit (Beckman Coulter, USA; reference B37419AB August 2016) according to manufacturer's instruction, followed by quantification in a Qubit® 3.0 fluorometer (Life Technologies, USA).

2.8.1 Library preparation, cluster generation and sequencing

To prepare the sequencing libraries a Nextera XT DNA Library Prep (Illumina, USA; Part#15031942, v03, February of 2018) was used. All necessary documents for the execution of this kit are provided by the manufacturer and are present in Illumina's Support Resources page (https://support.illumina.com/sequencing/sequencing_kits/nextera_xt_dna_kit/documentation.html). In the last step of library preparation, a PhiX Control v3 (Illumina, USA) was added to the library poll, a known sequence that acts as the sequencing internal quality control.

Before the sequencing step, we followed a bead-based normalization method from the MiSeq System Denature and Dilute Libraries Guide (Illumina, USA; Doc#15039740, v04, April 2018). Libraries were sequenced in a MiSeq equipment (Illumina, USA) using a 0.5 Gb flow cell, by

defining in the samples within the library in the sample sheet, and then loading them into the MiSeq reagent cartridge for automated cluster generation and sequencing. During sequencing, the equipment carried out a primary quality control (QC) analysis.

2.8.2 Data analysis – variant calling

Data analysis of the sequencing results comprehended three steps:

Quality control: Quality of the sequencing data was evaluated using the MultiQC 1.6.dev0 software, which provides an overview of the sequencing quality helping us spot failing samples or identify groups of samples behaving irregularly [129]. The Illumina InterOp module v1.0.26-src was run to obtain detailed statistics from the sequencing run, together with the tool FastQC v0.11.5 which assesses the quality of sequencing files [130].

Mapping reads to reference genome: The Human reference genome used was GRCh38 [131] (version GRCh38.p12, primary assembly, unmasked genomic DNA sequences), available in Ensembl [132]. Clean reads (demultiplexed sequences without adapters or indices) were mapped against the reference genome with the *Bowtie 2* v2.3.4 [133] (via Docker image miguelpmachado/bowtie2:2.3.4.3-02). The alignment was performed using the "very-sensitive-local" mode, for 1 maximum number of valid alignments per read, a maximum size of insertion for paired-end alignments of 1500 and deterministic behavior.

The mapping files were sorted by genomic coordinates using SAMtools software v1.9 [134] (via Docker image miguelpmachado/samtools:1.9), applying the modules *view* and *sort.* The following steps were executed using the sorted file for each sample. Duplicated reads were identified and marked by the tool GATK v4.1.2.0 [135] module MarkDuplicates (via Docker image broadinstitute/gatk:4.1.2.0).

Different statistics were calculated using the modules stats and depth from SAMtools. For each target region, the average depth of coverage and percentage of reference sequence covered with a minimum of 10 reads was assessed.

Variant calling: Base call quality values were corrected for systematic error with the software GATK (tools BaseRecalibrator and ApplyRecalibration) using a known germline variants database downloaded from Ensembl Release 96 (after removing "0", ambiguous and repeated alternative alleles, as well as entries without alternative alleles). The resulting alignment file was used for determination of variants.

Varying positions compared to the reference sequence were initially called for each sample using the tool GATK HaplotypeCaller in Genomic Variant Call Format (GVCF) mode. StandardAnnotation, StandardHCAnnotation and AS_StandardAnnotation annotation groups were added, using the information of germline variants present in Ensembl Release 96 database. Joint genotyping was performed on the samples using the tool GATK Genotype GVCFs for the variants found by HaplotypeCaller, but only for those with minimum call confidence of 20.

Varying positions were filtered for a variant quality (QUAL) < 130 (Phred scale) and sample genotypes were only considered when there was a minimum read depth (DP) 10x and genotype quality (GQ) < 90. In the end, after manual inspection of the results, DP was increased to 20x. In order to accept genotyped heterozygous positions, the allelic depth reads (AD) was also taken into consideration to account for allelic unbalance: genotypes with an allelic balance below 30 % were excluded ($0.3 \leq$ accepted allelic balance ≤ 70).

Additionally, manual depth of coverage analysis was performed resorting to the software Integrative Genomics Viewer v2.3.86 (http://software.broadinstitute.org/software/igv/home) [136]. Afterwards, regions with a high density of variants that turned out to be repetitive regions, and because of that more prone to polymerase stuttering, had their called variants also excluded. For the determination of these repetitive patterns, the online software Tandem Repeats Finder v4.09 (https://tandem.bu.edu/trf/trf.html) [137] was used.

Confirmation of called coding variants was done by automated Sanger sequencing of *TMPRSS6* exons containing such variants. For each confirmed variant, three subject's samples were sequenced, one for each genotype. PCR conditions for amplification of the exons' sequences are described in Tables S.10 and S.18 [115].

2.9 Variant effect prediction

Rare diseases can be difficult to diagnose on account of the low incidence and incomplete knowledge of the implicated alleles. Regardless of this fact, variant analysis of sequencing data can lead to the discovery of underlying genetic mutations as well as bring immediate benefit of diagnosis resulting in a more accurate prognosis, removing the burden of additional clinical investigations [138].

Called variants were crossed-referenced with information in online databases: Ensembl (https://www.ensembl.org/index.html); PubMed (https://www.ncbi.nlm.nih.gov/pubmed/); ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/); SNPedia (https://www.snpedia.com/); UniProt (https://www.uniprot.org/); and Protein Data Bank (https://www.rcsb.org/). Afterwards, we undertook *in silico* analysis of the called variants' effect in terms of splicing and protein changes.

When performing these analyses, we used as reference for the *TMPRSS6* gene the Ensembl sequence ENSG00000187045, the wild-type transcript sequence ENST00000346753.8 and the correspondent protein sequence following UniProtKB identifier Q8IU80.

2.9.1 Prediction of splice-affecting variants

Mammalian genes are composed of several relatively short exons interrupted by longer introns. To generate correct, mature mRNAs, these exons must be identified and joined in a process called splicing. Failure to recognise exon-intron boundaries or failing to remove an intron generates aberrant mRNAs that are either unstable or will translate into defective or deleterious protein isoforms [139].

A single SNP may lead to changes in the splicing. In this way, mutations responsible for modifications to this process account to up to 50 % of all mutations that lead to gene dysfunction. Therefore, and because these mutations can be disease-causing, the software Human Splicing Finder v3.1 (http://www.umd.be/HSF/) [140] was used to forecast the splicing impact of the called variants in the *TMPRSS6* gene of patients with suspicion of IRIDA.

2.9.2 Prediction of protein changes

Non-synonymous SNPs may introduce variation in the amino acid sequence of the translated protein. Incorrect protein function has a major influence on human health, and, for this reason, it becomes important to study of missense variants' impact in the final peptide. This can be done by studying the protein sequence alignments and its three-dimensional structure [141].

For us to study the impact of the found variants in the subjects' matriptase-2, we used the online software PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) [142] and Sorting Intolerant

From Tolerant (SIFT; https://sift.bii.a-star.edu.sg/) [143]. The first one predicts the possible impact of amino acid substitutions on the stability and function of human proteins using structural and comparative evolutionary considerations, while the second one presumes an important amino acid will be conserved in the protein family, and so changes at well-conserved positions tend to be predicted as deleterious [142], [143]. Additionally, to study the effects of missense variants on the 3D structure of matriptase-2, and since there was no structure already available in the databases, the protein structure was modeled in the online software Phyre2 v2.0 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) [144], followed by analysis on the software Missense3D (http://www.sbg.bio.ic.ac.uk/~missense3d/) [145].

2.10 Statistical analysis

2.10.1 Genetic association study

A database was constructed using the information regarding the three study groups: Control, BTT and IDA, including hematological parameters, iron status information and all the genetic variants analysed (*HBB*, *TMPRSS6* and *HFE* genes). So that we could better understand the data, additional information was added regarding the classification of the anemia (mild, moderate or severe) and the extent of iron stores (iron deficiency, normal or iron overload). Statistical analysis was performed using R v3.6.1 software (http://www.R-project.org) [146] after data clean up and preparation in Excel 2013 spreadsheets. List of database variables and their details is provided in Supplementary Materials Table S.19.

Data have been tested for normality using the Shapiro-Wilk test and non-parametric statistics were used when such condition was not met. Fisher's exact tests were used to test the allelic and genotypic associations of all the SNPs and Hardy–Weinberg equilibrium (HWE) of the genotypic frequencies among controls and cases (BTT and IDA) was examined using a chi-square test ($\chi 2$) for each group, using the R package 'gap'. The strength of the association was estimated by an odds ratio (OR) of risk and 95 % confidence intervals (CI), adjusted for sex and age, for the studied SNPs homozygosis, heterozygosis and wild type, using the R package 'questionr'. The adjusted variables were chosen based on the knowledge that the used hematological and iron status measurements differ for sex and age [39], [147].

The Mann-Whitney test was used to compare a continuous variable according to two categories. Analysis of a continuous variable according to more than two categories was performed using the Kruskal-Wallis test. Nominal variables were analyzed using Fisher's exact test. Regression analysis was performed separately, with a single predictor variable, using genotype as the dependent and each of the hematological and iron status variables as the independent, while being adjusted for sex and age, as well as for the thalassemic trait when analyzing association within the BTT group. Regressions were not performed if sample < 30 was not verified. A p-value < 0.05 (two-tailed) was considered statistically significant after the False Discovery Rate (FDR) method.

2.10.2 Microcytic anemia differentiation

The hematological parameters and iron status information of patients with microcytic anemia from groups BTT and IDA were compared with a Mann-Whitney test. Then, to differentiate them we evaluated the diagnostic performance of 13 distinct indices: RBC, England and Fraser (E&F), Mentzer, Srivastava, Shine and Lal (S&L), Bessman (RDW), Ricerca, Jayabose (RDWI; Red cell distribution width index), Green and King (G&K), Mean Density of Hemoglobin per Liter (MDHL), Mean Cell Hemoglobin Density (MCHD), Sirdah and Ehsani. These were chosen for

being the most cited in the literature and because they could be calculated with the available data. Mathematical formulas and cutoffs were those set by authors in original published reports (Table 2.5).

Table 2.5 – Discriminant indices used for distinguishing patients with microcytic anemia.

Hematological Index	Formula	In favor of BTT	In favor of IDA	Reference
RBC	RBC	> 5	< 5	[148]
England and Fraser (E&F)	MCV-RBC-(5Hb)-3.4	< 0	> 0	[148]
Mentzer	MCV/RBC	< 13	> 13	[149]
Srivastava	MCH/RBC	< 3.8	> 3.8	[150]
Shine and Lal (S&L)	MCV^2xMCH/100	< 1530	> 1530	[151]
Bessman (RDW)	RDW	< 15	> 15	[152]
Ricerca	RDW/RBC	< 4.4	> 4.4	[153]
Jayabose (RDWI)	MCVxRDW/RBC	< 220	> 220	[154]
Green and King (G&K)	MCV^2 x RDW/HBx100	< 65	> 65	[155]
MDHL	(MCH/MCV)xRBC	> 1.63	< 1.63	[156]
MCHD	MCH/MCV	> 0.3045	< 0.3045	[156]
Sirdah	MCV-RBC-(3Hb)	< 27	> 27	[157]
Ehsani	MCV-(10RBC)	< 15	> 15	[158]

RBC – Red blood cells; **RDW –** Red cell distribution width**; RDWI –** Red cell distribution width index; **MDHL** – Mean Density of Hemoglobin per Liter; **MCHD** – Mean Cell Hemoglobin; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean.

To determine the performance of the indices, the following confusion matrix-based measurements were calculated in Excel 2013: Accuracy (Efficiency), Sensitivity (SENS), Specificity (SPEC), Positive predictive values (PPV), Negative predictive values (NPV) and Youden's index (YI) (Equations 2.2 to 2.7) [159], [160]. Receiver operating characteristic (ROC) curve analysis was used to illustrate the diagnostic performance, using the R package 'Proc' [161]. An Area Under the Curve (AUC) value < 0.75 means that the test shows deficiencies in its diagnostic accuracy, but due to the high values of AUC obtained from the ROC analysis, a cut-off of 0.70 was established for the YI in order to determine the best formulas [162]. The best threshold was plotted in the curves with the corresponding 95 % CI computed with 2000 stratified bootstrap replicates.

$$\text{Accuracy (Efficiency)} = \left[\frac{\text{Correctly diagnosed}}{\text{Total}}\right] \tag{2.2}$$

$$\text{Sensitivity (SENS)} = \left[\frac{\text{true positive}}{(\text{true positive} + \text{false negative})}\right] \tag{2.3}$$

$$\text{Specificity (SPEC)} = \left[\frac{\text{true negative}}{(\text{true negative} + \text{false positive})}\right] \tag{2.4}$$

$$\text{Positive predictive values (PPV)} = \left[\frac{\text{true positive}}{(\text{true positive} + \text{false positive})}\right] \tag{2.5}$$

$$\text{Negative predictive values (NPV)} = \left[\frac{\text{true negative}}{(\text{true negative} + \text{false negative})}\right] \tag{2.6}$$

$$\text{Youden's index} = (\text{sensitivity} - \text{specificity}) - 1 \tag{2.7}$$

Chapter 3 – Results and Discussion

3.1 The basis of hemoglobinopathies

In this part of the dissertation, after observing out of range CBC parameters and detecting abnormal results in presumptive methods for the identification of hemoglobin variants, we report a series of 15 patients to whom molecular testing was performed to confirm and characterize their hematological disorder. The findings are summarized in Table 3.1.

Table 3.1 – Summary of the molecular lesions in the origin of hemoglobinopathies in the studied individuals.

Case	Name	Mutation	Mutation, HGVS nomenclature
1,2	Hb Leiden	beta 6(A3) Glu->0 or beta 7(A4) Glu->0	*HBB*:c.22_24delGAG
3	Hb Pôrto-Alegre	beta 9(A6) Ser>Cys	*HBB*:c.29C>G
3	Codon 27 GCC>GCT	rs74296717	*HBB:c.84 C>T*
4	Hb N-Baltimore	beta 95(FG2) Lys>Glu	*HBB*:c.286A>G
4,5	Hb S	beta 6(A3) Glu>Val	*HBB*:c.20A>T
5,6,12	$-\alpha^{3.7\ kb}$ deletion, α^{+}	deletion of 3804 nts from the *HBA*2 gene to *HBA*1 gene	NG_000006.1:g.34164_37967del3804
6	Hb Nigeria	alpha2 81(F2) Ser>Cys	*HBA2*:c.245C>G
7	Cape Verde deletion β^{0}	del of 7719 nts from *HBB* IVSII to downstream of the gene	U01317:g.71551_79269del7719
8	$^{A}\gamma$ IVSII $TG_{(4)}CG_{(5)}$ deletion	rs61080176	NM_000559.2:c.316-302_316-285del18
8	*Xmn*I polymorphism	rs7482144	*HBG2*:c.-158C>T
9	Hb Strasbourg	beta 23(B5) Val>Asp	*HBB*:c.71T>A
10	Hb J-Iran	beta 77(EF1) His>Asp	*HBB*:c.232C>G
11	Hb E, β^{+}	beta 26(B8) Glu>Lys	*HBB:*c.79G>A
12,13	Hb A2' (Hb B2)	delta 16(A13) Gly>Arg	*HBD*:c.49G>C
12,14,15	IVS-I-6; β^{+}	beta nt 148 T>C	*HBB*:c.92+6T>C

3.1.1 Cases 1 and 2: Heterozygous for Hb Leiden

Two apparently unrelated Portuguese cases arrived for hemoglobinopathy evaluation, their hematological findings are in Table 3.2. The first one is a 62 years old female presenting no anemia but low mean corpuscular hemoglobin concentration (MCHC = 31.1 g/dL), a parameter that shows how completely the erythrocyte space is filled with hemoglobin and that is why when low is normally associated with anemia [163], and anisocytosis, meaning elevated variability in RBCs size measured by the parameter RDW (17.2 %). The second case is a 71 years old male presenting with the same hematological findings with additional hypochromia (MCH = 24.7 fL MCHC = 31.8 g/dL; and RDW = 20.1 %).

Table 3.2 – Hematological features for Cases 1 and 2.

Sex	Age	RBC ($x10^{12}$/L)	Hb (g/dL)	Ht (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)
F	62 y	4.29	12.5	40.2	93.5	29.1	**31.1**	**17.2**
M	71 y	5.31	13.1	41.2	77.6	**24.7**	**31.8**	**20.1**

Bold – values outside normal range; **M** – Male; **F** – Female; **y** – Years; **RBC** – Red blood cells; **Hb** – Hemoglobin; **Ht** – Hematocrit; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **MCHC** – Mean corpuscular hemoglobin concentration; **RDW** – Red Cell Distribution Width.

Capillary electrophoresis was performed as an additional presumptive test for hemoglobin disorders. The results revealed similar electrophoretic profile for both cases, with the presence of a Hb variant migrating in "Hb D zone" accounting for 29.7 % and 23.7 % of the total hemoglobin, for cases 1 and 2 respectively, and normal levels of Hb A2 (Figure 3.1 A and B). Additionally, case 2 presented a vestigial amount (0.1 %) of hemoglobin migration in "Hb C zone", this was considered an artefact that posed no challenge to the interpretation of the electrophoretic profile, for this reason, it was disregarded.

Molecular investigation of the atypical phenotype was performed by complete DNA sequencing of the *HBB* gene, using different PCR reactions for each gene segment as previously described. These results from *HBB* sequencing revealed, for both cases, a heterozygous deletion of the nucleotide triplet GAG at codon 6 or 7 (*HBB*:c.22_24delGAG – beta 6(A3) Glu->0 OR beta 7(A4) Glu->0), corresponding to the deletion of a glutamic acid residue (Figure 3.1 C), since these two codons are identical, it cannot be further determined which has been deleted (Figure 3.1 C). This deletion is responsible for the synthesis of an unstable hemoglobin variant named Hb Leiden. The remaining parts of the *HBB* gene did not display any other variants or polymorphisms that could be responsible for the phenotype.

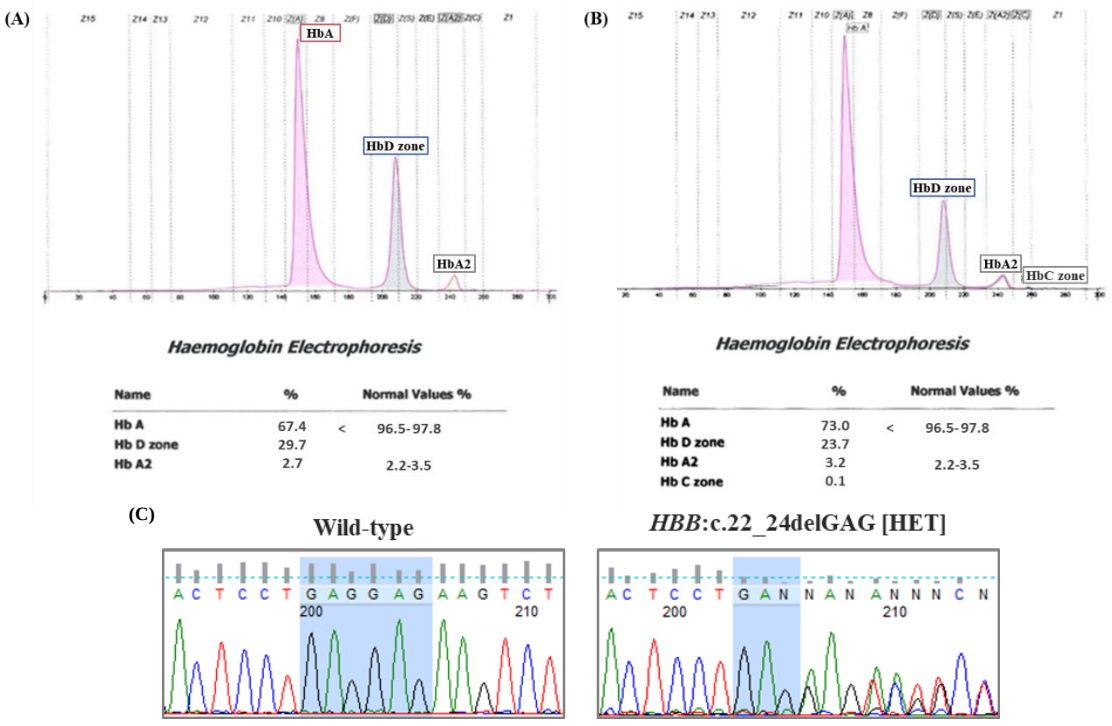


Figure 3.1 – Biochemical characterization and molecular identification of a hemoglobin variant, Hb Leiden. Electrophoretic profile by capillary electrophoresis, **(A)** Case 1 and **(B)** Case 2, showing Hb variant migrating in "Hb D zone" (blue); and **(C)** Partial electropherogram from automated Sanger sequencing of the sense strand of *HBB* gene at exon 1 showing the *HBB*:c22_24delGAG (beta6(A3)Glu>0 or beta7(A4)Glu>0) deletion responsible for the Hb variant. Wild-type nucleotides and mutated positions in an unaffected individual and the affected patient are indicated in a blue box. The figure depicts a heterozygous sample as it is present in both cases 1 and 2.

Originally described by De Jong *et al.* [164], Hb Leiden is an unstable hemoglobin that arises from the deletion of one amino acid near the N-terminus of the β-globin chain. This variant exhibits a slightly higher oxygen affinity than Hb A, with similar oxygen dissociation curves but different partial pressure of oxygen at which the hemoglobin is half saturated (p50) values, and decreased response to organic phosphates, which has been suggested to be a consequence of the

location of the deletion, since the N-terminal region of the β-chains participates in the binding of 2,3-DPG, and may be responsible for the change of oxygen affinity [165]. Generally, in unstable hemoglobin disorders, there is a balanced globin synthesis, which is not the case of Hb Leiden where we can observe in patients an excess of α-chains, yet the excess does not lead to a β-thalassemia like phenotype [166].

Hb Leiden represents 19 % to 31 % of the total hemoglobin and is associated with very mild hemolysis, and normal erythrocyte indices [166]. The deletion in the heterozygous state is defined by mild compensated hemolytic condition with normal Hb concentrations; reticulocytotic (shortened red cell survival - 39 days); anisopoikilocytosis; occasional jaundice; splenomegaly; and occasional hemolytic crises precipitated by drugs and infection [167], [168].

Although patients with this variant can be symptom-free, a precipitating cause for a severe hemolytic episode is always a danger in these patients. For this reason, is recommended education on how to avoid hemolysis provoking agents, such as drugs used to treat other illnesses. In these patients with mild continuous hemolysis, the spleen, even though slightly enlarged, apparently does not play a significant role, since there is no significant increase of RBC sequestration. Thus, there is no evidence that a splenectomy will benefit those patients, while it may be resorted to in a severe hemolytic episode [169]. Furthermore, Hb Leiden patients should have access to genetic counselling since if inherited simultaneously with β-thalassemia leads to chronic severe hemolytic anemia [170]. Also, clinicians should be aware that this variant may result in discordant glycemic profiles between normal glycated hemoglobin (HbA1c) and high blood sugar levels, which can delay diabetes diagnosis and treatment for several years. Treatment and monitoring options for this patient's diabetes and diabetes-related comorbid conditions will require careful management to avoid preventable complications [167], [171].

Finally, since this variant has been previously reported in several countries, the Netherlands [164], [165], [168], China [170], South Africa [169], Yugoslavia [172], and Thailand [173], and is believed to have arisen independently, there is a possibility that the two Portuguese subjects are related. For this reason and because these patients may have clinical complications, a family study is recommended.

3.1.2 Case 3: Heterozygous for Hb Pôrto-Alegre

A 26-year-old healthy female, with all hematological parameters within the normal range (Table 3.3), arrived for investigation of hemoglobin variant after presenting an abnormal electrophoretic profile by CE.

Table 3.3 – Hematological features for Case 3.

Sex	Age	RBC ($x10^{12}$/L)	Hb (g/dL)	Ht (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)
F	26 y	4.75	13.2	39.5	83.2	27.8	33.4	12.2

Bold – values outside normal range; **F** – Female; **y** – Years; **RBC** – Red blood cells; **Hb** – Hemoglobin; **Ht** – Hematocrit; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **MCHC** – Mean corpuscular hemoglobin concentration; **RDW** – Red Cell Distribution Width.

The subject presented an electrophoretic profile showing elevated Hb F (19.8 %), which can indicate the presence of a hemoglobin variant that migrates close to Hb F and may not be separated from it (Figure 3.2 A), as well as increased Hb A2 (4.2 %). The subject's profile also exhibited some hemoglobin migration in "Z12 zone", this was considered an artefact that posed no challenge to the interpretation of the electrophoretic profile, for this reason, it was disregarded.

In order to molecularly investigate the presence of a hemoglobin variant the *HBB* gene was sequenced. DNA sequencing confirmed the presence of a mutation in codon 9 leading to the substitution of serine for cysteine (*HBB*:c.29C>G – beta 9(A6) Ser>Cys) and resulting in a rare hemoglobin variant called Hb Pôrto-Alegre (PA) that it is also in *cis* with the intragenic polymorphism rs74296717 (codon 27 GCC>GCT – *HBB*:c.84 C>T), both mutations are present in the first exon and are in the heterozygous state (Figure 3.2 B).

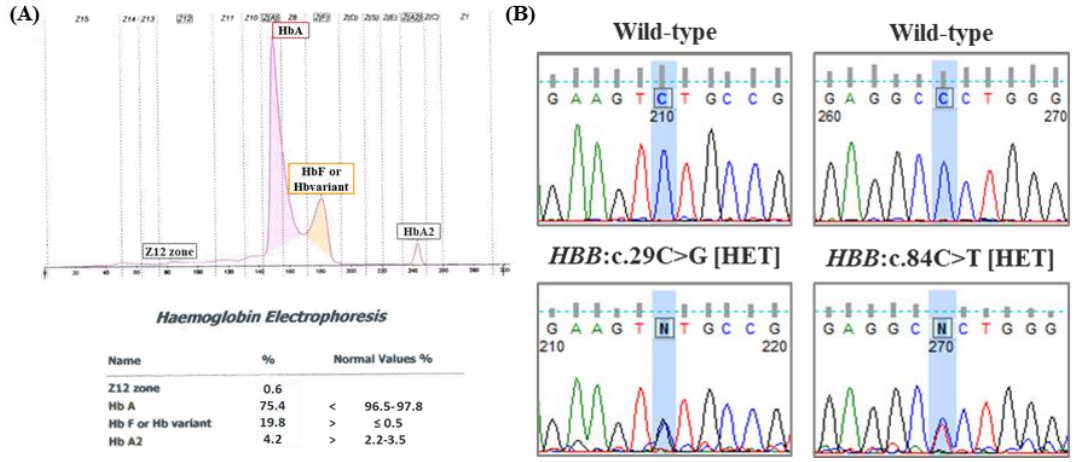


Figure 3.2 – Biochemical characterization and molecular identification of hemoglobin variant, Hb Pôrto-Alegre. **(A)** Electrophoretic profile by capillary electrophoresis showing Hb variant migrating in "Hb F zone" (yellow); **(B)** Partial electropherogram from automated Sanger sequencing of the sense strand of *HBB* gene at exon 1 showing the *HBB*:c.29C>G (beta 9(A6) Ser>Cys) the mutation responsible for the Hb variant (left) and the linked codon 27 polymorphism *HBB*:c.84 C>T - rs74296717 (right). Wild-type nucleotides and mutated positions in an unaffected individual and the affected patient are indicated in a blue box.

The hemoglobin variant being studied was first described in 1963 by Tondo *et al*. [174] and was later observed in members of several other families in Brazil [175], [176], Cuba [177] and the Canary Islands [178]. The presence of the intragenic polymorphism in codon 27 is suggestive that this mutation originated from a single mutational event in the Portuguese population and was then spread to South America, namely to Brazil [179].

Hb PA is a rare hemoglobin in which the mutation induces, during storage of the hemolysate, polymerization by forming intermolecular disulfide bridges (S-S) via the extra cysteine residue, that carries an extra thiol group (-SH) oriented towards the exterior of the Hb molecule (Figure 3.3) [180]. The normal electrophoretic mobility of a fresh, non-oxidized, hemolysate of homozygous patients indicates that the mutation does not alter the net charge of the molecule and that *in vivo* the molecule exists as a tetramer [181]. Maintenance of the tetramer form *in vivo* is suggested to be caused by the activity of glutathione reductase activity, which keeps the reducing environment preventing the formation of disulfide bridges [180]. This theory is supported by the fact that the enzyme's activity was increased in the RBCs of homozygous Hb PA patients compared to normal subjects and that their erythrocytes contained twice the amount of reduced glutathione [182]. The polymers Hb PA do not change significantly O2 binding, this type of hemoglobins that polymerize without changing oxygenation, occurs frequently in amphibians and reptiles as a response of hemolytic events [183], [184].

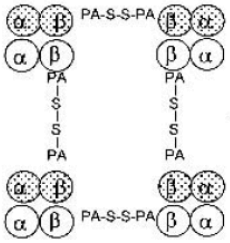


Figure 3.3 – Hemoglobin Pôrto-Alegre oligomers (T_4). *In vitro* polymerization Hb PA by spontaneous formation of intermolecular disulfide bridges via the extra cysteine residue extra thiol group (-SH) oriented towards the exterior of the molecule**.** Adapted from Baudin-Creuza *et al*. [180].

Individuals heterozygous or homozygous for Hb PA remain asymptomatic without clinical or hematological features [179], without any evidence of hemolytic disease, as is the case of our subject that has its electrophoretic profile justified by the presence of the variant. Hb A2 normally constitutes less than 3 % of the total hemoglobin in adults, while it has almost no physiological importance, the determination of Hb A2 is an important tool to diagnose of BTT [185]. Since the subject does not present any clinical features and complete sequencing of the *HBB* gene ruled out any β-thalassemia mutations, there is no need to further investigate the increased levels of Hb A2

3.1.3 Case 4: Compound heterozygous for Hb S/N-Baltimore

An eleven-month-old female baby with values within the normal range for her age group arrived for hemoglobin variant analysis since her father was known for being heterozygous for the Hb variant N-Baltimore (Table 3.4).

Table 3.4 – Hematological features for Case 4.

Sex	Age	RBC ($x10^{12}$/L)	Hb (g/dL)	Ht (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)
F	11 m	4.2	11.4	34.6	86.5	27.2	33.0	12.5

Bold – values outside normal range; **F** – Female; **m** – Months; **RBC** – Red blood cells; **Hb** – Hemoglobin; **Ht** – Hematocrit; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **MCHC** – Mean corpuscular hemoglobin concentration; **RDW** – Red Cell Distribution Width.

Apart from the complete blood count, as an additional presumptive test, HPLC was performed with elution chromatogram profile exposing a Hb variant migrating in "S-window" and elevated Hb F levels (15.0 %) (Figure 3.4 A). Usually, Hb F disappears from the red blood cells of infants after about 6 months [186]. Thereafter, molecular confirmation of both variants, Hb S and N- Baltimore, was carried out by automated Sanger sequencing of the amplified fragments of the complete *HBB* gene. Results revealed two substitutions in heterozygosity at exon 1 and 2, the mutation responsible for the Hb S (*HBB*:c.20A>T – beta 6(A3) Glu>Val) and the mutation responsible for Hb N-Baltimore (*HBB*:c.286A>G – beta 95(FG2) Lys>Glu), corroborating the suspicions raised by the presumptive tests (Figure 3.4 B).

Sickle cell trait (SCT) is an inherited condition in which both normal hemoglobin and sickle hemoglobin are produced in the RBCs, with subjects generally leading a healthy life [187].

Hb N-Baltimore was first described in an American Black family by Clegg *et al*. [188]. Since then, this hemoglobin has also been called Hb Jenkins, Hopkins-1, N-Memphis, and Kenwood [189]. Subjects with this variant present with normal hematological findings and have been previously reported in various ethnic groups [190]. One characteristic of this variant is that its concentration in the peripheral blood of simple heterozygotes is the same as hemoglobin A. This occurs because the α-globin chains have a relatively positive surface charge and so interact more

readily with relatively negatively charged β-globin variants to form αβ-dimers, as is the case of variant Hb N-Baltimore [189], [191]. In compound heterozygous, this is reflected in the higher percentage of the negatively charged β-globin variant Hb N-Baltimore, which is found in approximately 50 % in heterozygotes, when compared to β-globin variants with a positive surface charge, such as Hb S or Hb C, whose quantity in the heterozygote is 40-45 % [191].

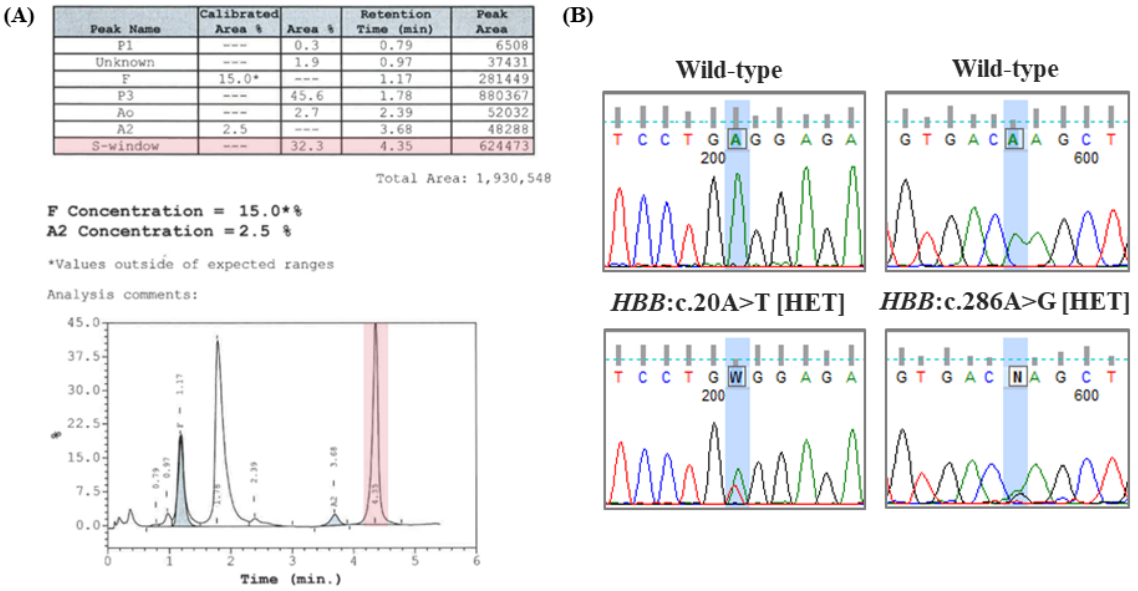


Figure 3.4 – Biochemical characterization and molecular identification of two hemoglobin variants present in compound heterozygoty, Hb S/N-Baltimore. (A) Elution chromatogram profile by HPLC showing Hb variant migrating in "S-window" (pink). **(B)** Partial electropherogram from automated Sanger sequencing of the sense strand of *HBB* gene at exon 1 and 2 showing the *HBB*:c.20A>T (beta 6(A3) Glu>Val) the mutation responsible for the Hb S (left) and the *HBB*:c.286A>G (beta 95(FG2) Lys>Glu) the mutation responsible for Hb N-Baltimore (right). Wild-type nucleotides and mutated positions in an unaffected individual and the affected patient are indicated in a blue box.

Hb F is the major genetic modulator of the hematologic and clinical features of sickle cell disease, an effect mediated by its exclusion from the sickle hemoglobin polymer. Fetal hemoglobin levels are inherited as a quantitative genetic trait and its distribution among sickle erythrocytes is highly variable, going from 1% to 30%, normally associated with different haplotypes within the β-globin gene cluster that is responsible for HPFH [192], [193].

In the present case, the found compound heterozygous for Hb S/N-Baltimore does not appear to be behind any present or future clinical complications, since both variant traits are asymptomatic, especially in the presence of HPFH. In the future, additional molecular tests can be performed to determine the direct cause of HPFH. Furthermore, it is important for the clinician to know that Hb N-Baltimore has been known to lead to a mistaken diagnosis of diabetes for producing a spuriously elevated HbA1c level [194].

3.1.4 Case 5: A sickle cell anemia patient with coinheritance of alpha-thalassemia

A 36-year-old male, known to have sickle cell disease (SCD), arrived presenting with microcytic (MCV = 72.0 fL) hypochromic (MCH = 21.3 pg) anemia (Hb = 10.9 g/dL), low MCHC (30.3 g/dL), elevated RDW (17.4 %) and reticulocytosis (2-3%), (Table 3.5). Additionally, the patient's blood smear did not display any drepanocytes, which was not expected in a sickle cell disease patient.

Table 3.5 – Hematological features for Case 5.

Sex	Age	RBC ($x10^{12}$/L)	Hb (g/dL)	Ht (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)	Retic (%)
M	36 y	5.0	**10.9**	**36.0**	**72.0**	**21.3**	**30.3**	**17.4**	**2-3**

Bold – values outside normal range; **M** – Male; **y** – Years; **RBC** – Red blood cells; **Hb** – Hemoglobin; **Ht** – Hematocrit; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **MCHC** – Mean corpuscular hemoglobin concentration; **RDW** – Red Cell Distribution Width; **Retic** – Reticulocytes.

The presented phenotype is milder than what was expected for a SCD patient. To understand what was behind this phenotype, hemoglobin analysis was performed by HPLC revealing the presence of hemoglobin migrating in the "S-window", as expected, levels of Hb F (1.9 %), which are not high enough to justify such SCD phenotype modelling, and elevated Hb A2 (6.3 %) that with this clinical presentation could indicate the presence of a β-thalassemia mutation (Figure 3.5 A). Furthermore, globin chain analysis was performed by RP-HPLC but exhibited no changes in globin chain ratios (Figure 3.5 B).

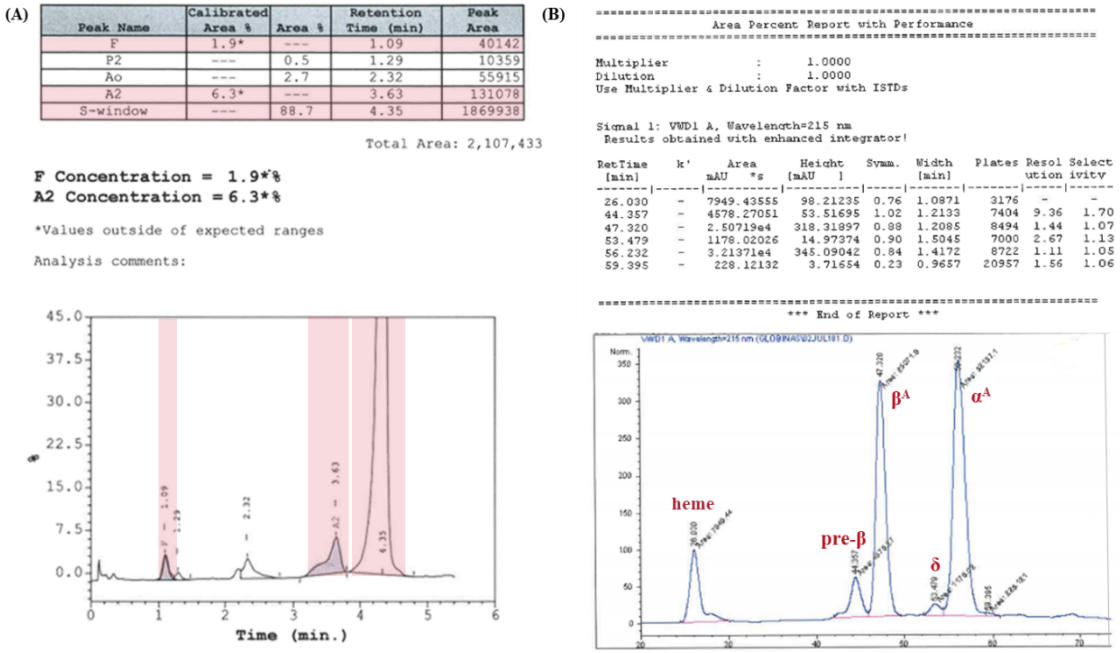


Figure 3.5 – Biochemical characterization of the hemoglobin variant Hb SS. Representative chromatograms of **(A)** elution profile by HPLC showing decreased Hb F, elevated Hb A2 and Hb variant migrating in "S-window" (pink); and **(B)** globin chain analysis by RP-HPLC revealing no apparent change in globin ratios.

Molecular analysis of the *HBB* gene was performed in search for thalassemic mutations or the presence of a Hb variant. Sequencing results only confirmed the presence of a homozygous mutation in codon 6 (*HBB*:c.20A>T – beta 6(A3) Glu>Val) that leads to the substitution of glutamic acid by valine and is responsible for Hb SS (Figure 3.6 A). Co-inheritance of α-thalassemia has been associated with a milder phenotype in SCD patients [195]. One prevalent deletion causing α-thalassemia removes about 3.7 kb of DNA, fusing the *HBA2* and the *HBA1* genes, resulting in the formation of a single α2α1 gene [191]. A gap-PCR was carried out to rapidly detect the presence of the $-\alpha^{3.7kb}$ alpha-thalassemic deletion. Direct detection by visualization of the gap-PCR product by gel electrophoresis revealed that the subject was homozygous for the studied deletion (Figure 3.6 B). The presence of homozygous α^{+}-thalassemia

trait (-$\alpha^{3.7}$/ -$\alpha^{3.7}$) is associated with microcytic erythrocytes and sometimes with mild anemia [191], which goes in line with the hematological features of the patient.

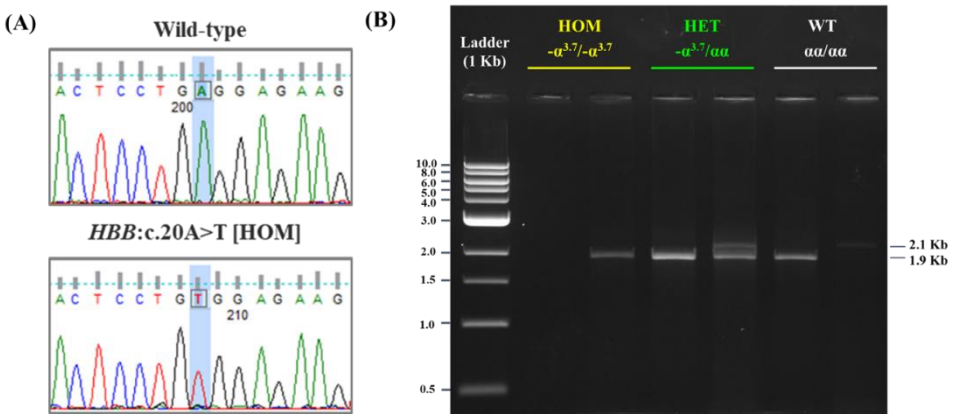


Figure 3.6 – Molecular identification of a sickle cell anemia patient homozygous for the alpha thalassemia -$\alpha^{3.7}$ deletion. (A) Partial electropherogram from automated Sanger sequencing of the sense strand of *HBB* gene at exon 1 showing the homozygous Hb S variant *HBB*:c.20A>T (beta 6(A3) Glu>Val). Wild-type nucleotides and mutated positions in an unaffected individual and the affected patient are indicated in a blue box. **(B)** Direct detection of alpha thalassemia deletion -$\alpha^{3.7}$ by visualization of the gap-PCR product by gel electrophoresis. The subject was homozygous for the studied deletion (yellow). For measuring the size of the expected fragments, a 1Kb DNA Ladder (New England BioLabs, USA) was used.

SCD is one of the most common severe monogenic disorders in the world. In deoxygenating or dehydrating conditions, Hb S polymerizes within the erythrocytes, leading to intracellular tactoids that deform the RBCs into the characteristic sickle shape. Hemoglobin polymerization leads to erythrocyte rigidity that induces microvascular obstruction, abnormal adhesion of leukocytes and platelets (Figure 3.7). Vaso-occlusion is central to the pathophysiology of this disease, although the importance of chronic anemia, hemolysis, and vasculopathy has been established [193], [195].

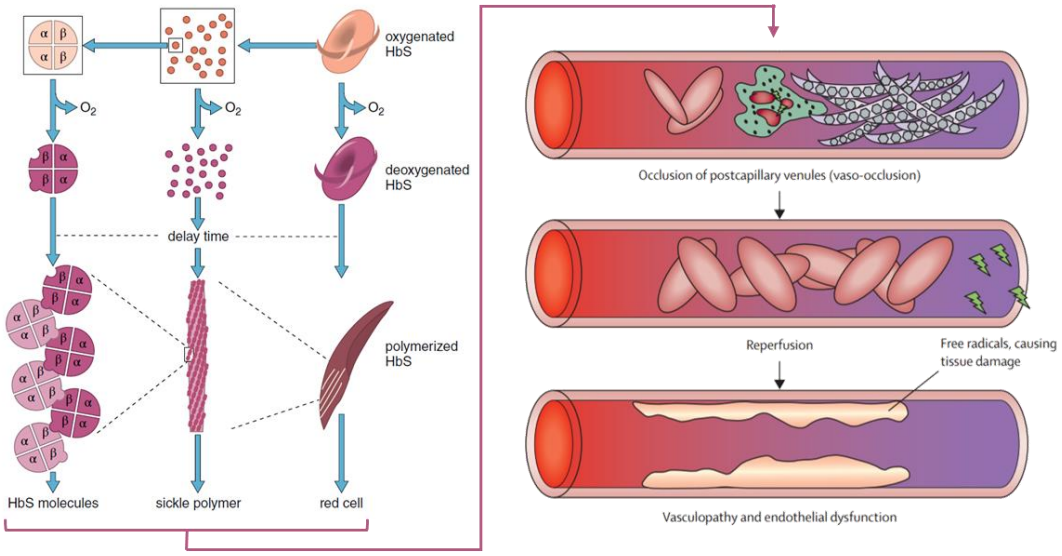


Figure 3.7 – Pathophysiology of sickle cell disease. The roles of Hb S polymerization, hyperviscosity, vaso-occlusion, hemolysis, and endothelial dysfunction are shown. Deoxygenation causes HbS to polymerase, leading to sickled erythrocytes. Vaso-occlusion results from the interaction of sickled erythrocytes with leucocytes and the vascular endothelium. Vaso-occlusion then leads to infarction, hemolysis, and inflammation; inflammation enhances the expression of adhesion molecules, further increasing the tendency of sickle erythrocytes to adhere to the vascular endothelium and to worsen vaso-occlusion. Reperfusion of the ischemic tissue generates free radicals and oxidative damage. The damaged erythrocytes release free hemoglobin into the plasma, which strongly binds to nitric oxide, causing functional nitric oxide deficiency and contributing to the development of vasculopathy. Adapted from Hoffbrand *et al*. [79]. and Rees *et al*. [193].

In the presence of α-thalassemia, the concentration of hemoglobin in each erythrocyte is reduced, decreasing the tendency of Hb S to polymerase, which in turn results in increased hemoglobin concentrations and decreased rates of hemolysis. For this reason, as it happens in this patient, the clinical effects of α-thalassemia are variable but generally beneficial for patients, decreasing the occurrence of stroke, gall stones, leg ulcers, and priapism. Although it improves the prognosis, α-thalassemia does not reduce the pain frequency [193].

3.1.5 Case 6: Compound heterozygous for Hb Nigeria + deletion -$\alpha^{3.7kb}$

A 50-year-old male presenting with microcytic (MCV = 73.8 fL) hypochromic (MCH = 26.2 pg) polycythemia (RBC = 5.84 x10^{12}/L), which means elevated RBC, and elevated MCHC (35.5 g/dL) and RDW (18.6 %).

Table 3.6 – Hematological features for Case 6.

Sex	Age	RBC (x10^{12}/L)	Hb (g/dL)	Ht (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)
M	50 y	**5.84**	15.3	43.1	**73.8**	**26.2**	**35.5**	**18.6**

Bold – values outside normal range; **M** – Male; **y** – Years; **RBC** – Red blood cells; **Hb** – Hemoglobin; **Ht** – Hematocrit; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **MCHC** – Mean corpuscular hemoglobin concentration; **RDW** – Red Cell Distribution Width.

Hemoglobin analysis was performed by RP-HPLC and revealed the presence of an abnormal α-chain (Figure 3.8 A). To investigate the presence of a hemoglobin variant we tried to amplify the α-genes so they could later be Sanger sequenced. The amplification of the α-genes is carried out using the same primers as the ones used in the gap-PCR for detection of the 3.7 kb α-thalassemia deletion. Upon observation of the amplified fragments in an agarose gel, we were able to ascertain that the subject was heterozygous for the α^{+}-thalassemia trait (αα/-$\alpha^{3.7}$), (Figure 3.6 B; green). Following this discovery, only the *HBA2* was Sanger sequenced. DNA sequencing confirmed the presence of a heterozygous mutation in codon 81 leading to the substitution of serine for cysteine (*HBA2*:c.245C>G – alpha2 81(F2) Ser>Cys) and resulting in a rare hemoglobin variant called Hb Nigeria (Figure 3.8 B).

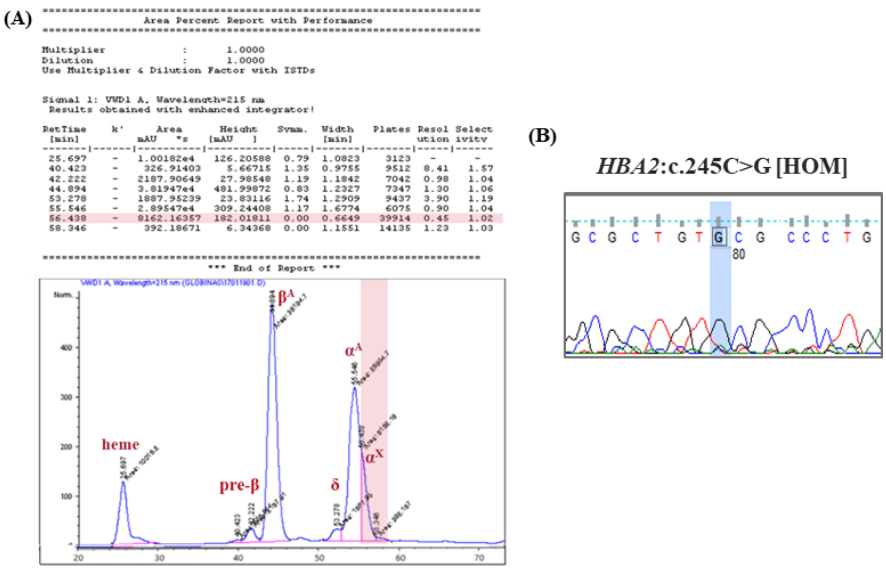


Figure 3.8 – Biochemical characterization and molecular identification of a compound heterozygous for Hb Nigeria and alpha thalassemia -$\alpha^{3.7}$ deletion. (A) Representative chromatograms of globin chain analysis by RP-HPLC revealing the presence of abnormal α chain (α^{X}-pink). **(B)** Partial electropherogram from automated Sanger sequencing of the sense strand of *HBA2* gene at exon 3 showing the *HBA2*:c.245C>G (alpha2 81(F2) Ser>Cys) the mutation responsible for Hb Nigeria. Mutated positions in the affected individual are indicated in a blue box. α-thalassemia trait confirmed by gap-PCR (see Figure 3.5 B - green).

Even though the rare variant Hb Nigeria presents with a new cysteine residue on the surface of the molecule, there is no evidence for the formation of abnormal Hb polymers, since this position is not involved in any of the interchain contacts [196], [197]. Previously, Honig *et al.* had already associated this variant with α-thalassemia, stating that the phenotypic presentation of the subject, microcytosis, hypochromia, and abnormal erythrocyte morphology, was only a product of the α-thalassemia [197].

In conclusion, the subject's phenotype can be justified by the presence of α^{+}-thalassemia trait with the simultaneous presence of the rare variant Hb Nigeria not being responsible for any additional clinical complications.

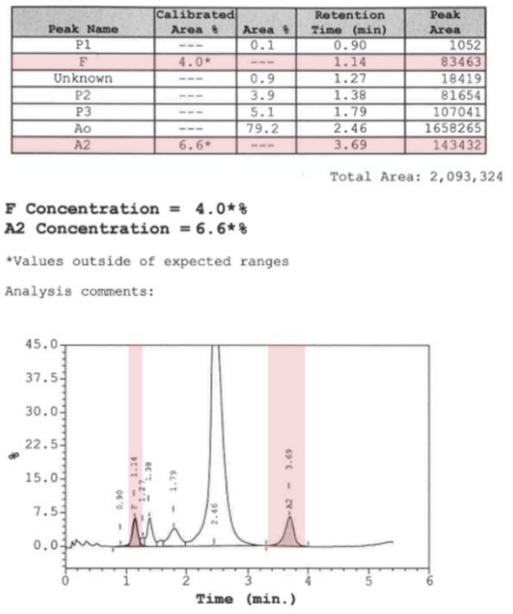
3.1.6 Case 7: Heterozygous for the Cape Verde deletion

A 12-year old female child arrived for hemoglobinopathy's studies exhibiting microcytic (MCV = 64.2 fL) hypochromic (MCH = 19.6 pg) anemia (Hb = 10.7 g/dL), with low MCHC (30.5 g/dL) and elevated RDW (17.4 %). Hemoglobin analysis was performed by HPLC and revealed elevated levels of Hb F (4.0 %) and Hb A2 (6.6 %) (Figure 3.9). Normally, upregulation of γ- and/or δ-globin genes, responsible for Hb F and HbA2, respectively, in association with microcytic anemia, are typical features of β-thalassemia trait [198].

Table 3.7 – Hematological and biochemical features for Case7.

Sex	Age	RBC (x10^{12}/L)	Hb (g/dL)	Ht (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)	SI (µg/dL)
F	12 y	5.5	**10.7**	35.0	**64.2**	**19.6**	**30.5**	**17.4**	122.0

Bold – values outside normal range; **F** – Female; **y** – Years; **RBC** – Red blood cells; **Hb** – Hemoglobin; **Ht** – Hematocrit; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **MCHC** – Mean corpuscular hemoglobin concentration; **RDW** – Red Cell Distribution Width; **SI** – Serum iron.



Peak Name	Calibrated Area %	Area %	Retention Time (min)	Peak Area
P1	---	0.1	0.90	1052
F	4.0*	---	1.14	83463
Unknown	---	0.9	1.27	18419
P2	---	3.9	1.38	81654
P3	---	5.1	1.79	107041
Ao	---	79.2	2.46	1658265
A2	6.6*	---	3.69	143432

Figure 3.9 – Biochemical characterization of unknown hemoglobin variant. Elution chromatogram profile by HPLC showing elevated Hb F and Hb A2 (pink).

The *HBB* gene was entirely amplified and sequenced in search of common β-thalassemia mutations or the presence of molecular lesions responsible for Hb variants, neither one was found. Although, the defects responsible for β-thalassemia are predominantly single base substitutions and small insertions or deletions (INDELs) in the *HBB* gene, that affect almost every known stage

of gene expression, larger deletions, either restricted to the β-globin gene or involving the β-LCR with or without the gene, may also be behind these disorders [72]. In this context, it is difficult to detect large deletions because the deletion might not be within the borders of the primers of the amplified segment, if this happens, in case of heterozygosity, the allele containing the deletion will otherwise not be amplified, and the analysis result will falsely appear to be homozygous wild-type [199]. For these reasons, to screen for large deletions, the copy number of target fragments in the β-globin gene cluster was determined using the MLPA assay with normal or carrier status being determined according to the ranges obtained from the ratios of normal samples.

In this case, the MLPA assay results revealed that 3 probes were not amplified (13619-L15073 [Exon 3]; 11885-L25666 [0.2 kb after Exon 3]; and 05836-L06321 [0.5 kb after Exon 3], making up for a deletion that could range from 0.6 kb to 10 kb (Figure 3.10 A). Andersson *et al*. reported the 5' and 3' endpoints of a 7.7 kb deletion, starting from the *HBB* IVSII to 3' downstream of the β-globin gene, and present a PCR strategy for rapid DNA diagnosis of the named Cape Verde-thalassemia deletion [124]. The missing probes, in this case, indicated a deletion matching the description of the Cape Verde deletion, therefore, the describing author's diagnosis protocol was followed. The presence of the Cape Verde deletion was at first sight confirmed, its detection was done by agarose gel electrophoresis of the PCR product of *HBB* gene containing the deletion and presenting a 419 bp fragment, normal controls present with no amplification of the target fragment (Figure 3.10 B). Additionally, the fragment was sequenced confirming a 7719 bp deletion (Figure 3.10 C).

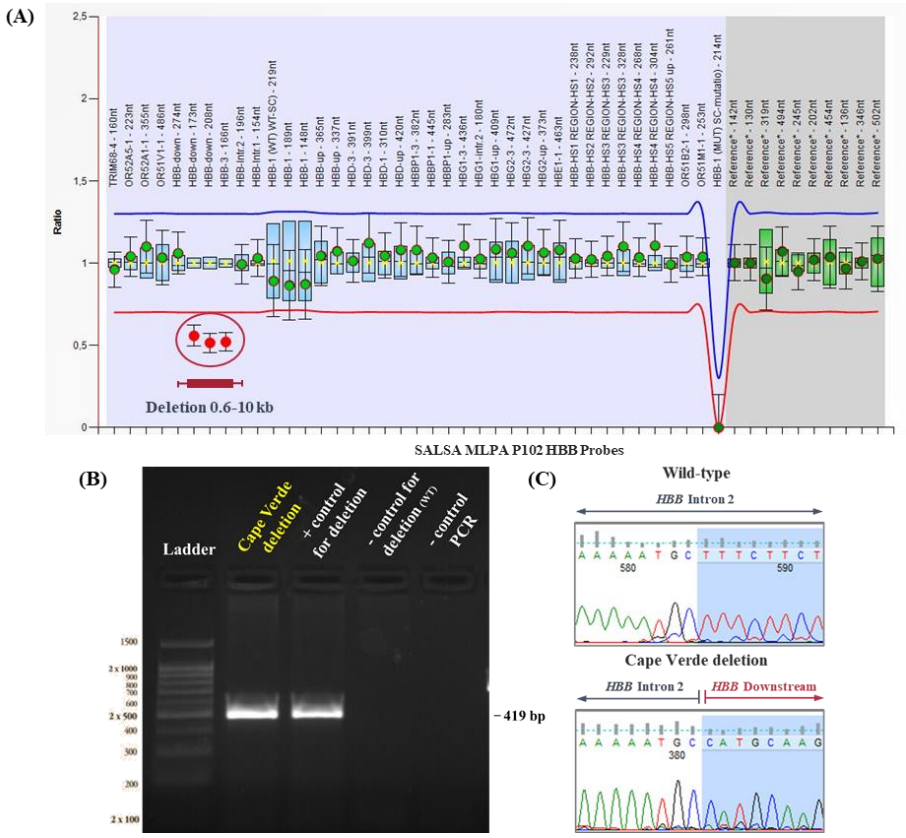


Figure 3.10 – Molecular identification of β-globin gene cluster Cape Verde deletion. (A) MLPA analysis showing the relative probe signals across the β-globin gene cluster, missing probes and estimated size of the deletion shown in dark red, blue and red lines represent the ranges obtained from the ratios of normal samples; **(B)** Detection of Cape Verde deletion by gel electrophoresis of the PCR product of *HBB* gene containing the deletion and presenting a 419 bp fragment (yellow), as it is shown in Andersson *et al*. 2006 [124]; and **(C)** Partial electropherogram from automated Sanger sequencing of the sense strand of *HBB* gene confirming a 7719 bp deletion U01317:g.71551_79269del7719

(beta nts 1007 - 8725 deleted). Wild-type nucleotides and mutated positions in an unaffected individual and the affected patient are indicated in a blue box. For measuring the size of the expected fragments, a DNA Ladder 100 bp plus (AppliChem, USA) was used.

The Cape Verde thalassemia deletion is one of two known deletions that remove the 3'-end of the *HBB* gene but preserve the integrity of its 5'-end [124]. However, this large deletion origins an absence of beta-chain synthesis from this allele. The identification of the deletion allows proper therapeutics and genetic counseling for the child's family.

3.1.7 Case 8: Hb F variability

This case reports a 26-year-old female of Portuguese ancestry from the mother side and Egyptian ancestry from the father side. The subjects had CBC parameters within the normal range but present low serum iron (SI = 48.0 µg/dL), which may indicate iron deficiency, however, SI is not the most reliable marker for iron deficiency (Table 3.8).

Table 3.8 – Hematological and biochemical features for Case 8.

Sex	Age	RBC ($x10^{12}$/L)	Hb (g/dL)	Ht (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)	SI (µg/dL)
F	26 y	4.5	13.3	41.6	91.8	29.4	32.1	12.9	**48.0**

Bold – values outside normal range; **F** – Female; **y** – Years; **RBC** – Red blood cells; **Hb** – Hemoglobin; **Ht** – Hematocrit; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **MCHC** – Mean corpuscular hemoglobin concentration; **RDW** – Red Cell Distribution Width; **SI** – Serum Iron.

Although hematological parameters were within the normal range, when performed hemoglobin analysis through HPLC, the results showed elevated Hb F (7.2 %) indicative of HPFH (Figure 3.11 A). Additionally, $^{G}\gamma$:$^{A}\gamma$ ratio was determined through globin chain analysis by RP-HPLC revealing an apparent ratio of 100:0 (Figure 3.11 B), suggesting that the Hb F present in excess is composed only of $^{G}\gamma$ chains, differing completely from the expected γ-globin chain ratio of 40:60 present in adult trace amounts of Hb F.

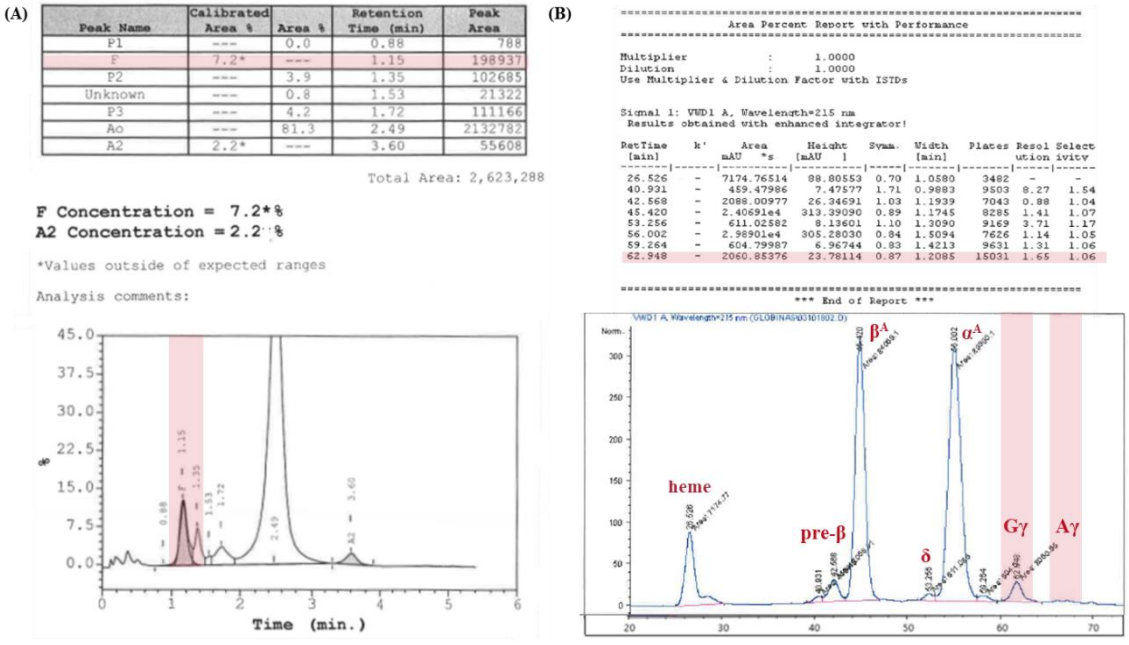


Figure 3.11 – Biochemical characterization of Hb F polymorphisms resulting in anormal chromatographic profiles. Representative chromatograms of: **(A)** elution profile by HPLC showing elevated Hb F (pink) and **(B)** globin chain analysis by RP-HPLC revealing an apparent absence of $^{A}\gamma$ globin chain (pink).

Initially, to rule out any defects that could dysregulated erythropoiesis and lead to secondary Hb F synthesis into adulthood, the *HBB* gene was entirely sequenced, but revealed no defects. Hereinafter, since there are several large deletions that are in the origin of HPFH [200] and because the $^{A}\gamma$-globin chain was apparently not being expressed, a MLPA assay was performed to detect large deletions in the beta cluster. The results revealed, by comparison with healthy controls, a normal copy number for the amplified sequences detected by the target and reference probes, concluding that a large deletion was not responsible for the present phenotype. The next step included the sequencing of the genes *HBG2* and *HBG1*, responsible for the γ-globin chains in Hb F, and purposely including the analysis of both their promoter regions. The analysis exposed the presence of the *Xmn*I polymorphism in the *HBG2* gene promoter region (*HBG2*:c.-158C>T – rs7482144) and a variation of the polymorphic region (TG)n(CG)m in the second intron of the *HBG1* gene region by deletion of the IVSII $TG_{(4)}CG_{(5)}$ motif (NM_000559.2:c.316-302_316-285del18 – rs61080176). Molecular findings are displayed in Figure 3.12 A and B.

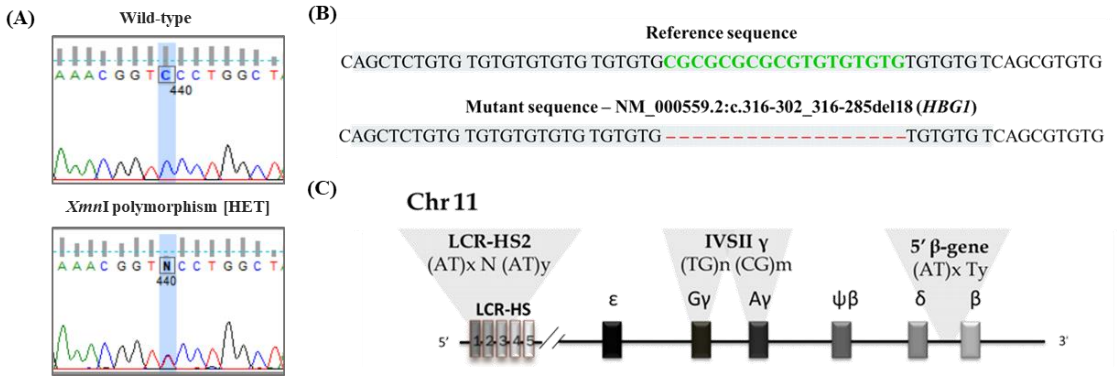


Figure 3.12 – Molecular identification of Hb F associated polymorphisms. (A) Partial electropherogram from automated Sanger sequencing of the sense strand of *HBG2* gene at the promotor showing heterozygosity for *Xmn*I polymorphic site *HBG2*:c.-158C>T (rs7482144) responsible for increased Hb F. Wild-type nucleotides and mutated positions in an unaffected individual and the affected patient are indicated in a blue box. **(B)** *HBG1* deletion of the IVSII $TG_{(4)}$ $CG_{(5)}$ motif (rs61080176); and **(C)** map of the β-globin gene cluster and the polymorphic regions, with respective possible microsatellite configurations. Adapted from Moumni *et al.* 2016 [201]

Normal inter-individual Hb F variation is a quantitative trait influenced by many loci inside or outside the β-globin gene cluster and can lead to a heterogeneous group referred to as hereditary persistence of fetal hemoglobin [202]. In some non-deletional forms of HPFH, a group of single point mutations in the γ-globin gene promoters are associated with increased γ-globin gene transcription, likely through mechanisms of increased competition for the LCR sequences against the adult globin genes or as a consequence of reduced binding of repressor factors [203]. The presence of *Xmn*I polymorphic site in the $^{G}\gamma$-globin promoter region has been positively correlated with elevated synthesis of Hb F and increased expression of $^{G}\gamma$-globin component in term newborn infants and is associated with a delayed switch over from fetal to adult hemoglobin [204]. The predisposition of carriers to increased Hb F concentrations is normally more accentuated when they are under conditions of erythropoietic stress, such as in sickle cell disease and β-thalassemia [202].

Microsatellites are short (typically less than 100 bp) DNA sequences in which motifs of 1–6 bp are tandemly repeated. These regions are highly polymorphic in their length, resulting in variability in repeat number, making them very useful as genetic markers [205]. In the β-globin gene cluster, there are several polymorphic microsatellite regions (Figure 3.12 C), which are possible binding sites of the transactivator factors involved in the chromatin structure of the β-globin locus. One example of this is intervening sequence II region of two fetal hemoglobin genes,

and variation on this site, as is the case of the motif found in the $^{A}\gamma$-globin gene of the case in study, alter the binding of transcriptional factors to the cis-regulatory elements and consequently activate or repress the transcription of certain β-globin genes [201]

The sum effect of the molecular findings in *HBG2* and *HBG1* genes are believed to be responsible for the variation in chain composition and elevated levels of Hb F but are not responsible for the presented low SI, indicating hemoglobinopathy unrelated iron deficiency. The subject is not anemic, however, in case of further concern by the clinician, complementary analysis of the iron status is required, in particular, the measurement of Ft and TSAT to assess the state of iron stores and traffic, respectively.

3.1.8 Case 9: Heterozygous for Hb Strasbourg

The subject is a 35-year-old female with polycythemia (RBC =5.07 $x10^{12}$/L) and the consequent increase of hemoglobin (16.0 g/dL) and hematocrit (48.4 %) that is the volume percentage of red blood cells in the blood (Table 3.9), this condition normally serves as a compensatory mechanism for impaired blood oxygenation and reduced oxygen saturation, preventing the tissues to obtain an adequate oxygen supply [206], which could be indicative of the presence of a hemoglobin variant with high oxygen affinity.

Table 3.9 – Hematological features for Case 9.

Sex	Age	RBC ($x10^{12}$/L)	Hb (g/dL)	Ht (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)
F	35 y	**5.07**	**16.0**	**48.4**	95.5	31.6	33.1	12.9

Bold – values outside normal range; **F** – Female; **y** – Years; **RBC** – Red blood cells; **Hb** – Hemoglobin; **Ht** – Hematocrit; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **MCHC** – Mean corpuscular hemoglobin concentration; **RDW** – Red Cell Distribution Width.

To confirm the presence of a hemoglobin variant, CE was performed, and its results substantiated the suspicion with an electrophoretic profile presenting a hemoglobin variant migrating in "Z11 zone", which represented 43.2 % of the total hemoglobin (Figure 3.13 A).

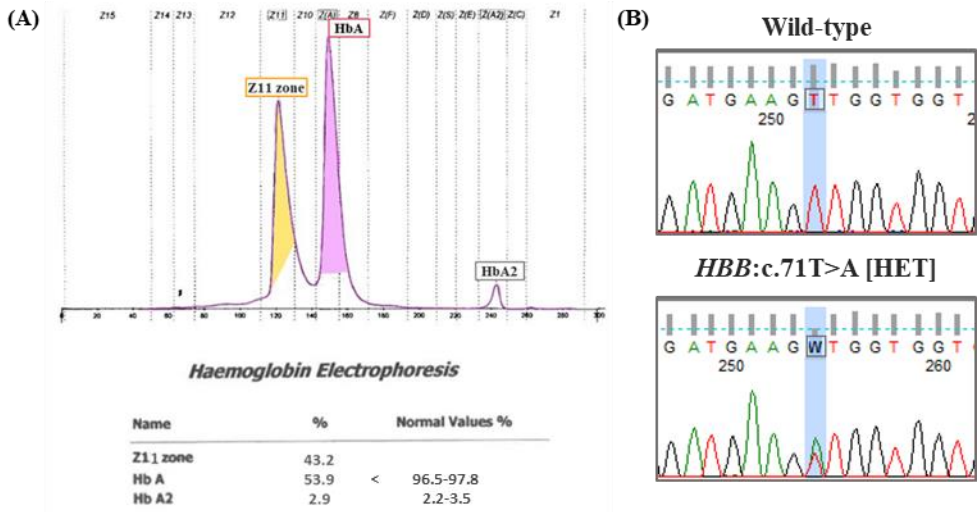


Figure 3.13 – Biochemical characterization and molecular identification of hemoglobin variant, Hb Strasbourg. **(A)** Electrophoretic profile by capillary electrophoresis showing Hb variant migrating in "Z11 zone" (yellow); **(B)** Partial electropherogram from automated Sanger sequencing of the sense strand of *HBB* gene at exon 1 showing the *HBB*:c.71T>A (beta 23(B5) Val>Asp) the mutation responsible for the Hb variant. Wild-type nucleotides and mutated positions in an unaffected individual and the affected patient are indicated in a blue box.

Molecular confirmation of the variant was performed by complete amplification and sequence of the *HBB* gene. Sequence analysis of the amplified DNA identified a heterozygote mutation at codon 23 that changed a GTT into a GAT causing a valine to be replaced by an aspartic acid in the globin chain (HBB:c.71T>A – beta 23(B5) Val>Asp), giving rise to a variant known as Hb Strasbourg (Figure 3.13 B).

Described originally in a Portuguese woman by Garel *et al.*, Hb Strasbourg is a hemoglobin that exhibits a high oxygen affinity and erythrocytosis [207], which is in agreement with the phenotype presented by this case's subject, therefore concluding the search for any other variant. In the presence of high oxygen affinity hemoglobin variants, the lower release of oxygen leads to tissue hypoxia prompting an erythropoietin output at renal level inducing both an increase in erythropoiesis and an offsetting secondary erythrocytosis. These symptoms are typically well tolerated by young patients, although thrombotic complications have been reported in elderly patients or when other vascular hazard factors are associated [208].

3.1.9 Case 10: Heterozygous for Hb J-Iran

The subject is a 40-year-old pregnant woman from Moldavia presenting with macrocytosis (MCV = 98.3 fL) arrived for hemoglobinopathy screening (Table 3.10). During pregnancy, microcytosis is defined as an MCV of < 80 fL and macrocytosis by an MCV of > 95 fL [209].

Table 3.10– Hematological and biochemical features for Case 10.

Sex	Age	RBC ($x10^{12}$/L)	Hb (g/dL)	Ht (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)	SI (µg/dL)
F	40 y	4.2	13.3	41.0	**98.3**	31.9	32.4	13.0	119.0

Bold – values outside normal range; **F** – Female; **y** – Years; **RBC** – Red blood cells; **Hb** – Hemoglobin; **Ht** – Hematocrit; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **MCHC** – Mean corpuscular hemoglobin concentration; **RDW** – Red Cell Distribution Width; **SI** – Serum iron.

To screen for hemoglobinophaties, hemoglobin analysis was performed by HPLC with additional globin chain analysis by RP-HPLC. The subject's chromatographic elution profile by HPLC presented quite elevated Hb F (33.9 %), most likely indicating the presence of a hemoglobin variant that migrates close to Hb F and may not be separated from it. Such elevated values are too great, even during pregnancy, and since it is possible to observe an abnormal asymmetric peak further indicate the presence of a Hb variant (Figure 3.14 A). The globin chain analysis by RP-HPLC confirmed the presence of abnormal β chain (β^X) and absence of γ-globin chains that constitute fetal hemoglobin, results are displayed in Figure 3.14 B. Thereafter, the subjet's *HBB* gene was amplified and sequenced unveiling a CAC>GAC heterozygous mutation in codon 77 in the second exon of the gene that results in a substitution of a histidine for an aspartic in globin chain (HBB:c.232C>G – beta 77(EF1) His>Asp). This mutation is responsible for the Hb J-Iran (Figure 3.14 C).

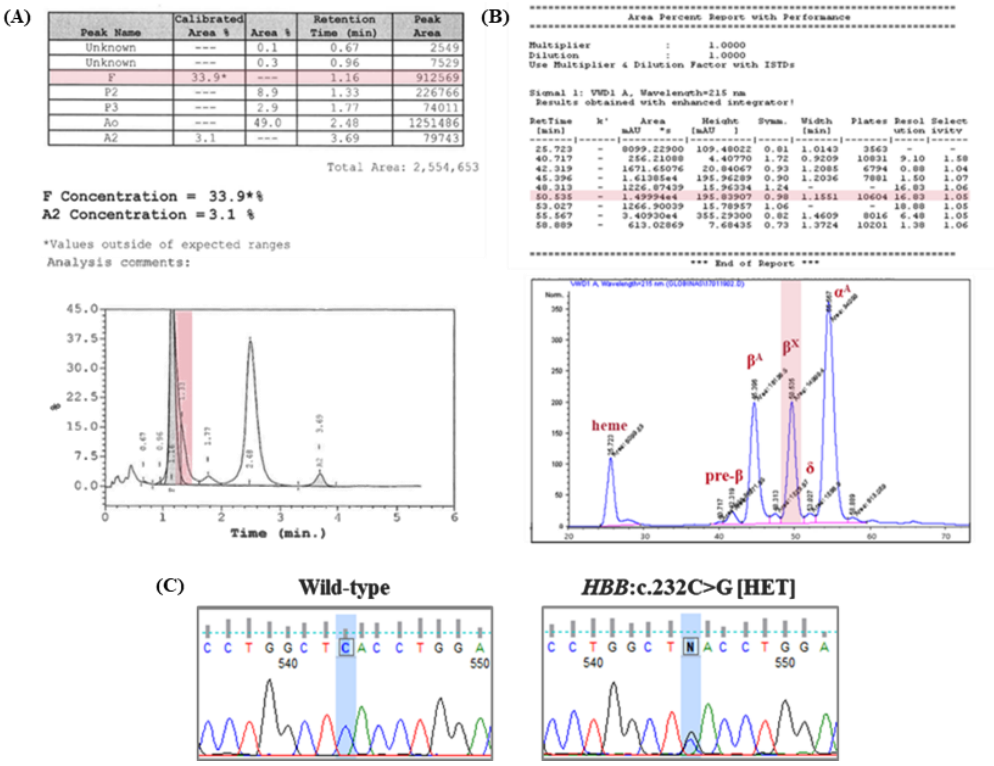


Figure 3.14 – Biochemical characterization and molecular identification of hemoglobin variant, Hb J-Iran. Representative chromatograms of **(A)** elution profile by HPLC showing Hb variant migrating in Hb F zone (pink) and **(B)** globin chain analysis by RP-HPLC revealing the presence of abnormal β chain (β^X-pink). **(C)** Partial electropherogram from automated Sanger sequencing of the sense strand of *HBB* gene at exon 2 showing the *HBB*:c.232C>G (beta 77(EF1) His>Asp) the mutation responsible for the Hb J-Iran. Wild-type nucleotides and mutated positions in an unaffected individual and the affected patient are indicated in a blue box.

Hb J-Iran is a variant reported manly in Iranian, Turkish and Russian-Armenian families and was originally described by Rahbar *et al.* in the ethnically diverse population of Iran [210]–[212]. This is a stable variant that varies between 30 % and 45 % in heterozygotes and is often missed for not affecting the hematologic parameters nor leads to clinical abnormalities [213].

Many pregnant women have enlarged red blood cells, which can be caused by whether by normal physiological changes pregnancy or by a pathological condition. Macrocytosis developing during pregnancy is a physiological change in most cases, where pregnant women present with normoblastic erythropoiesis [214]. Additionally, normal newborn infants have larger RBC than the healthy adult and it is well established by now that cells are exchanged between mother and fetus during gestation, which can contribute to the increased MCV levels [215]. Macrocytosis is the most sensitive index of a megaloblastic process, for example in pernicious anemia, which occurs normally in pregnancy [214]. Deficiency of vitamin B12 or folic acid induces megaloblastic changes in most actively replicating tissues, but these morphologic changes are most pronounced in bone marrow [216]. Contrary to this, the intake of folate supplements, in the absence of ID, has been associated with an increase in MCV [214].

The subject´s hemoglobin variant does not induce any megaloblastic changes that could be responsible for the increased MCV levels. The changes in the proband RBC's size are either a consequence of the pregnancy itself or of folate supplements, normally prescribed to pregnant women to help prevent birth defects since there are no other indications of megaloblastic anemia. Finally, is important to note that diabetes during gestation is associated with an increased risk of fetal and maternal complications [217], and although this variant has no clinical significance, Hb

J-Iran should be taken into account in the context of diabetes since it interferes with the measurement of HbA1c by the cation exchange (CE)-HPLC method, and so the clinician should opt for an alternative method to measure HbA1c [218].

3.1.10 Case 11: Heterozygous for Hb E

In this case we explain the phenotype of a 38-year-old male presenting with microcytic (MCV = 76.5 fL) and hypochromic (MCH = 25.5 pg) polycythemia (RBC = 7.75 x10^{12}/L; Hb = 19.4 g/dL; and Ht = 59.3 %). The complete blood count is displayed in Table 3.11. A high RBC count combined with a low mean volume is generally attributed to thalassemia minor, either α or β, or polycythemia vera, a type of blood cancer, with iron deficiency [219].

Table 3.11 – Hematological features for Case 11.

Sex	Age	RBC (x10^{12}/L)	Hb (g/dL)	Ht (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)
M	38 y	**7.75**	**19.4**	**59.3**	**76.5**	**25.5**	32.7	14.9

Bold – values outside normal range; **M** – Male; **y** – Years; **RBC** – Red blood cells; **Hb** – Hemoglobin; **Ht** – Hematocrit; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **MCHC** – Mean corpuscular hemoglobin concentration; **RDW** – Red Cell Distribution Width.

Hemoglobin analysis was performed by CE, the electrophoretic profile revealed a Hb variant (23.8 %) migrating in "E zone" (Figure 3.15 A). The Hb variant detected by presumptive methods was confirmed by molecular analysis of the *HBB* gene and confirmed the presence of Hb E in heterozygosity (HBB:c.79G>A – beta 26(B8) Glu>Ly) resulting from a G>A substitution in codon 26 (Figure 3.15 B), which produces structurally abnormal hemoglobin while activating a cryptic splice site, resulting in faulty mRNA processing. The quantity of normally spliced β^E mRNA is reduced due to the creation of a new stop codon, the abnormally spliced mRNA is nonfunctional. Consequently, Hb E is synthesized at a reduced rate and behaves like a mild form of β-thalassemia [220].

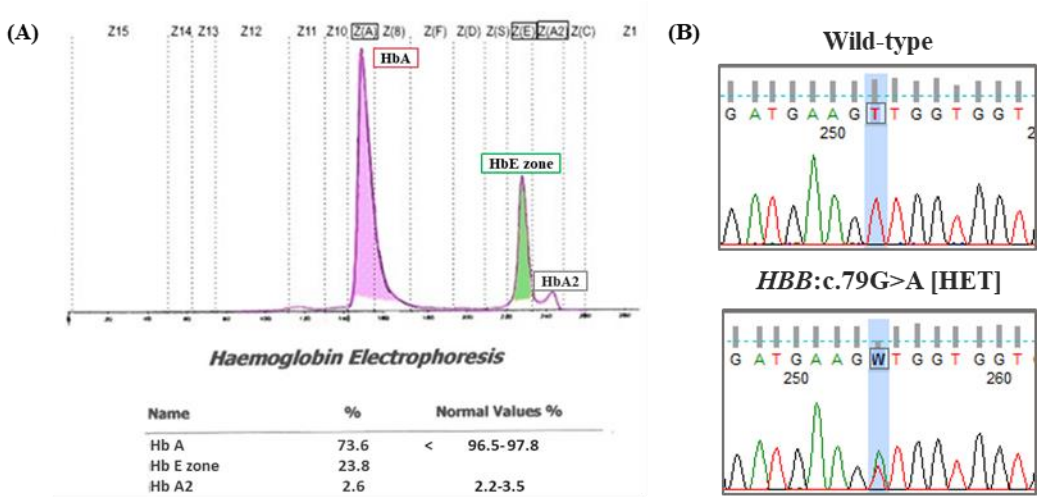


Figure 3.15 – Biochemical characterization and molecular identification of hemoglobin variant, Hb E. (A) Electrophoretic profile by capillary electrophoresis showing Hb variant migrating in "E zone" (green); **(B)** Partial electropherogram from automated Sanger sequencing of the sense strand of *HBB* gene at exon 1 showing the *HBB*:c.79G>A (beta 26(B8) Glu>Lys) Val>Asp) mutation confirming the Hb E suspicion. Wild-type nucleotides and mutated positions in an unaffected individual and the affected patient are indicated in a blue box.

Hb E is one of the world's most common and important mutations in Southeast Asia and Africa, with the genotype Hb E/β-thalassemia being responsible for approximately half of all

severe β-thalassemia cases worldwide [221], [222]. The disorder is characterized by high clinical variability, ranging from mild and asymptomatic anemia to a life-threatening disorder that requires transfusions from infancy [220]. Albeit, the cause of the remarkable variability in individuals with Hb E/β-thalassemia remains mostly unknown. Within a family, patients with the same mutations may show significant differences in clinical severity [221].

Having Hb E trait typically has no clinical significance. However, patients may present with mild microcytosis without anemia and, frequently, mild erythrocytosis [223]. This hemoglobin is also slightly unstable and may lead to hemolysis triggered by viral infections and medications [93]. These phenotypic features must be recognized so that people with this innocuous condition are not subjected to unnecessary medical examinations and inappropriate treatment [223]. In the end, we do not be the elevated number of RBCs is a direct result of the presence of this structural variant and consequent β-thalassemic trait. For this reason, further studies should be conducted to unveil the complete molecular causes behind this subject's phenotype and rule out polycythemia vera.

3.1.11 Cases 12, 13, 14 and 15: Segregation in a family of three hemoglobin defects

In this part of the dissertation, we describe a family study carried out in a family of Moroccan origin that presented segregation of three hemoglobin defects. The proposita is a 36-year-old female, mother of three, presenting with microcytic (MCV = 72.9 fL) hypochromic (MCH = 22.9 pg) anemia (Hb = 12.6 g/dL). There was no information regarding her iron status, but she was known to be heterozygous for the α^{+}-thalassemia trait ($\alpha\alpha/-\alpha^{3.7}$). The hematological features of the affected family are summarized in Table 3.12.

Table 3.12 – Hematological and biochemical features from individuals in the family study (Cases 12 to 15).

Family	Sex	Age	RBC ($x10^{12}$/L)	Hb (g/dL)	Ht (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)	SI (µg/dL)	Ft (µg/L)
Mother	F	36 y	5.5	**12.6**	40.3	**72.9**	**22.9**	31.4	14.4	102	-
Father	M	42 y	5.5	15.6	48.1	87.3	28.3	32.5	12.9	74	-
Daugther1	F	11 y	4.8	13.6	41.5	86.9	28.9	32.9	13.0	**59**	**17.36**
Daughther2	F	6 y	5.8	10.9	34.3	**59.2**	**18.9**	31.9	16.7	71	**7.45**
Son	M	5 y	5.3	11.3	35.7	**67.5**	**21.3**	31.5	15.3	72	**23.82**

Bold – values outside normal range; **M** – Male; **F** – Female; **y** – Years; **RBC** – Red blood cells; **Hb** – Hemoglobin; **Ht** – Hematocrit; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **MCHC** – Mean corpuscular hemoglobin concentration; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Ft** – Ferritin.

The proposita's hemoglobin analysis was performed by HPLC, with elution chromatogram profile revealing a small percentage of a Hb variant migrating in "S-window" (1.6 %) and decreased Hb A2 (1.9 %), (Figure 3.16 A). The Hb variant detected by presumptive methods led to the molecular analysis of the *HBB* gene, that in turn, unveiled the presence of β^{+}-thalassemia trait (IVS-I-6; *HBB*:c.92+6T>C– beta nt 148 T>C) brought on by an aa substitution in the exon-intron boundary of the first intron (Figure 3.16 B). Normally, β-thalassemia trait subjects have characteristically increased Hb A2 levels, since the proposita presented with decreased levels of this hemoglobin, additional molecular analysis of the *HBD* gene was performed to justify these results. Complete Sanger sequencing of the gene revealed the heterozygous presence of the variant Hb A2'(*HBD*:c.49G>C – delta 16(A13) Gly>Arg), also called B2, a clinically silent variant that results from a G>C substitution at codon 16 of the δ-globin gene, which entails an exchange of glycine for arginine (Figure 3.16 B). This genetic disorder has been detected in

homozygous and heterozygous states, and may also be co-inherited with β-thalassemia traits and other minor hemoglobinopathies [224].

Overall, the molecular analysis allowed us to discover that the proposita is a triple-heterozygous patient resulting from a combination of α^{+}- and β^{+}-thalassemia trait, and the variant Hb A2'.

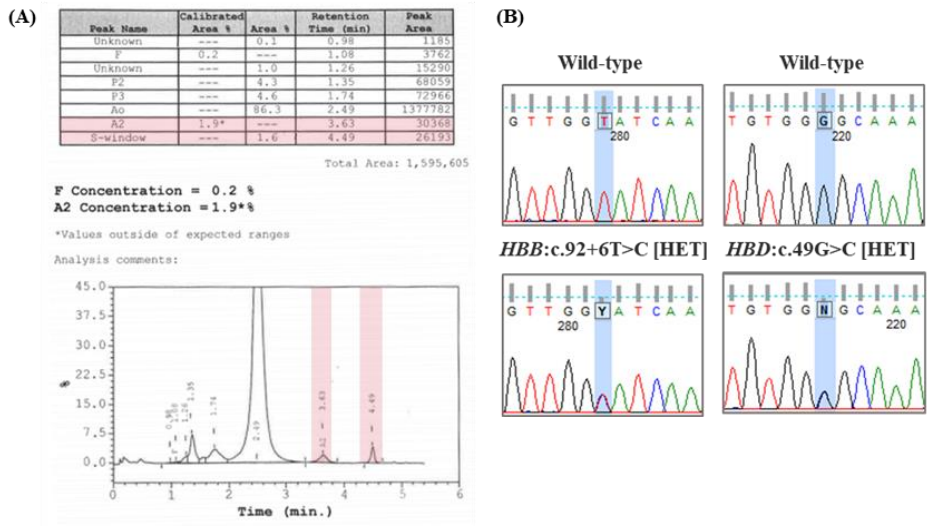


Figure 3.16 – Biochemical characterization and molecular identification of three hemoglobin defects in the proposita, HbA2' (HbB2) and α- and β-thalassemia trait simultaneously. (A) Elution chromatogram profile by HPLC showing Hb variant migrating in "S-window" and decreased Hb A2 (pink). **(B)** Partial electropherogram from automated Sanger sequencing of the sense strand of *HBB* gene at intron 1 showing HBB*:c.92+6T>C* (IVS-I-6 (T->C); β+) the mutation responsible for β-thalassemia trait (left) and *HBD* intron 1 showing *HBD:c.49G>C* (delta 16(A13) Gly>Arg) the mutation responsible for Hb A2' (right). Wild-type nucleotides and mutated positions in an unaffected individual and the affected patient are indicated in a blue box. α-thalassemia trait confirmed by gap-PCR (see Figure 3.6 B - green).

Following the analysis of the proposita, the same molecular analysis was performed for the remaining members of the family. At the time of the analysis, the husband, father of her children, presented with hematological parameters within the normal range, the older daughter presented with iron deficiency (SI = 59.0 µg/dL; Ft = 17.36 µg/L), and the two younger children presented with microcytosis and hypochromia accompanied by iron deficiency, according to the parameter ranges set for their age group. Results confirmed that the father was healthy with no molecular lesions being found. When it came to the children there was segregation of the lesions present in the mother's *HBB* and *HBD* genes, with the older daughter being heterozygous for the Hb A2' variant, and the younger daughter and son being β-thalassemia carriers. The pedigree of the family and laboratory findings are shown in Figure 3.17.

Molecular analysis must be used for the accurate diagnosis of double heterozygous α- and β-thalassemia for proper risk assessment, especially in geographical areas in which both have a high prevalence. This determines couples at risk of having children with severe α-thalassemia, such as Bart's hydrops fetalis, or β-thalassemia major [225]. In this case, the husband was proved to be healthy when it came to the globin chains, and so we conclude that in case the couple decides to have other children, those will not be at risk of having any severe thalassemia syndrome. The same cannot be said for the two younger children, carriers of β-thalassemia, when they choose to have children of their own, therefore at that time, genetic counselling should be provided. Additionally, all the children have iron deficiency not justified by the found molecular lesions, so iron therapy should be conducted.

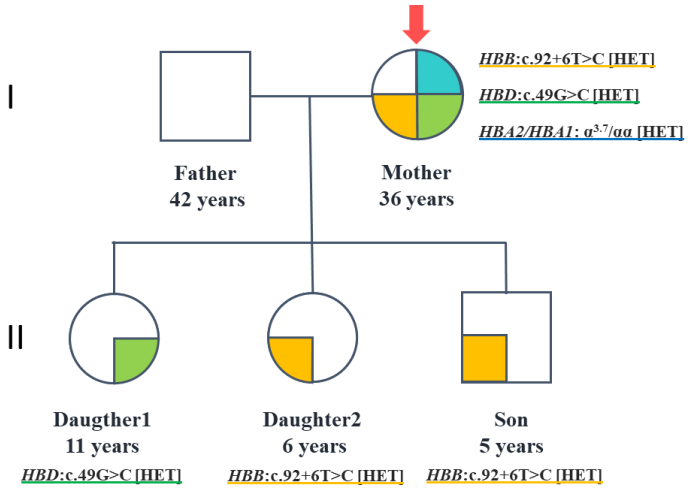


Figure 3.17 – Pedigree analysis and genotypes of the family studied. Mutation screening was done for the proposita (red arrow), her husband and offspring. Yellow represents the β^{+}-thalassemia trait (*HBB*:c.92+6T>C – beta nt 148 T>C); blue represents the α^{+}-thalassemia trait ($\alpha\alpha/-\alpha^{3.7}$), and green represents the variant Hb A2' or B2 (*HBD*:c.49G>C – delta 16(A13) Gly>Arg). **HET** – Heterozygous; **HOM** – Homozygous.

3.2 Cases with suspicion of IRIDA

3.2.1 Patients' genetic variants identification and validation

The genetic characterization of the samples of 10 patients with suspicion of IRIDA was carried out. The hematological and iron status features of the subjects are displayed in Table 3.13, outside the normal range values reveal that all patients presented with microcytic hypochromic iron deficiency anemia, according to normal range values adapted for each age group. Like the cases in hand, IRIDA patients normally present with extremely low TSAT levels and normal/borderline low Ft levels. For this reason, and because the subjects had no success with previous treatment, with oral and intravenous iron, and other causes for microcytic anemia were excluded, these cases were suitable for molecular analysis of the exons and exon-intron boundaries of the *TMPRSS6* gene in search for molecular lesions indicative of IRIDA.

Table 3.13 – Hematological and iron status features in subjects tested for IRIDA.

ID	Sex	Age	RBC ($x10^{12}$/L)	Hb (g/dL)	MCV (fL)	MCH (pg)	RDW (%)	SI (µg/dL)	Tf (mg/dL)	TIBC (µg/dL)	TSAT (%)	Ft (µg/L)
1	F	43 y	4.51	**7.9**	**57.9**	**17.4**	**19.0**	**25**	**500**	**625.00**	**4.00**	**4.2**
2	M	3 y	5.03	**7.9**	**56.7**	**15.7**	**22.9**	**22**	-	-	-	**6.0**
3	M	7 y	-	-	-	-	-	-	-	-	-	-
4	F	-	-	-	-	-	-	-	-	-	-	-
5	M	-	-	-	-	-	-	-	-	-	-	-
6	M	3 y	4.59	**9.0**	**63.0**	**19.6**	**17.1**	**20**	**381**	**476.19**	**4.20**	**2.5**
7	F	18 y	4.57	**10.5**	**74.1**	**23.0**	**17.4**	**37**	370	**462.50**	**8.00**	**6.3**
8	M	4 y	4.29	**9.9**	**69.9**	**23.1**	**16.6**	**34**	247	309.09	11.00	16.9
9	M	4 y	4.52	**9.8**	**70.3**	**21.7**	**17.7**	**19**	253	316.67	**6.00**	33.0
10	M	14 y	5.18	**8.7**	**57.4**	**16.8**	**18.1**	**20**	**381**	**476.19**	**4.20**	**2.2**

Bold – values outside normal range; **M** – Male; **F** – Female; **y** – Years; **m** – Month; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

The analysis of the *TMPRSS6* gene was performed by amplifying 3 long-PCR fragments of the gene that amounted for a total of 23,518 bp per subject. Fragments were then sequenced by NGS. Subjects 1 to 4 acted as positive controls of the NGS analysis, since, prior to this study, these patients had the coding regions of the *TMPRSS6* gene analysed by Sanger sequencing having been found common variants suitable for NGS data validation. The found variants were present in heterozygous or homozygous combinations as depicted in Table 3.14. The subject number 5 had also been previously studied for the same reason as the other subjects, but no significant variant was found, for this reason, it was added to the analysis to act as a negative control.

One major technical challenge associated with NGS technologies is the presence of repetitive DNA sequences, which result in ambiguities in alignment and assembly, that in turn, lead to biases and errors in interpretation [205]. Because of these difficulties, a manual depth of coverage analysis of the aligned sequencing data was performed using IGV to visually highlight anomalies and aid in the result interpretation process. Results revealed the presence of three regions with a high density of variants that turned out to be repetitive regions (Figure 3.18), one in fragment 1 (22:37097630-37098146) and the other two in fragment 3 (22: 37071959-37073220; 22:37073347-37073397). The two repetitive regions present in fragment 3 are shown as one in the fragment's graphical representation in Figure 3.18, due to their proximity and the inability to

visually separate them at the presented bp scale. With aid from the software Tandem Repeats Finder, the repetitive patterns in these regions were determined, with three tandemly repeated motifs being identified, $(CCACCGTCCTGTAACGGAGGGGCAGGAGCGGG)_{16}$, $(ATGG)_{325}$, and $(GATG)_{12}$, within each identified region respectively. These motifs leave the polymerase more prone to stuttering, therefore, called variants within these regions were excluded, with a total of 13 variants being left out.

Table 3.14 – Identification of variants present in the *TMPRSS6* gene of positive control subjects.

rs_ID	Location	MC	Region	NM_153609.3	NP_705837.1	GT	S_ID
rs2235321	22:37066886-37066886	Syn	Ex17	c.2217C>T	p.Tyr739=	0/1	3
						1/1	1;4
rs855791	22:37066896-37066896	Miss	Ex17	c.2207T>C	p.Val736Ala	0/1	3
						1/1	1;4
rs4820268	22:37073551-37073551	Syn	Ex13	c.1563C>T	p.Asp521=	0/1	3
						1/1	1;4
rs881144	22:37075250-37075250	Syn	Ex11	c.1254C>T	p.Tyr418=	0/1	2
rs2111833	22:37084757-37084757	Syn	Ex9	c.1083G>A	p.Ser361=	0/1	1;2;4
rs2235324	22:37089684-37089684	Miss	Ex7	c.757A>G	p.Lys253Glu	0/1	1;2;4
rs11704654	22:37103346-37103346	Syn	Ex2	c.99G>A	p.Pro33=	0/1	1;3

rs_ID – reference single nucleotide polymorphism ID; **NM** – Reference sequence based on a protein-coding RNA (mRNA); **NP** – Reference sequence based on a protein (amino acid) sequence; **MC** – Molecular consequence; **Miss** – Missense; **Syn** – Synonym; **Region** – Genomic Region; **GT** – Genotype; **0/1** – Heterozygous; **1/1** – Homozygous; **S_ID** – Subject ID.

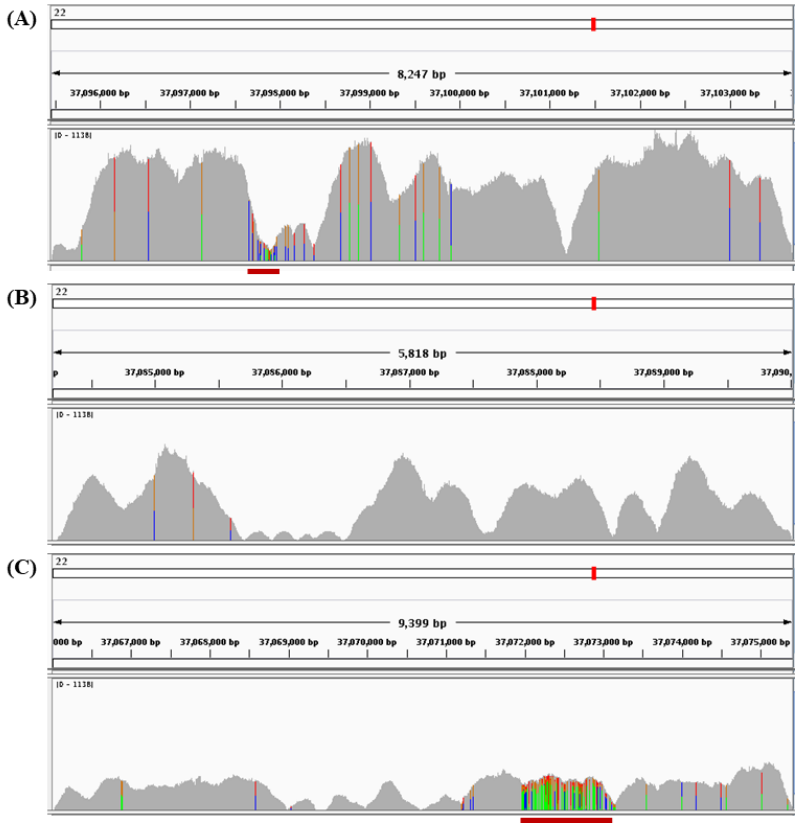


Figure 3.18 – NGS total coverage analysis. (A) Fragment 1 with 8266 bp containing exons 1-6 (22.37095447-371037129); **(B)** Fragment 2 with 5831 bp containing exons 7-10 (22.37084192—3709002); and **(C)** Fragment 3 with 9421 bp containing exons 11-18 (22.37065998—37075418). Reference base (equal) is depicted in grey with called variants in color (C: blue; G: orange - not black; A: green; T: red), and bars represent the number of reads supporting the genomic position. Above is marked the relative chromosomic position and bellow are underlined in red the high variation regions studied for exclusion purposes. Analysis performed in IGV v2.3.86.

After application of the exclusion criteria, the NGS analysis revealed a total presence of 81 variants (Table S.20): 2 upstream/5'UTR variants; 8 coding region variants resulting from 5 synonymous and 3 missense mutations (Figure 3.19); and 71 intronic variants, of those, 16 were in the exon-intron boundaries (up to 120 bp of the flanking intron) and 55 were deep intronic. All called variants had already received an rs identifier, with all variants found within the coding regions having already been spotted in control subjects (1 to 4), except for one of the missense mutation, c.871G>A – G291S, which had not been previously reported in the literature as a cause of IRIDA. The presence of all coding region mutations was confirmed by Sanger sequencing of the exons in which they were present. A subject with each of the three possible genotypes was selected for variant confirmation.

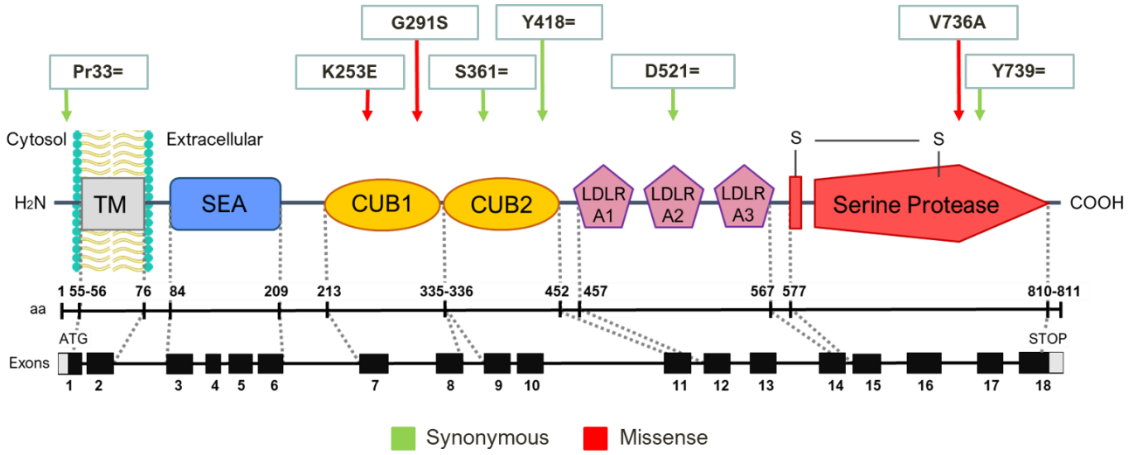


Figure 3.19 – Protein and gene and locations of found variants in the coding regions of the *TMPRSS6* gene. Identified IRIDA mutations are indicated in boxes, with green and red arrows indicating the location of the synonymous and missense mutations, respectively.

One can find polymorphisms all over the genome, rather than just in coding or splicing regions, in this way, a considerable number of SNPs are expected to be located within regulatory sequences, as can be the case of some of the found deep intronic variants. The potential effects of these variants on gene regulation depend on their location respect to the regulatory elements, which are always difficult to assess [226]. Because regulatory regions are difficult to identify and study, there were no available primers for Sanger sequencing confirmation of called variants, and a final IRIDA diagnosis requires only the identification of genetic lesions within exons and exon-intron boundaries of the *TMPRSS6* gene, deep intronic variants were not extensively studied in this dissertation.

3.2.2 *In silico* studies of pathogenicity of found variants

Analysis of possible splice site mutations: A precise pre-mRNA splicing is essential for an appropriate protein translation, in which consensus *cis*-sequences that define exon-intron boundaries and regulatory sequences are recognized by splicing machinery. The presence of mutations at these sequences can be responsible for improper exon and intron recognition and may result in the formation of an aberrant transcript of the mutated gene [227].

Synonymous mutations can change the sequence of a gene without directly altering the encoded protein. Due to the absence of protein change, synonymous mutations are often classified as neutral by mutation prediction tools and are sometimes ignored in this type of analysis. Yet, synonymous mutations can have an important role in mRNA splicing, folding, mRNA interaction and protein translation [228]. Taking that into account, the called synonymous variants were

analyzed together with intronic variants, using the HSF software, to predict the effects of mutations on splicing signals or to identify splicing motifs in the mutated *TMPRSS6* sequence.

According to the *in silico* splicing analysis, all synonymous and exon-intron boundary variants had either "Probably no impact on splicing" or "No significant splicing motif alteration detected". Therefore, we conclude that these render no pathogenicity when it comes to the splicing process. As for the deep intronic variant, results were similar to the other tested variants except for 3 variants that the software classified with "Potential alteration of splicing", being able to create exonic cryptic splice sites with potential for creating an aberrant protein [228]. No further investigation was conducted regarding these deep intronic variants, which the presence could not be confirmed by Sanger sequencing.

Analysis of missense mutations: The found missense mutations, two common polymorphisms and one novel undescribed variant, were analyzed by three online prediction software to determine their impact in the final peptide based on the differences between human matriptase-2 and the homologous proteins from other mammals (PolyPhen-2), the importance of one aa based on its level of conservation within the protein family (SIFT); and the 3D protein structure (Missense3D).

Sequence variation data of the human proteome can be used to analyze 3D protein structures to derive functional insights, and therefore be used to predict the phenotypic consequence of a missense variant [145], [229]. Seeing that in the online databases there was no available 3D structure model of matriptase-2, we used Phyre2 homology modelling tool to produce the 3D structure of the protein based on alignment to known protein structures, obtaining an amino acids structure with 88 % of residues modelled with 90% of confidence (Figure 3.20).

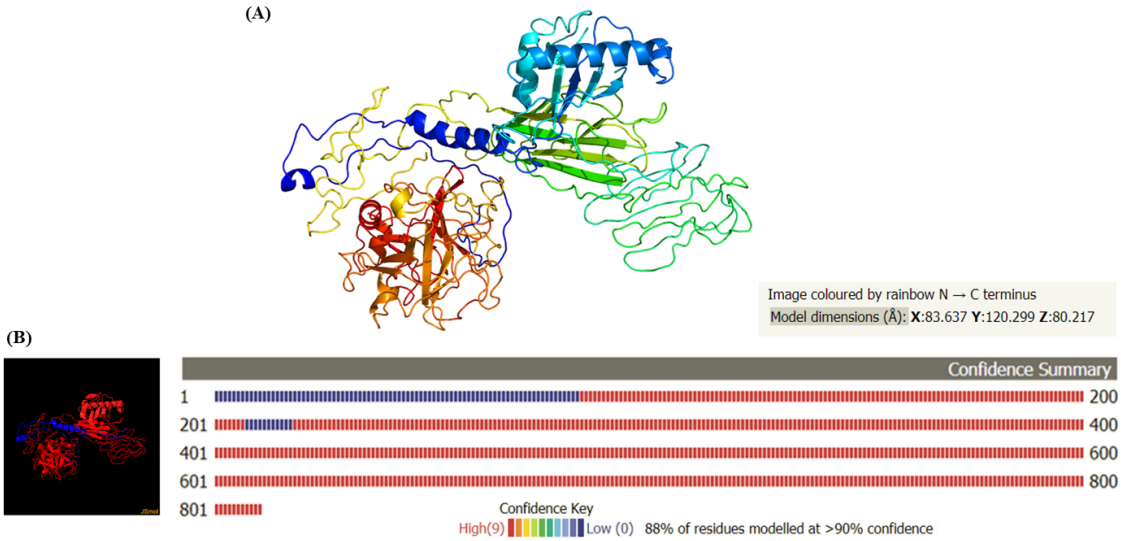


Figure 3.20 – Phyre2 protein model. (A) Ribbon representation of matriptase-2 3D predicted structure colored by rainbow N to C model. **(B)** Confidence of residue prediction.

V736A: This variant, located in the serine protease domain of the protein, is the result of the substitution of valine for alanine, and has been several times associated with changes in the hematological and iron parameter [57], [230], [231], with the allele T being established as the risk allele. Furthermore, functional studies of this variant reveal no differences in the shedding and proteolytic activity between variant V736A and the wild-type protein [232].

The valine and alanine amino acids are similar since both are non-polar capable of having van der Waals interactions, with very non-reactive side chains. For these reasons, they are rarely directly involved in protein functions like catalysis, although they can play a role in substrate recognition. [226]. We put to test the pathogenicity that this variant entailed for the protein, results from Poly-Phen2 predicted V736A to be benign for the protein (Figure 3.21). Additionally, to been predicted as benign, the variant was also classified as tolerated (0.73; ≥ 0.05 considered neutral) according to SIFT, and no structural damage resulting from the presence of the variant was detected with Missense3D.

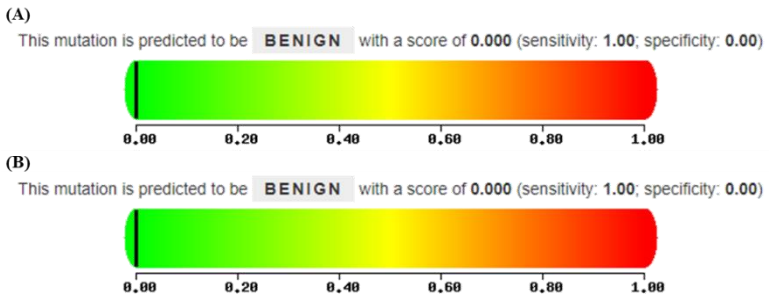


Figure 3.21 – PolyPhen-2 pathogenicity prediction of variant V736A on matriptase-2. Variant predicted as benign according to the score obtained for predictions (**A)** based on the HumDiv data and **(B)** the HumVar data.

The fact that the gene encoding for this protein is highly conserved across mammalian species [49], [52], would make one believe that the wild-type allele would be the conserved one. Instead, the conservation profile, resulting in a multiple sequence alignment with homologous protein sequences from closely related mammalian species (Figure 3.22), reveals that alanine is the conserved one.

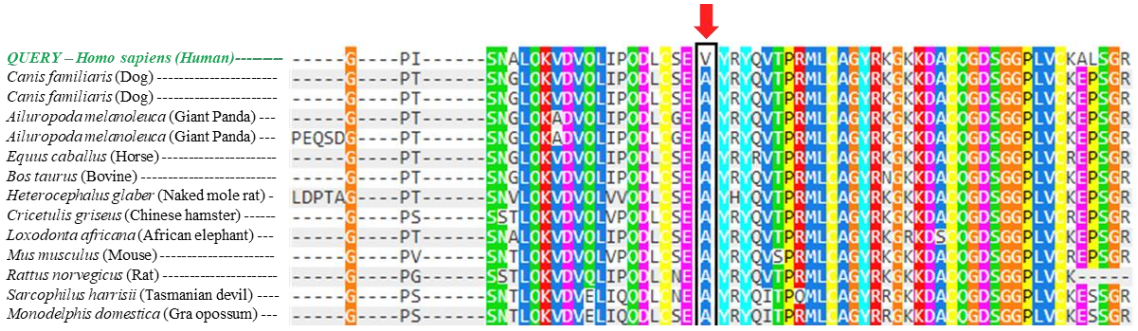


Figure 3.22 – Conservation profile of matriptase-2 at the position of the variant V736A. Multiple sequence alignment of protein containing variant and homologous protein sequences of closely related mammalian species. The pattern of amino acid substitutions is inside a black box and highlighted with red arrow. Results obtained from PolyPhen-2. **V** – Valine; **A** – Alanine.

The results support the data present in the literature, stating that this variant is not responsible for an IRIDA phenotype. On the contrary, the variant was predicted as benign when compared to the wild-type allele.

K253E: Lysine, a somewhat amphipathic amino acid, frequently plays an important role in protein structure and is quite often present in protein active or binding sites for being able to form salt-bridges, hydrogen bonds and van der Waals interactions. In the variant K253E present in the CUB1 domain, lysine is substituted by glutamic acid, although this change entails the change of a positive to a negative aa, both of them share the same biochemical properties such as having elevated side-chain flexibility and establishing the same molecular interactions [226]. Results from *in silico* analysis from Poly-Phen2 predict that K253E has a benign effect on the protein (Figure 3.23). This result is corroborated by the software SIFT that predicts the variant as tolerated

(1.00; ≥ 0.05 considered neutral) and Missense3D that concludes that its presence leads to no structural damage.

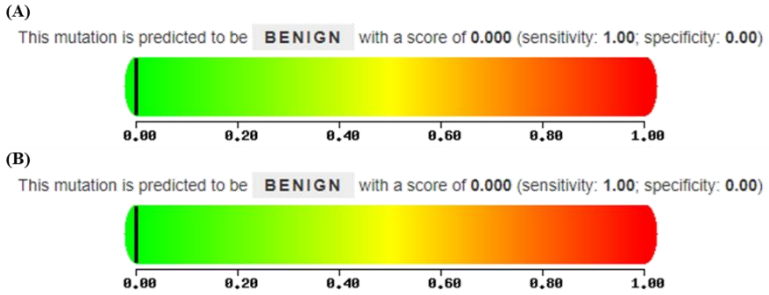


Figure 3.23 –PolyPhen-2 pathogenicity prediction of variant K253E on matriptase-2. Variant predicted as benign according to the score obtained for predictions (**A)** based on the HumDiv data and **(B)** the HumVar data.

Additionally, the conservation profile (Figure 3.24), also obtained from PolyPhen-2, reveals that the wild-type lysine present in the human protein appears not to be especially conserved across other close related mammalian species. The profile displays that the lysine residue can be substituted by glutamic acid, such as what happens in the variant in study, and by glutamine, which is known to be frequently substituted by other polar amino acids, such as the previously mentioned glutamate [226].

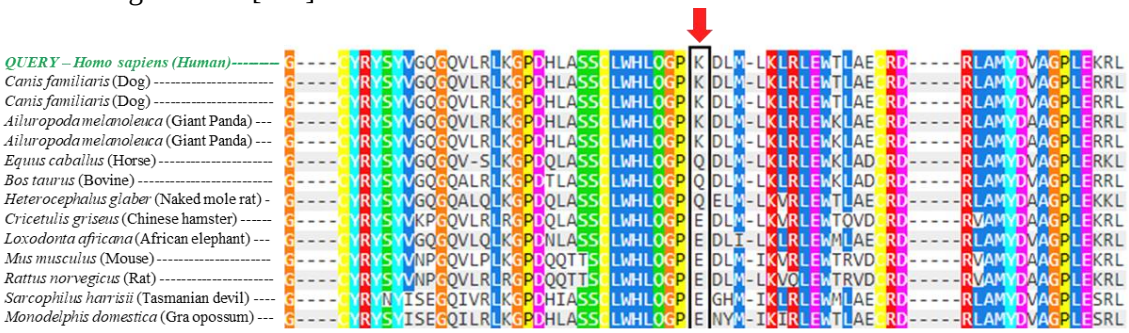


Figure 3.24 – Conservation profile of matriptase-2 at the position of the variant K253E. Multiple sequence alignment of protein containing variant and homologous protein sequences of closely related mammalian species. The pattern of amino acid substitutions is inside a black box and highlighted with red arrow. Results obtained from PolyPhen-2. **K** – Lysine; **Q** – Glutamine; **E** – Glutamic acid.

Originally thought to be responsible for a non-severe IRIDA phenotype [233], this common variant has never been associated with changes in hematological and iron parameters [234]–[236]. By predicting that K253E is a benign tolerated non- damaging variant, we conclude that this variant is not likely to be responsible for an IRIDA phenotype, corroborating the conclusions reached by Donker *et al.* [237].

G291S: This missense variant, which has never been described before to be in the origin of an IRIDA phenotype, is caused by the substitution of glycine for serine in the CUB1 domain of matriptase-2. Glycine's structure is unique for containing hydrogen as its side chain, which leads the residue to reside in parts of protein structures that are forbidden to all other amino acids, such as the tight turns in the peptide structures. Furthermore, this residue plays a distinct functional role by using its backbone, instead of its side chain, to bind to phosphates [226]. Contrarily, serine is a somewhat indifferent amino acid that can reside both within the interior of a protein or on the protein's surface. This aa also differs from glycine by being polar and being able to establish up to three hydrogen bonds, rather than just van der Waals interactions.

When attempting to determine the impact of G291S in the final peptide, PolyPhen-2 predicted the variant as a possibly damaging mutation (Figure 3.25), and so did the software SIFT that predicts the variant as deleterious (0.00; ≥ 0.05 considered neutral).

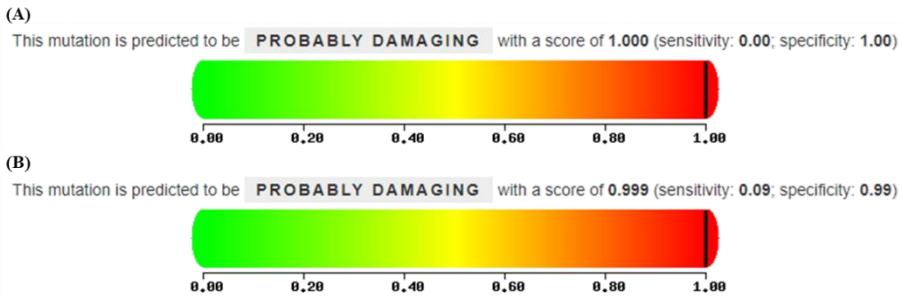


Figure 3.25 – PolyPhen-2 pathogenicity prediction of variant G291S on matriptase-2. Variant predicted as probably damaging according to the score obtained for predictions (**A)** based on the HumDiv data and **(B)** the HumVar data.

The conservation profile (Figure 3.26) analysis revealed that the glycine residue present at the position 291 is highly conserved among close related mammalian species. In case one sees a highly conserved glycine changing into another amino acid, it will mean that the change could have a drastic impact on protein function. This type of impact has already been described among protein kinases [238].

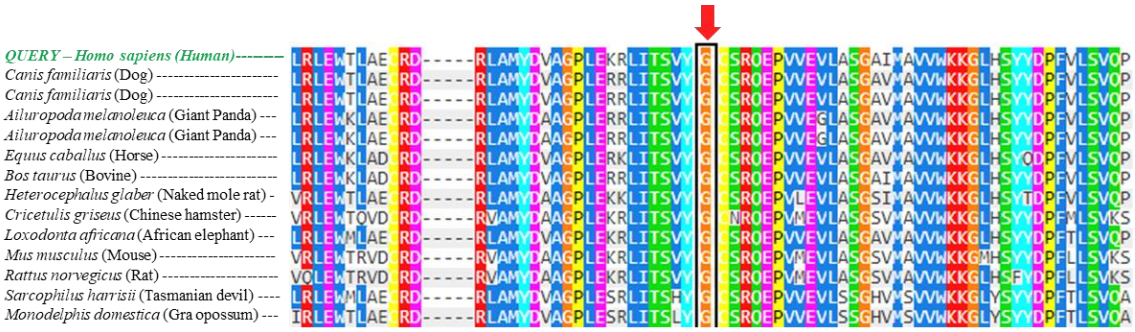


Figure 3.26 – Conservation profile of matriptase-2 at the position of the variant G291S. Multiple sequence alignment of protein containing variant and homologous protein sequences of closely related mammalian species. The pattern of amino acid substitutions is inside a black box and highlighted with red arrow. **G** – Glycine.

Also, based on the predicted structure for MT2, an analysis to determine the nature of the pathogenicity of G291S substitution was conducted using the software Missense3D (Figure 3.27). The software detected that the change of residues would potentially lead to two structural damages "Gly in a bend" and "Cavity altered". The former means that the wild-type residue is glycine and is in a bend curvature since it only has hydrogen as its side chain, glycine can adopt a far larger region in the protein backbone than other side chains, making it a characteristically found residue of loops and regions where a polypeptide chain makes a sharp turn [145]. For this reason, glycine is often conserved within a protein family for maintenance of a certain protein fold, and so its substitution to any amino acid is perceived as damaging.

If the structural damage detected is an altered cavity, also referred to as a pocket on the protein surface, the software means to suggest that the substitution leads to an expansion or contraction of the cavity volume of ≥ 70 $Å^3$ [145]. In general, residues that enclose cavities do not display any extra local mobility relative to their surrounding environments [239], particularly, glycine is far more conserved if located at a complemented pocket than if located within the rest of the interface for the lack of a side-chain that allows for tight packing [240]. The predicted changes in cavity volume can affect protein stability.

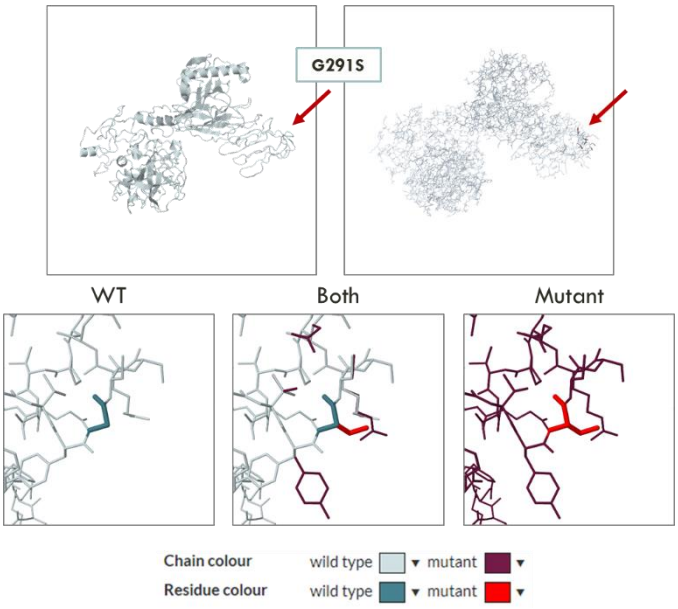


Figure 3.27 – Structural analysis: comparison between wild-type and mutant G291S.

The change of a vital residue within a protein domain is likely to disrupt the protein function by not allowing the protein to maintain its conformation or perform its biological functions, leading to a pathological phenotype [241] Despite the conserved structure of the proteins within the TTSP family, little is known about the specific functional contributions of the SEA, CUB, and LDLR domains in these proteins [242]. Matriptase-2 is expressed as an inactive zymogen, which needs a proteolytic cleavage to become active. Afterwards, the protease domain remains attached to the membrane anchor by a disulfide bond [243]. Oberst *et al.* has shown that matriptase activation requires glycosylation of the serine protease and CUB1 domains, strongly suggesting that matriptase glycosylation can differentially influence the level of protease activation [244], in the same way, Silvestri *et al.* suggested that the integrity of the CUB and LDLR domains is required for the activation of MT2 [245]. Furthermore, McDonald *et al.* showed that the SEA and LDLA domains support the protein transport to the cell surface, while the CUB, LDLR and protease domains facilitate the cleavage of its endogenous substrate hemojuvelin [242]. However, deletion of both CUB domains increases zymogen activation, suggesting that the CUB domains together may serve to prevent premature matriptase activation, but may also provide the structural basis for protein-protein interactions which are believed to be important for matriptase zymogen activation [246]. In this way, mutations within the CUB1 domain, such as G291S, can potentially have a nefarious effect on matriptase zymogen activation and subtract recognition.

When analyzing the pathogenicity of a novel missense mutation (c.871G>A; G291S), found in heterozygosity in one of the cases with suspicion of IRIDA, the obtained *in silico* results, presented as pathogenicity scores and functional data, reach the same conclusion that the variant is deleterious and potentially damaging for the protein stability. Moreover, by reviewing the function of the protein domain in which the mutation is inserted, we believe that the residue change may also affect the enzyme activation and substrate recognition. For the reasons stated above, is fair to assume that the presence of G291S is most likely, at least partially, the genetic cause for the phenotype of the patient in which it was found. In the future, actual functional and structural studies should be performed to confirm *in silico* results and further the knowledge about matriptase-2 and this novel mutation.

3.2.3 Case 6 to 8 and 10 – Unexplained phenotypes

In four of the not previously analyzed subjects we found a combination of the same variants found in the control subjects (Table 3.15). Of the seven found variants, two (Y418Y and P33P) have not been yet associated with either an IRIDA phenotype or any changes in the hematological and iron parameters. The two common missense variants, V736A and K253E, were previously described in this study as benign, with the former's wild-type allele being associated with a higher risk of IDA [57], [230], [231]. Similarly, the variant Y739Y risk allele has also been associated with a reduction in Hb and Ft concentrations and an increase in Tf in most populations investigated, as well as having a greater frequency in woman TSAT < 10 % [50], [234]. Regarding D521D, this variant was associated with a lower Ft level even after controlling for iron intake, with Shinta *et al.* showing that per copy of the minor allele (C), the Ft serum concentration was reduced by 5.00 μg/L [247]. Moreover, this variant is said to be associated with the levels of iron-related hematological parameters [236]. Finally, S361S has shown strong significant association with SI and TBIC, these results were obtained from subgroups of a multiethnic population suggesting that the variant may play different roles in different ethnicities [248].

Table 3.15 – Identification of variants present in the *TMPRSS6* gene of unexplained cases.

S_ID	GT	rs_ID	Location	MC	Region	NM_153609.3	NP_705837.1
6;7	0/1	**rs2235321**	22:37066886-37066886	Syn	Ex17	c.2217C>T	**p.Tyr739=**
6;7;8	0/1	**rs855791**	22:37066896-37066896	Miss	Ex17	c.2207T>C	**p.Val736Ala**
7;8	0/1	**rs4820268**	22:37073551-37073551	Syn	Ex13	c.1563C>T	**p.Asp521=**
6	1/1	rs881144	22:37075250-37075250	Syn	Ex11	c.1254C>T	p.Tyr418=
7	0/1	**rs2111833**	22:37084757-37084757	Syn	Ex9	c.1083G>A	**p.Ser361=**
6;8;10	0/1	**rs2235324**	22:37089684-37089684	Miss	Ex7	c.757A>G	**p.Lys253Glu**
7	1/1						
6	1/1	rs11704654	22:37103346-37103346	Syn	Ex2	c.99G>A	p.Pro33=

Bold – Relevant variants; **S_ID** – Subject ID; **rs_ID** – Reference single nucleotide polymorphism ID; **NM** – Reference sequence based on a protein coding RNA (mRNA); **NP** – Reference sequence based on a protein (amino acid) sequence; **MC** – Molecular consequence; **Miss** – Missense; **Syn** – Synonym; **Region** – Genomic Region; **GT** – Genotype; **0/1** – Heterozygous; **1/1** – Homozygous.

In whole, we believe that even the simultaneous presence of the found polymorphisms is not responsible for IRIDA, a hereditary autosomal recessive anemia. However, the subjects present with polymorphisms that have been associated with a greater risk of developing iron deficiency anemia. These polymorphisms may be modulating the subjects' phenotypes. However, each of the *TMPRSS6* variant explains only approximately 1 % of the variance in iron concentrations [249]. Since these subjects' clinical picture cannot be justified by the molecular findings, we suggest further investigation into their iron metabolism, with hepcidin quantification being the next logical step. When choosing the next steps for molecular investigations, one should be aware that the molecular techniques used in this study do not allow the detection of some INDELS, genetic rearrangements and epigenetic modifications.

3.2.4 Case 9 – IRIDA-like phenotype

Case 9 was a 4-year-old boy of sub-Saharan ancestry (Mozambique/Angola), presenting with microcytic hypochromic anemia, low TSAT, normal Ft, and having a partial response to intravenous iron treatment (Table 3.16).

Table 3.16 – Hematological and iron status features for Case 9 before and after intravenous iron treatment.

ID	Sex /Age	Iron treatment	RBC ($x10^{12}$/L)	Hb (g/dL)	MCV (fL)	MCH (pg)	RDW (%)	SI (µg/dL)	Tf (mg/dL)	TIBC (µg/dL)	TSAT (%)	Ft (µg/L)
9	M	Before	4.52	**9.8**	**70.3**	**21.7**	**17.7**	**19**	253	316.67	**6.00**	33.0
	4 y	After	4.38	**10.2**	**75.6**	**23.3**	**16.6**	**51**	218	273.02	18.68	**242.3**

Bold – values outside normal range; **M** – Male; **y** – Years; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Before the screening of *TMPRSS6* by NGS, the subject was already known to be an $\alpha^{3.7}$-thalassemia carrier, a contributing factor to the microcytic hypochromic anemia phenotype. The analysis revealed the presence of a novel missense mutation (c.871G>A; G291S) in heterozygosity, in *TMPRSS6* exon 8. *In silico* analysis indicated that the conserved amino acid change potentially results in damage to the protein stability. Due to its location in the CUB1 domain, it may also affect enzyme activation and substrate recognition. Additionally, 3 SNPs previously associated with a greater risk of developing iron deficiency anemia (K253E; S361S; and Y739Y) were also identified in *TMPRSS6* (Table 3.17).

Table 3.17 – Identification of variants present in the *TMPRSS6* gene of Case 9.

S_ID	rs_ID	Location	MC	Region	NM_153609.3	NP_705837.1	GT
9	rs2235321	22:37066886-37066886	Syn	Ex17	c.2217C>T	**p.Tyr739=**	0/1
	rs855791	22:37066896-37066896	Miss	Ex17	c.2207T>C	p.Val736Ala	1/1
	rs2111833	22:37084757-37084757	Syn	Ex9	c.1083G>A	**p.Ser361=**	1/1
	rs145053404	22:37086412-37086412	Miss	Ex8	c.871G>A	**p.Gly291Ser**	0/1
	rs2235324	22:37089684-37089684	Miss	Ex7	c.757A>G	**p.Lys253Glu**	1/1
	rs11704654	22:37103346-37103346	Syn	Ex2	c.99G>A	p.Pro33=	0/1

Bold – Relevant variants; **S_ID** – Subject ID; **rs_ID** – Reference single nucleotide polymorphism ID; **NM** – Reference sequence based on a protein-coding RNA (mRNA); **NP** – Reference sequence based on a protein (amino acid) sequence; **MC** – Molecular consequence; **Miss** – Missense; **Syn** – Synonym; **Region** – Genomic Region; **GT** – Genotype; **0/1** – Heterozygous; **1/1** – Homozygous.

IDA is the most common cause of anemia and is mainly caused by nutritional deficiency. Contrarily, IRIDA is known as an autosomal recessive disease and therefore less frequent. In this case, we conclude that the result of digenic inheritance of the novel damaging mutation (c.871G>A; G291S) and the 3 common modulating SNPs in the same gene and a co-inheritance of an α-thalassemia trait deletion may lead to an IRIDA-like phenotype.

Later in the investigation of this case, the DNA sample of the child's younger brother, also an $\alpha^{3.7}$-thalassemia carrier, presenting with hypochromic anemia (Table 3.18), characteristic of the genetic trait, arrived for molecular investigation of the *TMPRSS6* gene. The found and confirmed mutations in the older brother where searched for by Sanger sequencing of the exons in which they were present, confirming the presence of 3 relevant common polymorphisms (K253E – 1/1; V736A – 0/1; and Y739Y – 1/1).

Table 3.18 - Hematological and biochemical features in subjects tested for some IRIDA causing mutations.

ID	Sex	Age	RBC ($x10^{12}$/L)	Hb (g/dL)	MCV (fL)	MCH (pg)	RDW (%)	SI (µg/dL)	Tf (mg/dL)	TIBC (µg/dL)	TSAT (%)	Ft (µg/L)
11	M	22 m	4.42	**10.1**	73.8	**22.9**	18.0	65.0	273.0	341.00	19.06	16.3

Bold – values outside normal range; **M** – Male; **m** – Month; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

The absence of the potentially damaging missense mutation G291S in the younger brother leads to the conclusion that his phenotype is, most likely, just a consequence of the α-thalassemic trait, and further supports our conclusion that this mutation, even in the heterozygous state, is predominantly responsible for Case's 9 phenotype. Through the complete and partial analysis of the *TMPRSS6* gene, of both brothers respectively, we can infer the parents' genotype for some of the present variants. However, a complete family study should be conducted to determine from which parent the child has inherited the deleterious mutation so that assumptions can be made about the effects of this mutation on the subjects' phenotype in infancy and adulthood.

In conclusion, to complement the findings of this study, further functional and structural studies of the mutated protein as well as family studies, including the parents, should be conducted, in order to increase the knowledge of the iron metabolism and, ultimately, allow the identification of potential genetic markers for IDA. Moreover, this work suggests that carrying out an NGS analysis is an effective strategy to diagnose and study the rare disease IRIDA, especially as compared with conventional sequencing techniques.

3.3 Genetic association study

Genetic association studies (GAS) assess the association between phenotypic traits and genetic variants in a population. This type of analysis investigates the association without requiring information on inheritance, and therefore, is conducted on a sample of unrelated cases and controls, with studied variants being selected based on pathophysiological hypotheses [250]. A GAS was performed to investigate the role of four common variants in genes *TMPRSS6* [c.2207T>C, p.Val736Ala (V736A) and c.757A>G, p.Lys253Glu (K253E)] and *HFE* [c.187C>G, p.His63Asp (H63D) and c.845G>A, p.Cys282Tyr (C282Y)], associated with the iron metabolism, on the hematological and iron status parameters of patients from two different types of anemia (BTT and IDA) and healthy controls.

Demographic, hematological and iron status characteristics of controls and cases are reported in Table 3.29. In the results it is observable, as expected, that the mean values of all parameters of the control group are within normal ranges for both sexes. When considering the mean values of the BTT group, all hematological parameters are outside the normal range, corresponding with the pathophysiology in question. When it comes to the iron status of this group, only Ft was elevated, when considering the normal range, which is not surprising since some BTT subjects can develop iron overload due to dyserythropoiesis that leads to a variable increase of enteric iron absorption [251]. For this reason, is important to evaluate the iron status in BTT subjects since it may play a role in improving iron status in females, and in men can more easily lead to iron overload. Monitoring it could then avoid the harmful effects of iron overload in the early stages of the disorder in men [252]. Concerning the mean values of the IDA group parameters, they are also out of range, as expected for the iron deficiency anemia pathophysiology. Regarding the demographic, hematological and iron status profiles, when comparing the control and BTT groups, two iron status parameters, SI (p = 0.600) and TSAT (p = 0.108) do not show any statistically significant difference and are lower in the BTT group. All parameters were significantly different between IDA patients and healthy controls ($p < 0.001$).

Table 3.19 – Demographic, hematological and iron status parameters in control and case populations.

	Controls (N=134)	BTT (N=116)	IDA (N=67)	Mann-Whitney test	
Parameters	Mean ± SD	Mean ± SD	Mean ± SD	p^a	p^b
Gender (M/F)	42/92	63/53	10/57	**< 0.001**	**0.016**
Age (y)	39±16	45±18	44±16	**0.041**	**0.012**
RBC (x10^12/L)	4.34±1.24	5.70±0.61	4.45±0.43	**< 0.001**	**0.047**
Hb (g/dL)	13.74±4.52	11.82±1.20	10.23±1.45	**< 0.001**	**< 0.001**
MCV (fL)	91.45±1.93	64.80±4.21	72.40±7.67	**< 0.001**	**< 0.001**
MCH (pg)	31.71±1.56	20.79±1.39	23.11±3.35	**< 0.001**	**< 0.001**
RDW (%)	12.67±41.44	15.55±4.33	24.11±12.47	**< 0.001**	**< 0.001**
HbA2 (%)	-	4.13±0.43	-	-	-
SI (µg/dL)	113.53±42.41	108.06±31.05	33.35±17.68	0.600	**< 0.001**
Tf (mg/dL)	285.57±53.00	260.18±41.05	376.07±70.29	**< 0.001**	**< 0.001**
TIBC (µg/dL)	356.96±12.70	326.97±54.90	470.04±87.81	**< 0.001**	**< 0.001**
TSAT (%)	32.48±73.24	33.88±10.73	7.41±4.54	0.108	**< 0.001**
Ft (µg/L)	81.03±12.70	169.36±205.21	8.61±10.82	**< 0.001**	**< 0.001**

[a] – Comparison between the parameters of the total number of subjects in the control and BTT groups; **[b] –** Comparison between the parameters of the total number of subjects in the control and IDA groups; **BTT** – Beta thalassemia trait; **IDA** – Iron deficiency anemia; **SD** – Standard deviation; ***p* – *p*-value (significant < 0.05); y** – Years; **m** – Month; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Additionally, and based on the knowledge that the used hematological and iron status measurements differ for sex and age [39], [147], the differences between sexes were analyzed in the studied populations (Table S.21). The small sample size did not allow stratification of the population according to age groups. Statistically significant differences were observed for the RBC, Hb and Ft parameters ($p < 0.001$), of both control and BTT groups. Furthermore, when comparing sex, the TSAT levels differed within the control group ($p = 0.048$), as well as the Tf and TBIC levels ($p = 0.029$) within the BTT group. Within the IDA group, there were no statistically significant differences between sex for the hematological and iron status parameters, probably due to the small number of men within this group (10 males; 57 females), the only observable difference was between the mean age of the two groups ($p = 0.028$). The results from these comparisons highlight the importance of adjusting, when possible, the statistical association methods, carried out later in this work, for sex and age, to avoid confounding, by eliminating plausible alternative explanations for an observed relationship between independent and dependent variables.

To provide a surrogate of GAS quality, in terms of design and conduct, it was checked whether the genotype frequencies were in conformity with Hardy-Weinberg equilibrium [250]. In theory, disease-free control groups from outbred populations, as well as combined groups of cases and controls from studies where all subjects have a specific disease, should follow the HWE [253]. Deviations from it lead to type I error (false positive) for gene-disease associations and can be due to genotyping errors, population stratification, inbreeding, selection bias in the choice of controls and confounding factors unaccounted for [250], [253].

There was no difference in the frequency distribution of all studied gene variants between the controls and the two case groups, and in all three groups frequencies agreed with the HWE (Table 3.20). Compliance with the HWE indicates that the genotype frequencies are consistent with the two alleles being independently sampled from a population of alleles [254]. Normally, in the presence of a marker to disease susceptibility allele, the HWE is not expected to hold, yet, for association with a recessive susceptibility allele, HWE may hold [254]. In all the investigated groups there was a low frequency of the disease risk allele for the variants of the *HFE* gene, further justifying why the equilibrium was preserved for these variants.

When looking at the allele frequencies of the 3 groups, on the one hand, the *TMPRSS6* variants presented a lower minor allele frequency than the European population; V736A (28.85-31.30 < 39.0 %) and K253E (24.05-26.95 < 43.0 %), on the other hand, *HFE* variants presented similar minor allele frequency values to the ones described for the European and Portuguese population; H63D (12.70-19.90 %) and C282Y (2.06-4.05 %) [80], [81].

In case-control studies, in which the ratio of cases to controls is controlled by the investigator, it is not possible to make direct estimates of disease penetrance, illustrated by the risk ratio (RR) [255]. In this type of study, the magnitude and significance of the association can be estimated with the odds ratio (OR) and its 95 % confidence interval (CI) for the various genetic contrasts, representing the probability that the disease is present compared with the probability that it is absent [250], [255]. Advantageously, when disease penetrance is small, there is little difference between RRs and ORs [255].

Table 3.20 – Frequency distribution of *TMPRSS6* V736A and K253E and *HFE* H63D and C282Y gene variants. Frequencies expressed as % in the three groups control, beta-thalassemia trait, and iron deficiency anemia.

Locus, SNP (risk allele) Genotypes / Alleles	Controls (N=134)	BTT (N=116)	IDA (N=6)	Fisher's p^a	Fisher's p^b
TMPRSS6, V736A (T)					
TT	6.50	10.87	6.90		
TC	49.59	35.87	48.28	0.112	0.985
CC	43.90	53.26	44.83		
T/C	31.30/68.70	28.85/71.25	31,05/68.95	0.878	0.878
HWE (χ2, p)	15,000 (0.451)	6.333 (0.387)	6.333 (0.387)		
TMPRSS6, K253E (G)					
AA	61.34	65.43	67.31		
AG	23.53	16.05	17.31	0.418	0.677
GG	15.13	18.52	15.38		
A/G	73.05/26.95	73.40/26.60	75.95/24.05	0.955	0.746
HWE (χ2, p)	6.333 (0.387)	6.333 (0.387)	9,000 (0.532)		
HFE, H63D (G)					
CC	64.66	68.10	69.84		
CG	30.83	31.90	25.40	0.070	0.735
GG	4.51	0.00	4.76		
C/G	80.10/19.90	84.05/15.95	87.30/12.70	0.581	0.253
HWE (χ2, p)	12.000 (0.285)	2.000 (0.572)	12.000 (0.285)		
HFE, C282Y (A)					
GG	92.54	94.83	95.24		
GA	6.72	5.17	4.76	0.888	0.765
AA	0.75	0.00	0.00		
G/A	95.95/4.05	97.94/2.06	97.60/2.40	0.683	0.683
HWE (χ2, p)	6.000 (0.423)	4.000 (0.261)	4.000 (0.261)		

[a] – Comparison between the allele and genotype frequency of variants of the total number of subjects in the control and BTT groups; **[b] –** Comparison between the allele and genotype frequency of variants of the total number of subjects in the control and IDA groups; **BTT** – Beta thalassemia trait; **IDA** – Iron deficiency anemia; **SNP** – Single- nucleotide polymorphism; ***p* – *p*-value (significant < 0.05); HWE** – Hardy Weinberg equilibrium; **χ2** – Qui squared.

In Table 3.21 are displayed the genotype distributions of the four studied polymorphisms from *TMPRSS6* and *HFE* genes in control and IDA patients. An OR of less than 1 means that the first group was less likely to experience the event. After adjusting for sex and age, statistically significant OR values were obtained for all variants revealing that all the polymorphisms had an association with the IDA status when compared to the same genotypic exposure in controls. Looking at V736A, both heterozygous (TC) and homozygous (CC) genotypes present themselves as a protective factor against IDA, with subjects with these genotypes being 74 % less likely to develop IDA (OR = 0.26, 95% CI [0.09,0.68], p = 0.008; OR = 0.26, 95% CI [0.09,0.72], p = 0.010). These results reveal that the presence of only one protective allele (C), in heterozygous individuals, is enough to render a protective effect against IDA. In the literature, it has already been described that the T allele appears to inhibit hepcidin more efficiently than the C allele in *in vitro* experiments [55].

Concerning the K253E variant only the homozygote for the risk allele (GG) did not show any association with the disease outcome, with subjects presenting the wild-type (AA) and heterozygous (AG) genotype being 67 % and 82 %, respectively, less likely to present with IDA

(OR = 0.33, 95% CI [0.12,0.88], p = 0.030; OR = 0.18, 95% CI [0.05,0.59], p = 0.006). Sato *et al.* associated this variant with a non-severe IRIDA phenotype in a Japanese child [233], but Donker *et al.* argued that it was not very likely that the presence of K253E, simultaneously with V736A, was responsible for the clinical phenotype of the child [237]. Since then, no other author made an association between this variant and the hematological and iron parameters [234]–[236], although Lee *et al.* has found that K253E frequency was greater in women with TSAT ≥ 10 % [234].

For the H63D variant, only the homozygote for the iron overload risk allele (GG) did not show any association with the anemia outcome, with subjects presenting the wild-type (CC) and heterozygous (CG) genotype being 73 % and 81 %, respectively, less likely to present with IDA (OR = 0.27, 95% CI [0.19,0.65], P = 0.005; OR = 0.19, 95% CI [0.06,0.55], p = 0.003). The C282Y wild-type genotype (GG) is also 75 % less likely to present with IDA (OR = 0.25, 95% CI [0.10,0.61], p = 0.003). The OR results for the *HFE* variants represent an inverse association to what was expected since the presence of the risk alleles of these polymorphisms leads to hemochromatosis by affecting positively the iron uptake and storage [256], therefore, the heterozygous and homozygous genotypes, for both mutations, should be the ones conferring protection against IDA, instead of the wild-type and heterozygous states.

Table 3.21 – Distribution of V736A, K253E, H63D and C282Y genotype frequencies in IDA patients and controls

Genotype	WT			HET			HOM		
SNP	C/IDA	OR (95% CI)	*p*	C/IDA	OR (95% CI)	*p*	C/IDA	OR (95% CI)	*p*
V736A	8/4	0.23 (0.04,0.98)	0.056	61/28	0.26 (0.09,0.68)	**0.008**	54/26	0.26 (0.09,0.72)	**0.010**
K253E	73/35	0.33 (0.12,0.88)	**0.030**	28/9	0.18 (0.05,0.59)	**0.006**	18/8	0.35 (0.10,1.18)	0.095
H63D	86/44	0.27 (0.10,0.65)	**0.005**	41/16	0.19 (0.06,0.55)	**0.003**	6/3	0.17 (0.02,1.04)	0.066
C282Y	124/60	0.25 (0.10,0.61)	**0.003**	9/3	0.39 (0.07,1.96)	0.262	1/0	NC	0.986

C – Controls; **IDA** – Iron deficiency anemia; **SNP** – Single- nucleotide polymorphism; **WT –** Wild-type; **HET** – Heterozygous; **HOM –** Homozygous; **OR –** Odds ratio, from logistic regression models, adjusted for sex and age (continuous); **CI** – Confidence interval; ***p* –** *p*-value (**significant < 0.05**); **NC** – Not calculated.

The association results from this analysis can only be taken into consideration up to a point, since all results present statistically significant or borderline significant results, which may be a consequence of the small frequency of risk alleles in the population, especially regarding the *TMPRSS6* polymorphisms that are reported to be present in higher frequencies in IDA patients [235], and the fact that there is roughly twice the amount of control subjects when compared to the IDA group (134 controls; 67 IDA), and can be possibly responsible for the unexpected results regarding the *HFE* gene. This analysis was not performed for control *versus* BTT because, while changes in the genes involved in the iron metabolism can directly influence whether subjects develop or not iron deficiency anemia, these changes can modulate the iron status in BTT subjects but cannot influence whether they are or not carriers of β-thalassemia.

GAS usually assesses various contrasts of genotypes such as the allele contrast, the additive, recessive and dominant models. These models are attractive for researchers because the sample size becomes larger, generating more power to detect significant associations [250]. In this study, to assess the status of a single SNP we tested the null hypothesis of no association between the three genotypes (the two homozygotes and the heterozygote) among cases and controls. Additionally, when it was allowed by the sample size, the K253E genotypes were compared according to the recessive model, and the *HFE* variants according to the dominant model. The OR analysis revealed that, although IRIDA is an autosomal recessive disease caused by variants

within the *TMPRSS6* gene, the presence of only one alternative allele (C) of V736A is enough to confer protection against IDA, not showing a great protective difference from homozygous genotype for the variant (CC). So, neither the recessive nor dominant models could be applied in the genotypic comparisons, and consequently, V736A genotypes were compared with each other individually.

When analyzing the V736A variant in the female control group we found that MCV (p = 0.048), MCH (p = 0.048) and TSAT (p = 0.048) parameters were significantly increased in subjects with the homozygous genotype (CC) when compared to the wild type (TT), (Table S.22). Tanaka *et al*. had already reported increased values of Hb, MCV and Ft in subjects with the C allele [249], and inversely Benyamin *et al*. has found decreased Hb, MCV, TSAT and Ft in subjects with the T allele [57]. Like in our study, Chamber *et al.* and Kullo *et al.* have found decreased values of MCV in subjects with the T allele [257], [258]. All in all, our results are in line with what was described in the literature. The same was not observed for the male control subjects (Table S.22) and the subjects of both sexes from BTT (Table S.23) and IDA (Table S.24) groups.

In all studied groups, no statistically significant differences were observed in hematologic data and iron status according to the genotype of K253E, H63D and C282Y ($p > 0.05$) variants. Results for K253E are in agreement with results found in the literature [234]–[236], reinforcing the benign effect of this variant. The fact that there was a low frequency of the disease risk allele for the variants of the *HFE* gene, in all the investigated groups, may be a factor masking possible differences and associations. Detailed results are displayed in Tables S.25 to S.33.

Considering only IDA patients, no significant differences were observed in hematologic data and iron status according to the genotype of the polymorphism. Other than the statistical limitations the lack of results can probably be related to the impact of previous iron supplementation, which inevitably modifies the hematologic parameters, minimizing the differences between the groups.

Through logistic regression analysis in the control group, we confirm positive association between H63D and the parameter RDW (β = 0.53; SE = 0.26, 95% CI [0.09,1.08], p = 0.042), (Table S.34). If an association was made between this variant and the analyzed parameters, we would expect it to be with the iron status, and not with RDW that has been described to be within the normal range values in control subjects grouped by *HFE* genotype [259].

In the BTT group we also confirmed association between H63D and three iron parameters SI (β = -0.02; SE = 0.01, 95% CI [-0.04,-0.01], p = 0.006), TBIC (β = 0.01; SE = 4.89×10^{-03}, 95% CI [7.78×10^{-04},0.02], p = 0.042), and TSAT (β = -0.08; SE = 0.02, 95% CI [-0.13,-0.04], p = 0.001), (Table S35). Yet, the estimated coefficients are near 0 implying that the effect of the predictor is small. These findings in the BTT group, indirectly support the hypothesis that the erythroid regulator (β-thalassemia) might represent a stronger stimulus than the stores regulator (defective *HFE* gene) in determining the degree of iron absorption [260]. The small sample size of the IDA group did not allow regression analysis for any of the studied variants.

In sum, although some positive results have been found, the case-control groups used in this dissertation to assess the potential of genetic markers on iron deficiency anemia and beta-thalassemia trait groups failed to completely portray what was previously stated in the literature or to obtain significant associations between all four variants from *TMPRSS6* and *HFE* genes and the hematological and iron status parameters, clinically relevant for anemia differentiation. These

type of studies are sometimes unable to establish association due to intricate features of the populations, either genetic or environmental [50], [261]. Therefore, this analysis should be expanded to include a larger number of subjects, especially in the IDA group, while guaranteeing that the genotypic frequencies for each variant are robust enough to determine any statistically significant associations between genotype and phenotype, and taking into account sex, age and inter-ethnic variations to avoid confounding and stratification within the population.

3.4 Differentiating microcytic anemia

The final objective of this dissertation emerged from the difficulties we came across when trying, in the previous objective, to determine to which case group, BTT or IDA, subjects with microcytic anemia and missing data regarding the Hb A2 levels and iron status parameters, belonged to. Thenceforth, an opportunity arose to put into practice and study the performance of 13 commonly used mathematical indices for the screening of microcytic anemia patients, while ascertaining which are best suited for a Portuguese population.

A total of 172 subjects with microcytic anemia were retrospectively analyzed, of those, 111 were BTT subjects (60 males, 51 females) and 61 were IDA subjects (10 males, 51 females), their status was confirmed by molecular analysis of the mutations responsible for the thalassemia trait. The comparison of hematological and biochemical data of the subjects in each group (Table 3.22) revealed that all the parameters are significantly different ($p < 0.001$) from one case group to the other, this is especially important for the hematologic parameters in which the indices are based on. The hematological differences between males and females should be considered when using the mathematical indices and formulas for discriminating between BTT and IDA, to minimize variance in the discrimination power observed [262]. Thus, when determining the diagnostic accuracy of the used discrimination indices, the analysis was performed separately for males and females and then as one group.

Table 3.22 – Hematological and biochemical data and comparison of studied groups as well as sex groups.

	BTT (microcytic anemia)			IDA (microcytic anemia)			
	Total (N=111)	Male (n=60)	Female (n=51)	Total (N=61)	Male (n=10)	Female (n=51)	MW
Parameters	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	p[a]
Age (y)	45±18	44±17	45±18	44±16	59±16	41±15	< **0.001**
RBC (x10^12/L)	5.71±0.60	6.07±0.41	5.28±0.48	4.50±0.41	4.68±0.46	4.47±0.40	< **0.001**
Hb (g/dL)	11.86±1.17	12.53±0.94	11.07±0.88	10.09±1.44	10.39±1.67	10.03±1.41	< **0.001**
MCV (fL)	64.87±4.25	64.10±4.08	65.78±4.30	71.02±6.57	71.22±6.34	70.99±6.68	< **0.001**
MCH (pg)	20.82±1.40	20.64±1.20	21.04±1.59	22.46±2.74	22.23±2.45	22.51±2.82	< **0.001**
RDW (%)	15.58±4.36	16.40±5.69	14.61±1.40	25.13±12.60	31.64±16.13	23.86±11.56	< **0.001**
HbA2 (%)	4.11±0.42	4.08±0.37	4.15±0.48	-	-	-	-
SI (µg/dL)	108.39±31.56	110.05±28.84	106.57±34.50	30.21±14.28	27.89±10.45	30.72±15.03	< **0.001**
Tf (mg/dL)	261.36±40.71	251.66±33.70	272.22±45.26	383.73±69.88	407±94.92	378.50±63.31	< **0.001**
TIBC (µg/dL)	328.53±54.62	314.58±42.12	344.15±62.69	479.61±87.29	508.75±118.66	473.05±79.05	< **0.001**
TSAT (%)	33.80±10.78	35.40±9.74	32.02±11.68	6.43±3.24	5.96±2.98	6.53±3.32	< **0.001**
Ft (µg/L)	168.97±207.48	208.74±187.06	126.29±221.77	8.02±10.98	14.73±20.37	6.50±7.26	0.068

[a] – Comparison between the parameters of the total number of subjects in BTT and IDA groups; **BTT** – Beta thalassemia trait; **IDA** – Iron deficiency anemia; **SD** – Standard deviation; ***p* –** *p*-value (**significant < 0.05**); **MW –** Mann-Whitney test; **y** – Years; **m** – Month; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

The analysis was based on a binary classification model, which categorizes into one of two classes (true and false), giving rise to four possible classifications: a true positive, a true negative, a false positive, or a false negative [263]. The index used to distinguish between BTT and IDA with original publication thresholds, the number and proportion of correctly identified patients (true positives), as well as their discrimination power shown through measures of accuracy: SENS, SPEC, PPV, NPV, YI, and AUC values are presented in Tables S.36-S37, and Table 3.23,

to represent the combined data, male and females subjects, respectively. The ROC curves that represent the AUC measurement of the different indices and formulas and the accuracy findings for all calculated new best thresholds used when applying the indices to this study populations are detailed in Supplementary Material's Figures S.7-S.10 and Table S.38-S.40.

The analysis confirmed that the sample of male subjects was disproportionally small and did not allow the draw of conclusions in that group. The effects of the disproportion in case frequency and small sample size were observable on the ROC's curve behavior, systematically showing enhanced index performance (Table S.37 and Figures S.7-S.10). The same was observable for the combined data that also reflected this deceptive tendency, for this reason, conclusions were drawn only for the female group (Table S.36 and Figures S.7-S.10). An AUC value < 0.75 means that the test shows deficiencies in its diagnostic accuracy [162], in this study the size of the sample did not provide a sufficiently reliable estimate of this area presenting consistently high values that made difficult the determination of the best formula, thus, a cut-off of 0.70 was established for the YI.

Receiver operating characteristics graphs are constructed by plotting the true positive rate against the false-positive rate, making possible the identification of several regions of interest in the graph. The diagonal line from the bottom left corner to the top right corner denotes random classifier performance, a test that falls into this line has an AUC of 0.50 and represents a worthless differentiation, with as many false-positive responses as it produces true positive responses [157], [263]. Additionally, there are two important regions in the ROC graph, the left bottom that represents the conservative performance region where few false positive errors are committed, and the top left region representing classifiers with liberal performance where classifiers have a good true positive rate but also commit substantial numbers of false-positive errors. In the middle of these regions, at the point in the top left corner, we have the representation of a perfect differentiation classifier with AUC of 1.00 [263].

The results show wide differences in the discrimination power of the mathematical indices when applied to female subjects. Although the original published indices and formulas revealed very high discrimination power (almost 100 %) in terms of the correctly diagnosed case, none of the studied indices renders an absolute discrimination power when applied thereafter in other settings. In this case, the three best-performing indices to differentiate the two groups (IDA vs BTT) in the female population were RBC (YI = 0.71; AUC = 0.902), RDWI (YI = 0.84; AUC = 0.973) and G&K (YI = 0.82; AUC = 0.972), these indices correctly diagnosed 85 %, 92 % and 91 % of the cases, respectively. The worst performing index for case differentiation in this population was MCHD (Y I= 0.02; AUC = 0.569) showing almost no predictive value.

Smoothed ROC curves were constructed to visualize the AUC for each of the three best tested hematologic indices to distinguish BTT cases from IDA cases, respectively (Figure 3.28). Marked on the ROC curve is the new estimated best threshold, this indicted criterion value is the cut-off value corresponding to the highest accuracy (minimal false-negative and minimal false-positive results) [157]. In the graphical representation, we can verify that all new thresholds from the best performing indices are in the top left region of the graph, they are all liberal classifiers. When using the indices in a real-life context of screening, meaning that the use of liberal thresholds may be prefer by guarantying that a bigger number of β-thalassemia carriers are either confirmed by additional Hb A2 levels measurements or molecular analysis and receive the correct genetic counselling, preventing severe and lethal forms of thalassemia syndromes and unnecessary oral iron supplementation [264]. The sensitivity and specificity findings of these

cutoffs are summarized in Table 3.24, showing that for RBC and RDWI thresholds apparently did not deviate much from the value of original publication, and for G&K there might be a difference between our threshold and the published one (cutoff = 65; predictive cutoff = 71.26).

Table 3.23 – Diagnostic accuracy findings for all considered indices used in microcytic anemia female subjects.

Indices	Female									
	BTT (n=51)	IDA (n=51)	Correctly diagnosed	Accuracy (Efficiency)	SENS	SPEC	PPV	NPV	YI	AUC
RBC										
BTT > 5	40	4	87	**0.85**	0.78	0.92	0.91	0.81	**0.71**	**0.902**
IDA < 5	11	47			0.92	0.78	0.81	0.91		
England and Fraser (E&F)										
BTT< 0	19	0	70	0.69	0.37	1.00	1.00	0.61	0.37	0.915
IDA > 0	32	51			1.00	0.37	0.61	1.00		
Mentzer										
BTT< 13	37	7	81	0.79	0.73	0.86	0.84	0.76	0.59	0.886
IDA > 13	14	44			0.86	0.73	0.76	0.84		
Srivastava										
BTT< 3.8	19	1	69	0.68	0.37	0.98	0.95	0.61	0.35	0.846
IDA > 3.8	32	50			0.98	0.37	0.61	0.95		
Shine and Lal (S&L)										
BTT< 1530	51	42	60	0.59	1.00	0.18	0.55	1.00	0.18	0.712
IDA > 1530	0	9			0.18	1.00	1.00	0.55		
Bessman (Rdw)										
BTT< 15	36	5	82	0.80	0.71	0.90	0.88	0.75	0.61	0.917
IDA > 15	15	46			0.90	0.71	0.75	0.88		
Ricerca										
BTT< 4.4	51	26	76	0.75	1.00	0.49	0.66	1.00	0.49	0.970
IDA > 4.4	0	25			0.49	1.00	1.00	0.66		
Jayabose (RDWI)										
BTT< 220	44	1	94	**0.92**	0.86	0.98	0.98	0.88	**0.84**	**0.973**
IDA > 220	7	50			0.98	0.86	0.88	0.98		
Green and King (G&K)										
BTT< 65	42	0	93	**0.91**	0.82	1.00	1.00	0.85	**0.82**	**0.972**
IDA > 65	9	51			1.00	0.82	0.85	1.00		
MDHL										
BTT> 1.63	34	1	84	0.82	0.67	0.98	0.97	0.75	0.65	0.913
IDA < 1.63	17	50			0.98	0.67	0.75	0.97		
MCHD										
BTT> 0.3045	42	41	52	0.51	0.82	0.20	0.51	0.53	0.02	0.569
IDA < 0.3045	9	10			0.20	0.82	0.53	0.51		
Sirdah										
BTT< 27	26	1	76	0.75	0.51	0.98	0.96	0.67	0.49	0.897
IDA > 27	25	50			0.98	0.51	0.67	0.96		
Ensani										
BTT< 15	36	8	79	0.77	0.71	0.84	0.82	0.74	0.55	0.869
IDA > 15	15	43			0.84	0.71	0.74	0.82		

Bold – Best performing indices; **BTT** – Beta thalassemia trait; **IDA** – Iron deficiency anemia; **RBC** – Red blood cells; **RDW –** Red cell distribution width**; RDWI –** Red cell distribution width index; **MDHL** – Mean Density of Hemoglobin per Liter; **MCHD** – Mean Cell Hemoglobin Density **SENS** – Sensitivity; **SPEC** – Specificity; **PPV** – Positive predictive values; **NPV** – Negative predictive values; **YI** – Youden's index; **AUC** – Area under the curve.

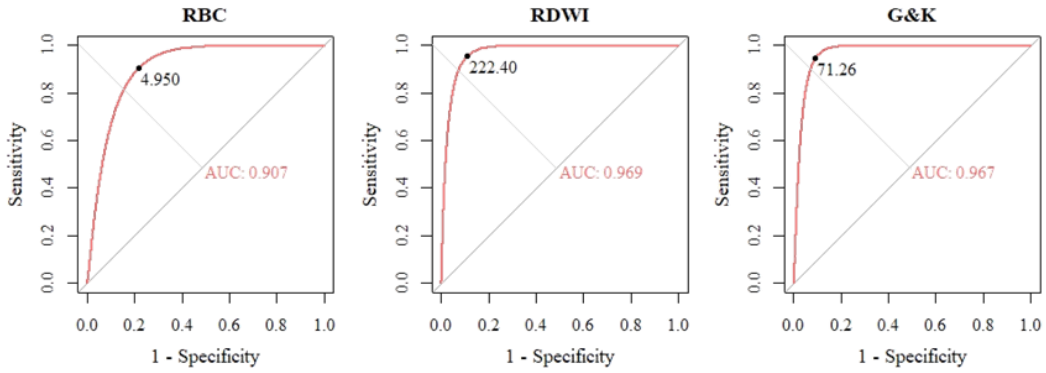


Figure 3.28 – Smoothed receiver operating characteristics (ROC) curves of the three best performant discriminant indices for discrimination of beta-thalassemia trait and iron deficiency anemia in a female Portuguese population. In the curve is marked the calculated best thresholds of each index.

Table 3.24 – Accuracy findings for calculated new best thresholds of the best performing indices for female microcytic anemia subjects.

Indices	Female					
	Cutoff	Predicted Cutoff	SENS	SPEC	PPV	NPV
RBC	5	4.95	0.92	0.80	0.82	0.91
Jayabose (RDWI)	220	222.40	0.98	0.90	0.91	0.98
Green and King (G&K)	65	71.26	0.94	0.92	0.92	0.94

RBC – Red blood cells; **RDWI –** Red cell distribution width index; **SENS** – Sensitivity; **SPEC** – Specificity; **PPV** – Positive predictive values; **NPV** – Negative predictive values.

The discriminative power of the indices, either simple indices or resulting from a complex formula, did not reach maximum diagnostic performance, the same was described in the literature and continuously stimulates authors to devise new, and supposedly, better indices to apply in their local patient population. Across these studies, there is a vast variation in patient selection criteria, such as: age; sex; the types of thalassemia included; the geographical origin of the patients; the type of hematology analyzer used; and cutoff value used for the respective discriminant indices [162].

In this study, the data used to generate the ROC curves represent uncombined cases of BTT and IDA, since ROC analysis has the principle of discriminating between two population, one population with the disease and the other population without the diseases. However, it is possible to use these discrimination formulas to check how they classify the combined cases [157] since we can see described in the literature studies that included both α- and β-thalassemia carriers. Albeit, the numbers reported are too small to make a solid conclusion as to the utility of the discriminant indices in these conditions [162]. Although our analysis only looks upon BTT subjects, the mutations responsible for it were different, with some resulting in a mild reduction in the output of β-chains (β^+) and other in no β-globin gene product (β^0). This led us to inquire about the differences between the parameters of the female BTT subjects used for the analysis (Table 2.25). No differences were observed between the hematological parameters used for the formulation of discriminant mathematical indices, ensuring that the molecular cause for BTT was not a confounding factor of the results. Additionally, it was found that there was no significant difference in the iron profile among the β^0 and β^+ thalassemia carriers, in accordance with what is described in the literature [265].

Overall, the analysis' results are in agreement with recently published evaluations of the indices revealing that RBC count is the best single measurement, while the G&K index shows

consistently to be most reliable index; both the Ricerca and S&L indices have almost 100% sensitivity but poor specificity [157], [266]–[268].

Table 3.25 – Difference between parameters of β^+- and β^0-thalassemia trait female subjects.

	BTT (female n=51)		
Thalassemia Trait	β^+ (n=12)	β^0 (n=39)	MW
Parameters	Mean ± SD	Mean ± SD	*p*
RBC (x10^12/L)	5.35±0.56	5.26±0.46	0.764
Hb (g/dL)	11.63±0.65	10.86±0.91	0.057
MCV (fL)	67.67±5.71	65.01±3.56	0.166
MCH (pg)	21.96±2.06	20.69±1.28	0.053
RDW (%)	14.70±1.31	14.58±1.45	0.593
HbA2 (%)	3.79±0.48	4.30±0.42	**0.003**
SI (µg/dL)	103.85±25.13	107.38±36.51	0.748
Tf (mg/dL)	291.83±47.75	263.50±44.94	0.146
TIBC (µg/dL)	380.94±75.69	329.38±56.18	0.058
TSAT (%)	28.25±9.03	33.64±12.41	0.192
Ft (µg/L)	128.30±175.83	129.93±234.43	0.537

BTT – Beta thalassemia trait; **SD** – Standard deviation; ***p* – *p*-value (significant < 0.05)**; **MW –** Mann-Whitney test; **y** – Years; **m** – Month; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

The mathematical indices accuracy analysis performed with a female adult Portuguese population suggests a similarity with other Mediterranean thalassemia endemic countries such as Spain [266], [269], Turkey [268], [270], Iran [271], Greece [272] and Palestine [262], where G&K and RDWI performed best. The same is observed in Brazil [264] and France [273], probably due to its Portuguese ancestry and population migration that have spread thalassemia genes to other original non-endemic parts of the world [162].

Differentiating BTT from IDA can be a diagnostic dilemma, as both conditions share many characteristics. Although, one cannot make definite diagnosis without it being 100 % sensitive and specific, there is value in identifying patients with microcytic anemia for which molecular diagnostic tests for confirming thalassemia trait are indicated, representing a useful tool in the doctor's guidance about the initial approach to be adopted. Moreover, these formulas can be the only differential tool in situations where other specific confirmatory tests are not available [264]. In this context, not only do specialists make use of the mathematical formulas to solve the diagnostic problem, but recently, there has been an increasing interest in the use of artificially intelligent automated medical diagnostic systems [274].

Chapter 4 – Conclusions

Diseases affecting hemoglobin synthesis and function are extremely common worldwide, altering hemoglobin structure and biochemical properties with physiological effects ranging from insignificant to severe [275]. In the first part of this dissertation, when investigating atypical hemoglobinopathies' phenotypes, we were able to shed a light in the molecular causes for the phenotype of 15 patients, while understanding the biology and clinical implications of the found variants. In total, we were able to find: 8 rare hemoglobin structural variants, present in the α-, β- and δ-globin genes; 3 polymorphisms, two that influenced the expression of γ-globin genes and one that helped understand the geographical origin of one of the *HBB* variants; 3 mutations responsible for β-thalassemia, an intronic point mutation, a structural variant that leads to a thalassemic hemoglobinopathy and a large deletion in the β-globin gene cluster; and 1 gene deletion responsible for α-thalassemia.

In some of the analyzed cases, the major molecular cause responsible for the patient's phenotype was unraveled but some phenotypic features could not be justified only by the molecular lesions found. That is why, in the future, if saw fit by the clinician, these cases could be revisited for further investigation. Additionally, cases where the lesions may lead to serious clinical complications, a family study is recommended.

In children and younger IDA patients with no reason for iron deficiency but unresponsiveness to routine iron treatment, the diagnosis of IRIDA needs to be considered and *TMPRSS6* gene mutations should be searched for. Out of the ten patients we studied for IRIDA, five had already the *TMPRSS6* gene previously analyzed by a less comprehensive sequencing technique. NGS results revealed no difference from the previously found variants in these control subjects, which were proved not to be responsible for the IRIDA phenotype. The same was verified for 4 other subjects, never analyzed before, in which the variants do not justify the phenotype, but may be modelling their hematological and iron status phenotype. Overall, there was only one subject whose molecular lesions could justify the presented phenotype. Further investigation should be performed on unresolved cases.

This is the first study to investigate the G291S (c.871G>A) genetic variant of the *TMPRSS6* gene in a patient with a suspicion of IRIDA. Although IRIDA is noted as an autosomal recessive disease, we infer that, in this case, the result of digenic inheritance of the novel damaging mutation and 3 common modulating SNPs in the same gene (K253E, S361S, and Y739Y), in addition to the co-inheritance of the α-thalassemia allele, may add up to an IRIDA-like phenotype. Further family studies should be conducted, as well as functional studies of the mutated protein, since the understanding of the molecular mechanisms of transmembrane protein action is impossible without detailed knowledge of their structure and how they interact with other proteins, nucleic acids and lipids [276].

The genetic association study results suggest that both heterozygous (TC) and homozygous (CC) genotypes of V736A present themselves as a protective factor against IDA, revealing that the presence of only one protective allele (C), in heterozygous individuals, is enough to render a protective effect. Additionally, we found that, in the female control group, the ones with homozygous genotype (CC) had significantly increased MCV, MCH and TSAT parameters when compared to the wild type (TT). Logistic regression results have detected an association between H63D and RDW in control subjects, and a small association between the same variant and the iron parameters of BTT subjects.

In the end, the data generated from this study will provide insight for future studies related to BTT and IDA in the Portuguese population. However, these results are recommended to be elucidated in a larger more curated population, for us to confirm possible associations between the genetic *TMPRSS6* and *HFE* polymorphisms and the phenotype of the subjects within our study groups.

Both, the clinical case study of IRIDA suspects and the genetic association study, highlight the importance of improving the knowledge of genetic factors influencing iron status, so that clinicians can improve care and advice given to populations at risk for iron deficiency, especially if combined with unfavorable environmental conditions such as diets with low bioavailable iron and high inflammation burden [50].

Finally, when trying to use and assess the performance of hematological mathematical indices that discriminate microcytic anemia patients (BTT and IDA), we found that for a female adult Portuguese population, although none of the studied indices rendered an absolute discrimination power, the best performing indices where RBC, G&K and RDWI. The obtained results presented some similarities with other Mediterranean countries, where G&K and RDWI also performed above our set cutoff. The same is observed in Brazil and France probably due to its Portuguese ancestry and migration. For these reasons, we conclude that when aiming to diagnose the condition underlying microcytic anemia, there is value in using these indices to recognize the individuals suspected of BTT and forward them for Hb A2 measurement or *HBB* molecular test.

In the future, it should be determined how large sample sizes should be to ensure that one can statistically detect differences in the accuracy of diagnostic techniques and to provide a sufficiently reliable AUC estimate, adding a robust group of male patients to extrapolate which of these indices would best apply to this sex and the whole adult Portuguese population. This effort would ensure that, in time, these indices would facilitate the screening of microcytic anemias and potentially be the basis of an artificially intelligent automated medical diagnostic system adapted for the Portuguese population, providing significant positive implications for hematology practice.

Overall, the statistical analysis performed during this dissertation, for both genetic association studies and diagnostic accuracy determination of mathematical indices for discrimination of microcytic anemia subjects was limited by the small sample size, and larger studies should be carried out to address the investigated issues.

In conclusion, although there is still work to be done regarding the four objectives set for this dissertation, we believe that by integrating various genomic techniques in clinical practice we have improved the understanding of the molecular etiology behind hereditary anemias. By studying this diverse group of disorders, which sometimes poses as a diagnostic challenge for clinicians, we made possible a correct diagnostic of the studied patients, allowing for correct clinical management and genetical counselling for the studied patients, and in turn, improving their quality of life.

Bibliographic references

[1] J. L. Badano and N. Katsanis, "Beyond mendel: An evolving view of human genetic disease transmission," *Nature Reviews Genetics*, vol. 3, no. 10, pp. 779–789, Oct 2002.

[2] H. Blencowe, S. Moorthie, M. Petrou, H. Hamamy, S. Povey, A. Bittles, S. Gibbons, M. Darlison, and B. Modell, "Rare single gene disorders: estimating baseline prevalence and outcomes worldwide," *Journal of Community Genetics*, vol. 9, no. 4, pp. 397–406, Oct. 2018.

[3] S. L. Thein, "Genetic insights into the clinical diversity of β thalassaemia," *British Journal of Haematology*, vol. 124, no. 3. pp. 264–274, Feb 2004.

[4] S. N. Wakap, D.M. Lambert, A. Olry, C. Rodwell, C. Gueydan, V. Lanneau, D. Murphy, Y. Le Cam, and A. Rath, "Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database," *European Journal of Human Genetics*, vol. 28, no. 2, pp. 165–173, Feb. 2020.

[5] G. Kuhlenbäumer, J. Hullmann, and S. Appenzeller, "Novel genomic techniques open new avenues in the analysis of monogenic disorders," *Human Mutation*, vol. 32, no. 2, pp. 144–151, Feb 2011.

[6] D. R. Higgs, J. D. Engel, and G. Stamatoyannopoulos, "Thalassaemia," *Lancet*, vol. 379, no. 9813, pp. 373–383, Jan 2012.

[7] J. C. Lin, "Approach to Anemia in the Adult and Child," in *Hematology: Basic Principles and Practice*, 7th ed., R. Hoffman, E.J. Benz, H. Heslop and J. Weitz, Elsevier Inc., 2018, pp. 458–467.

[8] M. H. Steinberg, E. J. Benz, A. H. Adewoye, and B. L. Ebert, "Pathobiology of the Human Erythrocyte and Its Hemoglobins," in *Hematology: Basic Principles and Practice*, 7th ed., R. Hoffman, E.J. Benz, H. Heslop and J. Weitz, Elsevier Inc., 2018, pp. 447–457.

[9] H. Drakesmith, E. Nemeth, and T. Ganz, "Ironing out Ferroportin," *Cell Metabolism*, vol. 22, no. 5. pp. 777–787, Nov 2015.

[10] J. A. Lukin and C. Ho, "The structure-function relationship of hemoglobin in solution at atomic resolution," *Chemical Reviews*, vol. 104, no. 3, pp. 1219–1230, Mar 2004.

[11] A. N. Schechter, "Hemoglobin research and the origins of molecular medicine," *Blood*, vol. 112, no. 10, pp. 3927–3938, Nov 2008.

[12] B. J. Bain, "Haemoglobin and the genetics of haemoglobin synthesis," in *Haemoglobinopathy Diagnosis*, 2nd ed, B. J. Bain, Fellow of the Royal College of Pathologists of Australasia, 2006, pp. 1–25.

[13] J. G. Betts, P. Desaix, E. Johnson, J. E. Johnson, O. Korol, D. Kruse, B.Poe, J. A. Wise, M.Womble and K. A. Young, "Anatomy and Physiology," Texas, USA, OpenStax, 2013, pp. 794.

[14] A. X. Fan, M. Hossain, M. A. Hossain, J. Stees, E. Gavrilova, and J. Bungert, "Regulation of erythroid cell differentiation by transcription factors, chromatin structure alterations, and noncoding RNA," *Elsevier*, pp.237-264, Jan 2015.

[15] S. J. Fuller and J. S. Wiley, "Heme Biosynthesis and Its Disorders: Porphyrias and Sideroblastic Anemias," in *Hematology: Basic Principles and Practice*, 7th ed., R. Hoffman, E.J. Benz, H. Heslop and J. Weitz, Elsevier Inc., 2018, pp. 497-513.e6.

[16] R. J. Ouellette and J. D. Rawn, "Amino Acids, Peptides, and Proteins," in *Organic Chemistry*, Elsevier, 2018, pp. 929–971.

[17] S. A. Sarnaik, "Thalassemia and related hemoglobinopathies," *The Indian Journal of Pediatrics*, vol. 72, no. 4, pp. 319–324, Apr 2005.

[18] J. F. Storz, J. C. Opazo, and F. G. Hoffmann, "Gene duplication, genome duplication, and the functional diversification of vertebrate globins," *Molecular Phylogenetics and Evolution*, vol. 66, no. 2., pp. 469–478, Feb 2013.

[19] K. Ryan, B. J. Bain, D. Worthington, J. James, D. Plews, A. Mason, D. Roper, D. Rees, B. de la Salle, and A. Streetly, "Significant haemoglobinopathies: Guidelines for screening and diagnosis," *British. Journal of Haematology*, vol. 149, no. 1, pp. 35–49, Sep 2010.

[20] G. M. Brittenham, "Pathophysiology of Iron Homeostasis," in *Hematology: Basic Principles and Practice*, 7th ed., R. Hoffman, E.J. Benz, H. Heslop and J. Weitz, Elsevier Inc., 2018, pp. 468–477.

[21] C. Camaschella, A. Nai, and L. Silvestri, "Iron metabolism and iron disorders revisited in the hepcidin era," *Haematologica*, vol. 105, no. 2, pp. 260–272, Jan 2020.

[22] J. P. Kehrer, "The Haber-Weiss reaction and mechanisms of toxicity," *Toxicology*, vol. 149, no. 1, pp. 43–50, Aug 2000.

[23] D. J. R. Lane, A. M. Merlot, M. H. Huang, D. H. Bae, P. J. Jansson, S. Sahni, D. S. Kalinowski and D. R Richardson, "Cellular iron uptake, trafficking and metabolism: Key molecules and mechanisms and their roles in disease," *Biochimica et Biophysica Acta - Molelular Cell Research*, vol. 1853, no. 5, pp. 1130–1144, May 2015.

[24] G. Papanikolaou and K. Pantopoulos, "Iron metabolism and toxicity," *Toxicology and Applied Pharmacology*, vol. 202, no. 2. pp. 199–211, Jan 2005.

[25] M. U. Muckenthaler, S. Rivella, M. W. Hentze, and B. Galy, "A Red Carpet for Iron Metabolism," *Cell*, vol. 168, no. 3, pp. 344–361, Jan 2017.

[26] T. Ganz and E. Nemeth, "Iron homeostasis in host defence and inflammation," *Nature Reviews Immunology*, vol. 15, no. 8. pp. 500–510, Aug 2015.

[27] B. Silva and P. Faustino, "An overview of molecular basis of iron metabolism regulation and the associated pathologies," *Biochimica et Biophysisica Acta - Molecular Basis of Disiese.*, vol. 1852, no. 7, pp. 1347–1359, Jul 2015.

[28] M. W. Hentze, M. U. Muckenthaler, B. Galy, and C. Camaschella, "Two to Tango: Regulation of Mammalian Iron Metabolism," *Cell*, vol. 142, no. 1, pp. 24–38, Jul 2010.

[29] G. J. Anderson, D. M. Frazer, and G. D. McLaren, "Iron absorption and metabolism," *Current Opinion in Gastroenterology*, vol. 25, no. 2, pp. 129–135, Mar 2009.

[30] P. Brissot and O. Loréal, "Iron metabolism and related genetic diseases: A cleared land, keeping mysteries," *Journal of Hepatology*, vol. 64, no. 2. pp. 505–515, Feb 2016.

[31] C. P. Anderson, M. Shen, R. S. Eisenstein, and E. A. Leibold, "Mammalian iron metabolism and its control by iron regulatory proteins," *Biochimica et Biophysica Acta - Molecular Cell Research*, vol. 1823, no. 9. pp. 1468–1483, Sep 2012.

[32] C. Y. Wang and J. L. Babitt, "Liver iron sensing and body iron homeostasis," *Blood*, vol. 133, no. 1. pp. 18–29, Jan 2019.

[33] R. Szabo and T. H. Bugge, "Membrane-anchored serine proteases in vertebrate cell and developmental biology," *Annual Review of Cell Developmental Biology*, vol. 27, no. 1, pp. 213–235, Nov 2011.

[34] D. Meynard, J. L. Babitt, and H. Y. Lin, "The liver: Conductor of systemic iron balance," *Blood*, vol. 123, no. 2. pp. 168–176, Jan 2014.

[35] R. E. Fleming and P. Ponka, "Mechanisms of disease: Iron overload in human disease," *New England Journal of Medicine*, vol. 366, no. 4, pp. 348–359, Jan 2012.

[36] S. Rivella, "Iron metabolism under conditions of ineffective erythropoiesis in β-Thalassemia," *Blood*, vol. 133, no. 1. pp. 51–58, Jan 2019.

[37] T. Ganz, "Hepcidin," *The Japanese Journal of Clinical Hematology - Rinsho. Ketsueki.*, vol. 57, no. 10, pp. 1913–1917, Jan 2016.

[38] E. McLean, M. Cogswell, I. Egli, D. Wojdyla, and B. De Benoist, "Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993-2005," *Public Health Nutrition*, vol. 12, no. 4, pp. 444–454, Apr 2009.

[39] C. Fonseca, F. Marques, A. Robalo Nunes, A. Belo, D. Brilhante, and J. Cortez, "Prevalence of anaemia and iron deficiency in Portugal: The EMPIRE study," *Internal Medical Journal*, vol. 46, no. 4, pp. 470–478, Apr 2016.

[40] WHO, "Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity,", (No. WHO/NMH/NHD/MNM/11.1), *Geneva, World Health Organanization.*, 2011.

[41] E. McLean, M. Cogswell, I. Egli, D. Wojdyla, and B. De Benoist, "Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993-2005," *Public Health Nutrition.*, vol. 12, no. 4, pp. 444–454, Apr 2009.

[42] S. L. James, D. Abate, K. H. Abate, S.M. Abay, C. Abbafati, N. Abbasi, H. Abbastabar, F. Abd-Allah, J., Abdela, , Abdelalim, A. and I. Abdollahpour, "Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016," *The Lancet*, vol. 390, no. 10100, pp. 1211–1259, Sep 2017.

[43] C. Hershko and C. Camaschella, "How I treat unexplained refractory iron deficiency anemia," *Blood*, vol. 123, no. 3, pp. 326–333, Jan 2014.

[44] C. Camaschella, "Iron deficiency: New insights into diagnosis and treatment," *Hematology*, vol. 2015, no. 1, pp. 8–13, Dec 2015.

[45] G. M. Brittenham, "Disorders of Iron Homeostasis: Iron Deficiency and Overload," in *Hematology: Basic*

Principles and Practice, 7th ed., R. Hoffman, E.J. Benz, H. Heslop and J. Weitz, Elsevier Inc., 2018, pp. 478–490.

[46] A. Lopez, P. Cacoub, I. C. Macdougall, and L. Peyrin-Biroulet, "Iron deficiency anaemia," *The Lancet*, vol. 387, no. 10021, pp. 907–916, Feb 2016.

[47] D. L. Longo, C. Camaschella, "Iron-deficiency anemia," *New England Journal of Medicine*, vol. 372, no. 19, pp. 1832–1843, May-2015.

[48] C. Camaschella, "New insights into iron deficiency and iron deficiency anemia," *Blood Reviews*, vol. 31, no. 4. pp. 225–233, Jul 2017.

[49] M. Cau, M. A. Melis, R. Congiu, and R. Galanello, "Iron-deficiency anemia secondary to mutations in genes controlling hepcidin," *Expert Reviews of Hematology*, vol. 3, no. 2, pp. 205–216, Apr 2010.

[50] W. N. Gichohi-Wainaina, G. W. Towers, D. W. Swinkels, M. B. Zimmermann, E. J. Feskens, and A. Melse-Boonstra, "Inter-ethnic differences in genetic variants within the transmembrane protease, serine 6 (TMPRSS6) gene associated with iron status indicators: a systematic review with meta-analyses," *Genes & Nutrition*, vol. 10, no. 1, pp. 1–15, Jan 2015.

[51] C. Y. Wang, D. Meynard, and H. Y. Lin, "The role of TMPRSS6/matriptase-2 in iron regulation and anemia," *Frontiers in Pharmacology*, vol. 5, pp. 1–6, May 2014.

[52] M. Stirnberg, E. Maurer, A. Horstmeyer, S. Kolp, S. Frank, T. Bald, K. Arenz, A. Janzer, K. Prager, P. Wunderlich and J. Walter, "Proteolytic processing of the serine protease matriptase-2: Identification of the cleavage sites required for its autocatalytic release from the cell surface," *Biochemical Journal*, vol. 430, no. 1, pp. 87–95, Aug 2010.

[53] G. Velasco, S. Cal, V. Quesada, L. M. Sánchez, and C. López-Otín, "Retractions: Matriptase-2, a membrane-bound mosaic serine proteinase predominantly expressed in human liver and showing degrading activity against extracellular matrix proteins," *Journal of Biological Chemistry*, vol. 294, no. 4. p. 1430, Oct 2019.

[54] P. Lee, "Role of matriptase-2 (TMPRSS6) in iron metabolism," *Acta Haematologica*, vol. 122, no. 2–3. pp. 87–96, Nov 2009.

[55] A. Nai, A. Pagani, L. Silvestri, N. Campostrini, M. Corbella, D. Girelli, M. Traglia, D. Toniolo, and C. Camaschella, , "TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals," *Blood*, vol. 118, no. 16, pp. 4459–4462, Oct 2011.

[56] D. A. Khuong-Quang, J. Schwartzentruber, M. Westerman, P. Lepage, K. E. Finberg, J. Majewski and N. Jabado, "Iron refractory Iron deficiency anemia: Presentation with hyperferritinemia and response to oral iron therapy," *Pediatrics*, vol. 131, no. 2, Feb 2013.

[57] B. Benyamin, M. A.Ferreira, G. Willemsen, S. Gordon, R. P. Middelberg, B. P. McEvoy, J. J. Hottenga, A. K. Henders, , M. J. Campbell, L. Wallace, and I. H. Frazer, "Common variants in TMPRSS6 are associated with iron status and erythrocyte volume," *Nature Genetics.*, vol. 41, no. 11, pp. 1173–1175, Nov 2009.

[58] S. Killip, J. M. Bennett and M. D. Chambers, "Iron deficiency anemia.," *American family physician.*, vol. 75, no. 5, pp. 671-678, Mar 2007.

[59] F. Bermejo and S. García-López, "A guide to diagnosis of iron deficiency and iron deficiency anemia in digestive diseases," *World Journal of Gastroenterology*, vol. 15, no. 37, pp. 4638–4643, Oct 2009.

[60] WHO, "Serum ferritin concentrations for the assessment of iron status and iron deficiency in populations. Vitamin and Mineral Nutrition Information System.," (WHO/NMH/NHD/MNM/11.2), *Geneva, World Health Organanization.*, 2011.

[61] L. Harmse, "Understanding the management of iron deficiency anaemia," *South African Family Practice*, vol. 58, no. 6, pp. 40–44, 2016..

[62] S. F. Clark, "Iron deficiency anemia: Diagnosis and management," *Currunt. Opininion in Gastroenterology*, vol. 25, no. 2, pp. 122–128, Mar 2009.

[63] M. Alleyne, M. K. Horne, and J. L. Miller, "Individualized Treatment for Iron-deficiency Anemia in Adults," *American Journal of Medicine*, vol. 121, no. 11. pp. 943–948, Nov 2008.

[64] J. D. Weatherall, J. B. Clegg, "Inherited haemoglobin disorders: an increasing global health problem,"*Bulletin of the World Health Organization*, no. 79, pp.704–712, 2001.

[65] M. Angastiniotis, J. L. Vives Corrons, E. S. Soteriades, and A. Eleftheriou, "The impact of migrations on the health services for rare diseases in Europe: The example of haemoglobin disorders," *The Scietific World Journal*, vol. 2013, Jan 2013.

[66] D. J. Weatherall, "Phenotype-genotype relationships in monogenic disease: Lessons from the thalassaemias," *Nature Reviews Genetics*, vol. 2, no. 4, pp. 245–255, Apr 2001.

[67] J. Chapin and P. J. Giardina, “Thalassemia Syndromes,” in *Hematology: Basic Principles and Practice*, 7th ed., R. Hoffman, E.J. Benz, H. Heslop and J. Weitz, Elsevier Inc., 2018, pp. 546-570.e10.

[68] D. L. Loukopoulos, “Haemoglobinopathies,” in *Encyclopedia of Life Sciences*, Chichester, UK: John Wiley & Sons, Ltd, 2003.

[69] E. J. Benz and B. L. Ebert, “Hemoglobin Variants Associated With Hemolytic Anemia, Altered Oxygen Affinity, and Methemoglobinemias,” in *Hematology: Basic Principles and Practice*, 7th ed., R. Hoffman, E.J. Benz, H. Heslop and J. Weitz, Elsevier Inc., 2018, pp. 608–615.

[70] M. Cyrklaff, C. P. Sanchez, N. Kilian, C. Bisseye, J. Simpore, F. Frischknecht, and M. Lanzer, “Hemoglobins S and C interfere with actin remodeling in Plasmodium falciparum-infected erythrocytes,” *Science*, vol. 334, no. 6060, pp. 1283–1286, Dec 2011.

[71] R. L. Nagel, M. E. Fabry, and M. H. Steinberg, “The paradox of hemoglobin SC disease,” *Blood Reviews*, vol. 17, no. 3, pp. 167–178, Sep 2003.

[72] S. L. Thein, “Molecular basis of β thalassemia and potential therapeutic targets,” *Blood Cells, Molecules, and Diseases*, vol. 70. pp. 54–65, May 2018.

[73] N. Ben Salah, R. Bou-Fakhredin, F. Mellouli, and A. T. Taher, “Revisiting beta thalassemia intermedia: past, present, and future prospects,” *Hematology*, vol. 22, no. 10, pp. 607–616, Nov 2017.

[74] M. D. Cappellini, V. Viprakasit, and A. T. Taher, “An overview of current treatment strategies for β-thalassemia,” *Expert Opinion in Orphan Drugs*, vol. 2, no. 7, pp. 665–679, Jul 2014.

[75] H. Y. Luo and D. H. K. Chui, “Diverse hematological phenotypes of β-thalassemia carriers,” *Annals of the New York. Acadademy of Sciences*, vol. 1368, no. 1, pp. 49–55, Mar 2016.

[76] P. Faustino, P. Pacheco, P. Loureiro, P. J. Nogueira, and J. Lavinha, “The geographic pattern of β-thalassaemia mutations in the Portuguese population,” *British Journal of Haematology*, vol. 107, no. 4, pp. 903–904, Dec 1999.

[77] P. Faustino, L.Osório-Almeida, J. Barbot, D. Espírito-Santo, J. Gonçalves, L. Romão, M. C. Martins, M. M. Marques, and J. Lavinha, “Novel promoter and splice junction defects add to the genetic, clinical or geographic heterogeneity of β-thalassaemia in the Portuguese population,” *Human Geneics.*, vol. 89, no. 5, pp. 573–576, Jul 1992.

[78] A. Piperno, R. Mariani, C. Arosio, A. Vergani, S. Bosio, S. Fargion, M. Sampietro, D. Girelli, M. Fraquelli, D. Conte and G. Fiorelli, “Haemochromatosis in patients with β-thalassaemia trait,” *British Journal of Haematology*, vol. 111, no. 3, pp. 908–914, Dec 2000.

[79] K. G. A. Victor Hoffbrand , Paresh Vyas , Elias Campo , Torsten Haferlach, *Color Atlas of Clinical Hematology: Molecular and Cellular Basis of Disease*. 2019.

[80] C. S. Cardoso, P. Oliveira, G. Porto, C. Oberkanins, M. Mascarenhas, P. Rodrigues, F. Kury, and M. de Sousa, “Comparative study of the two more frequent HFE mutations (C282Y and H63D): Significant different allelic frequencies between the North and South of Portugal,” *European Journal of Human Genetics*, vol. 9, no. 11, pp. 843–848, Nov 2001.

[81] C. Spínola, A. Brehm, and H. Spínola, “Prevalence of H63D, S65C, and C282Y hereditary hemochromatosis gene variants in Madeira Island (Portugal),” *Annals of Hematology*, vol. 90, no. 1, pp. 29–32, Jan 2011.

[82] R. Galanello and A. Cao, “Alpha-thalassemia,” *Genetics in Medicine*, vol. 13, no. 2, pp. 83–88, Feb 2011.

[83] R. Das and P. Sharma, “Molecular Genetics of Thalassemia Syndromes,” *Colloquium Series on Genomic and Molecular Medicine* vol. 5, no. 1, pp. 1–57, Aug 2016.

[84] C. L. Harteveld and D. R. Higgs, “α-thalassaemia,” *Orphanet Journal of Rare Diseases*, vol. 5, no. 1, pp.13, Dec 2010.

[85] B. J. Bain, “Theα,β,δ andγ Thalassaemias and Related Conditions,” in *Haemoglobinopathy Diagnosis*, 2nd ed., Oxford, UK: Blackwell Publishing Ltd, 2007, pp. 63–138.

[86] C. Vrettou, G. Kakourou, T. Mamas, and J. Traeger-Synodinos, “Prenatal and preimplantation diagnosis of hemoglobinopathies,” *International Journal of Laboratory Hematology*, vol. 40, pp. 74–82, May 2018.

[87] J. Old, *Prevention and Diagnosis of Haemoglobinopathies a Short Guide for Health Professionals and Laboratory Scientists.*, 21. Nicosia, Cyprus: Thalassaemia International Federation TIF, 1-79, 2013, pp. 1–79.

[88] S. N. Costa, S. Madeira, M. A. Sobral, and G. Delgadinho, “Hemoglobinopatias em Portugal e a intervenção do médico de família,” *Revista Portuguesa de Clínica Geral e Familiar*, vol. 32, no. 6, pp. 416–424, Dec 2016.

[89] Association of Public Health Laboratories, “Hemoglobinopathies: Current Practices for Screening , Confirmation and Follow-up,” *Association of Public Health*, December, 2015.

[90] C. Shen, Y. M. Jiang, H. Shi, J.H. Liu, W. J. Zhou, Q. K. Dai and H. Yang, "Evaluation of indices in differentiation between iron deficiency anemia and β-thalassemia trait for chinese children," *Journal of Pediatric Hematology and Oncology*, vol. 32, no. 6, pp. e218–e222, Aug 2010.

[91] A. T. Taher and A. N. Saliba, "Iron overload in thalassemia: Different organs at different rates," *Hematology*, vol. 2017, no. 1, pp. 265–271, Dec 2017.

[92] Committee for standards in haematology, "Guideline: the laboratory diagnosis of haemaoglobinopathies," 1998.

[93] E. Kohne, "Hemoglobinopathies," *Deutches Äerzteblatt International*, vol. 108, no. 31–32, p. 532, Aug 2011.

[94] Y. Sauntharajah and E. P. Vichinsky, "Sickle Cell Disease: Clinical Features and Management," in *Hematology: Basic Principles and Practice*, 7th ed., R. Hoffman, E.J. Benz, H. Heslop and J. Weitz, Elsevier Inc., 2018, pp. 584-607.e5.

[95] R. Martins, I. Picanco, A. Fonseca, L. Ferreira, O. Rodrigues, M. Coelho, T. Seixas, A. Miranda, B. Nunes, L. Costa, and L. Romão, "The role of HFE mutations on iron metabolism in beta-thalassemia carriers," *Journal of Human Genetics*, vol. 49, no. 12, pp. 651–655, Dec 2004.

[96] M. C. Martins, G. Olim, J. Melo, H. A. Magalhaes, and M. O. Rodrigues, "Hereditary anaemias in Portugal: Epidemiology, public health significance, and control," *Journal of Medical Genetics*, vol. 30, no. 3, pp. 235–239, Mar 1993.

[97] S. E.Hunt, W. McLaren, L. Gil, A. Thormann, H. Schuilenburg, D. Sheppard, A. Parton, I. M. Armean, S. J. Trevanion, P. Flicek, and F. Cunningham, "Ensembl variation resources," Database, vol. 2018, pp.1-12, Jan 2028.

[98] M. Wick, G.-M. Pinggera, and P. Lehmann, *Clinical Aspects and Laboratory. Iron Metabolism, Anemias: Novel concepts in the anemias of malignancies and renal and rheumatoid diseases*. 2013.

[99] B. J. Bain, *Haemoglobinopathy Diagnosis*, 2nd Editio. Oxford, United Kingdom: John Wiley and Sons Ltd, 2020.

[100] J. Riou, J.Szuberski, C. Godart, H. Wajcman, J. L. Oliveira, J. D. Hoyer and J. Bardakdjian-Michau, "Precision of CAPILLARYS 2 for the detection of hemoglobin variants based on their migration positions," *American. Journal of Clinical Pathology*, vol. 149, no. 2, pp. 172–180, Feb 2018.

[101] T. Higgins, M. Mack, and A. Khajuria, "Comparison of two methods for the quantification and identification of hemoglobin variants," *Clinical Biochemistry.*, ol. 42, no. 7–8, pp. 701–705, May 2009.

[102] J. H. Wan, P. L.Tian, W. H. Luo, B. Y. Wu, F. Xiong, W. J. Zhou, X. C. Wei. and X. M. Xu, "Rapid determination of human globin chins using reversed-phase high-performance liquid chromatography," *Journal of Chromatography*, vol. 901, pp. 53–58, Jul 2012.

[103] M. R. Green and J. Sambrook, "Isolation of high-molecular-weight DNA using organic solvents," *Cold Spring Harbor Protocols*, vol. 2017, no. 4, pp. 356–359, Apr 2017.

[104] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, and H. Erlich, "Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction," *Cold Spring Harbor SLaboratory Press*, vol. 51, no. 1, pp. 263–273, Jan. 1986.

[105] M. T. Rahman, M. S. Uddin, R. Sultana, A. Moue, and M. Setu, "Polymerase Chain Reaction (PCR): A Short Review," *Anwer Khan Modern Medical College Journal*, vol. 4, no. 1, pp. 30–36, Feb 2013.

[106] D. Voytas, "Agarose gel electrophoresis.," *Current Protocols in Protein Science*, vol. Appendix 4, no. 4, pp. 629–638, May 2001.

[107] P. Y. Lee, J. Costumbrado, C. Y. Hsu, and Y. H. Kim, "Agarose gel electrophoresis for the separation of DNA fragments," *Journal of Visualized Experiments*, no. 62, p. e3923, Apr 2012.

[108] A. R. Thornhill and K. Snow, "Molecular diagnostics in preimplantation genetic diagnosis," *Journal of Molecular Diagnostics*, vol. 4, no. 1, pp. 11–29, Feb 2002.

[109] H. Lee, S. A. Morse and O. Olsvik, *Nucleic Acid Amplification Technologies Application to Disease Diagnosis.*, 1996.

[110] E. P. T. Lam, C. M. L. Chan, N. B. Y. Tsui, T. C. C Au, K. F. Wong, H. T. Wong, K. Y. Chiu, L. W. C. Chan, B. Y. M. Yung and S. C. C. Wong, "Clinical Applications of Molecular Technologies in Hematology," *Journal of Medical Diagnostic Methods*, vol. 02, no. 04, 2013.

[111] Y. Q. Zhou, G. F. Xiao, L. Y. Li, W. D. Li, Z.Y. Liu, L. F. Zhu, Q. H. Mo, X. J. Qu. and X. M. Xu, "Evaluation of clinical application of gap-PCR as a routine method for alpha-thalassemia carrier detection.," *Di Yi Jun Yi Da Xue Xue Bao*, vol. 22, no. 5, pp. 434–436, May 2002.

[112] C. Dodé, R. Krishnamoorthy, J. Lamb, and J. Rochette, "Rapid analysis of -α3.7 thalassaemia and αααanti 3.7

triplication by enzymatic amplification analysis," *British Journal of Haematology*, vol. 83, no. 1, pp. 105–111, Jan 1993.

[113] P. A. Davies and G. Gray, "Long-range PCR.," *PCR Mutation Detetion Protocols - Humana Press*, vol. 187, pp. 51–55, 2002.

[114] S. Uribe-Convers, J. R. Duke, M. J. Moore, and D. C. Tank, "A Long PCR–Based Approach for DNA Enrichment Prior to Next-Generation Sequencing for Systematic Studies," *Applications in Plant Science* vol. 2, no. 1, p. 1300063, Jan 2014.

[115] A. Kloss-Brandstätter, G. Erhart, C. Lamina, B. Meister, M. Haun, S. Coassin, M. Seifert, A. Klein-Franke, B. Paulweber, L. Kedenko and B. Kollerits, "Candidate gene sequencing of SLC11A2 and TMPRSS6 in a family with severe anaemia: Common SNPs, rare haplotypes, no causative mutation," *PLoS One*, vol. 7, no. 4, pp. 1–8, Apr 2012.

[116] G. Duta-Cornescu, A. Simon-Gruita, N. Constantin, F. Stanciu, M. Dobre, D. Banica, R. Tuduce, P. Cristea and V. Stoian, "A comparative study of ARMS - PCR and RFLP - PCR as methods for rapid SNP identification,"*Biotechnological*, vol. 14, no. 6, pp. 4845–4850, 2009.

[117] J. Hoorfar, B. Malorny, A. Abdulmawjood, N. Cook, M. Wagner, and P. Fach, "Practical Considerations in Design of Internal Amplification Controls for Diagnostic PCR Assays," *Journal of Clinical Microbiology*, vol. 42, no. 5, pp. 1863–1868, May 2004.

[118] L. Valenti, A. L. Fracanzani, R. Rametta, M. Fraquelli, G. Soverini, S. Pelusi, P. Dongiovanni, D. Conte and S. Fargion, "Effect of the A736V TMPRSS6 polymorphism on the penetrance and clinical expression of hereditary hemochromatosis," *Journal of Hepatology*, vol. 57, no. 6, pp. 1319–1325, Dec 2012.

[119] D. Baty, A. Terron Kwiatkowski, D. Mechan, A. Harris, M. J. Pippard, and D. Goudie, "Development of a multiplex ARMS test for mutations in the HFE gene associated with hereditary haemochromatosis," *Journal of Clinical Pathology*, vol. 51, no. 1, pp. 73–74, Jan 1998.

[120] L. Y. Chuang, C. H. Yang, K. H. Tsui, Y. H. Cheng, P. L. Chang, C. H. Wen and H. W. Chang, "Restriction enzyme mining for SNPs in genomes," *Anticancer Research*, vol. 28, no. 4 A, pp. 2001–2007, Jul 2008.

[121] T. Vincze, J. Posfai, and R. J. Roberts, "NEBcutter: A program to cleave DNA with restriction enzymes," *Nucleic Acids Research.*, vol. 31, no. 13, pp. 3688–3691, Jul 2003.

[122] L. Stuppia, I. Antonucci, G. Palka, and V. Gatta, "Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases," *International Journal of Molecular Science*, vol. 13, no. 3, pp. 3245–3276, Mar 2012.

[123] J. P. Schouten, C. J. McElgunn, R. Waaijer, D. Zwijnenburg, F. Diepvens, and G. Pals, "Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification," *Nucleic Acids Research*, vol. 30, no. 12, p. e57, Jun 2002.

[124] B. A. R. Andersson, M. E. Wering, H. Y. Luo, R. K. Basran, M. H. Steinberg, H. P. Smith and D. H. Chui, "Sickle cell disease due to compound heterozygosity for Hb S and a novel 7.7-kb β-globin gene deletion," *European Journal of Haematology*, vol. 78, no. 1, pp. 82–85, Jan 2007.

[125] B. Sikkema-Raddatz, L. F. Johansson, E. N. de Boer, R. Almomani, L. G. Boven, M. P van den Berg, K. Y. van Spaendonck-Zwarts, J. P. van Tintelen, R. H. Sijmons, J. D. Jongbloed and R. J. Sinke, "Targeted Next-Generation Sequencing can Replace Sanger Sequencing in Clinical Diagnostics," *Human Mutation*, vol. 34, no. 7, pp. 1035–1042, Jul 2013.

[126] M. L. Metzker, "Emerging technologies in DNA sequencing," *Genome Research*, vol. 15, no. 12. pp. 1767–1776, Dec 2005.

[127] S. Chakravorty and M. Hegde, "Gene and Variant Annotation for Mendelian Disorders in the Era of Advanced Sequencing Technologies," *Annual Reviews of Genomics and Human Genetic*, vol. 18, no. 1, pp. 229–256, Aug 2017.

[128] Illumina, "An Introduction to Next-Generation Sequencing Technology," no. illumina. pp. 1–16, 2016.

[129] P. Ewels, M. Magnusson, S. Lundin, and M. Käller, "MultiQC: Summarize analysis results for multiple tools and samples in a single report," *Bioinformatics*, vol. 32, no. 19, pp. 3047–3048, Oct 2016.

[130] S. Andrews, F. Krueger, A. Seconds-Pichon, F. Biggins, and S. Wingett, "FastQC. A quality control tool for high throughput sequence data. Babraham Bioinformatics," *Babraham Institute*, 2015. [Online]. Available: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/%0Ahttp://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/. [Accessed: 20-Feb-2020].

[131] V. A. Schneider, T. Graves-Lindsay, K. Howe, N. Bouk, H. C. Chen, P. A. Kitts, T. D. Murphy, K. D. Pruitt, F. Thibaud-Nissen, D. Albracht and R. S. Fulton, "Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly," *Genome Research*, vol. 27, no. 5, p. 849–864, May 2016.

[132] D. R. Zerbino, P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins and A. Gall, "Ensembl 2018," *Nucleic Acids Research.*, vol. 46, no. D1, pp. D754–D761, 2018.

[133] B. Langmead and S. L. Salzberg, "Fast gapped-read alignment with Bowtie 2," *Nature Methods*, vol. 9, no. 4, pp. 357–359, Apr 2012.

[134] H. Li and A. Wysoker, "The Sequence Alignment/Map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009.

[135] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky and M. A, DePristo, "The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data," *Genome Research*, vol. 20, no. 9, pp. 1297–1303, Sep 2010.

[136] J. T. Robinson, H. Thorvaldsdóttir, A. M. Wenger, A. Zehir, and J. P. Mesirov, "Variant review with the integrative genomics viewer," *Cancer Research*, vol. 77, no. 21, pp. e31–e34, Nov 2017.

[137] G. Benson, "Tandem repeats finder: A program to analyze DNA sequences," *Nucleic Acids Research*, vol. 27, no. 2, pp. 573–580, 1999.

[138] W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, P. Flicek and F. Cunningham, "The Ensembl Variant Effect Predictor," *Genome Biology*, vol. 17, no. 1, Dec 2016.

[139] L. Cartegni, S. L. Chew, and A. R. Krainer, "Listening to silence and understanding nonsense: Exonic mutations that affect splicing," *Nature Reviews Genetics*, vol. 3, no. 4, pp. 285–298, Apr 2002.

[140] F. O. Desmet, D. Hamroun, M. Lalande, G. Collod-Bëroud, M. Claustres, and C. Béroud, "Human Splicing Finder: An online bioinformatics tool to predict splicing signals," *Nucleic Acids Research*, vol. 37, no. 9, pp. e67–e67, May 2009.

[141] P. C. Ng and S. Henikoff, "Predicting the effects of amino acid substitutions on protein function," *Annual Review of Genomics and Human Genetics*, vol. 7, no. 1., pp. 61–80, Sep 2006.

[142] I. Adzhubei, D. M. Jordan, and S. R. Sunyaev, "Predicting functional effect of human missense mutations using PolyPhen-2," *Curr. Protocols in Human Genetics*, vol. 76, no. 1, pp. 7–21, Jan 2013.

[143] P. C. Ng and S. Henikoff, "SIFT: Predicting amino acid changes that affect protein function," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3812–3814, Jul 2003.

[144] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, "The Phyre2 web portal for protein modeling, prediction and analysis," *Nature Protocols*, vol. 10, no. 6, pp. 845–858, Jun 2015.

[145] S. Ittisoponpisan, S. A. Islam, T. Khanna, E. Alhuzimi, A. David, and M. J. E. Sternberg, "Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated?," *Journal of Molecular Biology*, vol. 431, no. 11, pp. 2197–2212, May 2019.

[146] RC Team - Statistical Computing, U. Vienna, 2017, "R: A language and environment for statistical computing."

[147] B. J. Bain, *Blood Cells: A Practical Guide*, 4th Editio. 2014.

[148] J. M. England and P. M. Fraser, "differentiation of iron deficiency from thalassæmia trait by routine blood-count," *The Lancet*, vol. 301, no. 7801, pp. 449–452, Mar 1973.

[149] B. M. Ricerca, S. Storti, G. d'Onofrio, S. Mancini, M. Vittori, S. Campisi, G. Mango and B. Bizzi, "Differentiation of iron deficiency from thalassaemia trait: a new approach.," *Haematologica*, vol. 72, no. 5, pp. 409–413, 1987.

[150] H. G. Schriever and P. C. Srivastava, "differentiation of thalassæmia minor from iron deficiency," *The Lancet*, vol. 302, no. 7821. pp. 154–155, Jul 1973.

[151] I. Shine and S. Lal, "A strategy to detect β-thalassæmia minor," *The Lancet*, vol. 309, no. 8013, pp. 692–694, 1977.

[152] J. D. Bessman and D. I. Feinstein, "Quantitative anisocytosis as a discriminant between iron deficiency and thalassemia minor," *Blood*, vol. 53, no. 2, pp. 288–293, 1979.

[153] B. M. Ricerca, S. Storti, G. d'Onofrio, S. Mancini, M. Vittori, S. Campisi, G. Mango and B. Bizzi, "Differentiation of iron deficiency from thalassaemia trait: a new approach.," *Haematologica*, vol. 72, no. 5, pp. 409–413, 1987.

[154] S. Jayabose, J. Giamelli, O. LevondogluTugal, C. Sandoval, F. Ozkaynak, and P. Visintainer, "Differentiating iron deficiency anemia from thalassemia minor by using an RDW-based index," *Journal of Pediatric Hematology and Oncology*, vol. 21, no. 4, p. 314, Jul 1999.

[155] R. Green and R. King, "A new red cell discriminant incorporating volume dispersion for differentiating iron deficiency anemia from thalassemia minor," *Blood Cells*, vol. 15, no. 3, pp. 481–495, 1989.

[156] O. Elmissani, S. A. K. Halil, and G.. Oberts, “Mean density of hemoglobin per liter of blood : a new hematologic parameter with an inherent discriminant function,” *Carden Jennings*, no. May, pp. 5–8, 1999.

[157] M. Sirdah, I. Tarazi, E. Al Najjar, and R. Al Haddad, “Evaluation of the diagnostic reliability of different RBC indices and formulas in the differentiation of the β-thalassaemia minor from iron deficiency in Palestinian population,” *International Journal of Laboratory Hematology*, vol. 30, no. 4, pp. 324–330, Aug 2008.

[158] M. A. Ehsani, E. Shahgholi, M. S. Rahiminejad, F. Seighali, and A. Rashidi, “A new index for discrimination between iron deficiency anemia and beta-thalassemia minor: Results in 284 patients,” *Pakistan Journal of Biological Sciences*, vol. 12, no. 5, pp. 473–475, May 2009.

[159] M. O’Flaherty, K. Lerum, P. Martin, and D. Grassi, “Low agreement for assessing the risk of postoperative deep venousthrombosis when deciding prophylaxis strategies: A study using clinicalvignettes,” *BMC Health Services Research*, vol. 2, pp. 1–3, Dec 2002.

[160] A. Vehapoglu, G. Ozgurhan, A. D. Demir, S. Uzuner, M. A. Nursoy, S. Turkmen and A. Kacan, “Hematological indices for differential diagnosis of beta thalassemia trait and iron deficiency anemia,” *Anemia*, vol. 2014, Jan 2014.

[161] X. Robin, N. Turck, A. Hainard, T. Natalia, J.C. Sanchez, and M. Mueller, “pROC: an open-source package for R and S+ to analyze and compare ROC curves,” *BMC Bioinformatics*, vol. 8, pp. 12–77, Dec 2011.

[162] J. J. M. L. Hoffmann, E. Urrechaga, and U. Aguirre, “Discriminant indices for distinguishing thalassemia and iron deficiency in patients with microcytic anemia: A meta-analysis,” *Clinical Chemistry and Laboratory Medicine*, vol. 53, no. 12, pp. 1883–1894, Nov 2015.

[163] M. S. Rose, “Epitaph for the M.C.H.C,” *British Medical Journal*, vol. 4, no. 5780, p. 169, Oct.1971.

[164] W. W. W. De Jong, L. N. Went, and L. F. Bernini, “Haemoglobin Leiden: Deletion of β6 or 7 Glutamic Acid,” *Nature*, vol. 220, no. 5169. pp. 788–790, Nov 1968.

[165] R. L. Nagel, R. F. Rieder, R. M. Bookchin, and G. W. James, “Some functional properties of hemoglobin Leiden,” *Biochemical and Biophysical Research Communications*, vol. 53, no. 4, pp. 1240–1245, Aug 1973.

[166] J. G. Adams, L. A. Boxer, R. L. Baehner, B. G. Forget, G. A. Tsistrakis, and M. H. Steinberg, “Hemoglobin Indianapolis (β112[G14] arginine). An unstable β-chain variant producing the phenotype of severe β-thalassemia,” *Journal of Clinical Investigation*, vol. 63, no. 5, pp. 931–938, May 1979.

[167] R. F. Arakaki and B. Changcharoen, “Glycemic Assessment in a Patient with HB Leiden and Type 2 Diabetes,” *AACE Clinical Case Reports*, vol. 2, no. 4, pp. e307–e310, Sep. 2016.

[168] R. F. Rieder and G. W. James, “Imbalance in α and β globin synthesis associated with a hemoglobinopathy,” *Journal of Clinical Investigations*, vol. 54, no. 4, pp. 948–956, Oct 1974.

[169] M. M. Perlman, B. G. Wiltshire, K. Stevens, R. Cassel and H. Lehmann, “Haemoglobin Leiden in a South African negro. A case report.,” *South Africa Medical Journal.*, vol. 59, no. 15, pp. 537–540, Apr 1981.

[170] L. E. Lie-Injo, Z. I. Randhawa, J. Ganesan, D. Peterson, and J. P. Kane, “Hb leiden-β° thalassemia in a chinese with severe hemolytic anemia,” *American Journal of Hematology*, vol. 2, no. 4, pp. 335–342, Jan 1977.

[171] R. Hellman, “When are HBA1C Values Misleading?,” *AACE Clin. Case Reports*, vol. 2, no. 4, pp. e377–e379, Sep 2016.

[172] D. Juritie, I. Ruzdic, Z. Beer, G. D. Efremov, R. Casey, and H. Lehmann, “Hemoglobin Leiden [β6 or 7 (a3 or a4) GLU → 0] in a yugoslavian woman arisen by a new mutation,” *Hemoglobin*, vol. 7, no. 3, pp. 271–277, 1983.

[173] P. Sanguansermsri, D. Shimbhu, R. Wongvilairat, C. Pimsorn, and T. Sanguansermsri, “Spontaneous mutation of the hemoglobin Leiden (beta 6 or 7 Glu-->0) in a Thai girl,” *Haematologica*, vol. 88, no. 12, Jan 2003.

[174] C. V. Tondo, f. M. Salzano, and D. L. Rucknagel, “Hemoglobin Porto Alegre, a possible polymer of normal hemoglobin in a Caucasian Brazilian family.,” *American Journal of Human Genetics*, vol. 15, no. 3, pp. 265–279, Sep.1963.

[175] M. Seid-Akiiavan, M. Ayres, F. M. Salzano, W. P. Winter, and D. L. Rucknagel, “Two more examples of HB porto alegre, α2β29 ser→cys in Belém, Brazil,” *Human Hereditary*, vol. 23, no. 2, pp. 175–181, 1973.

[176] M. S. Gonçalves, M. F. Sonati, M. Kimura, V. R. Arruda, F.F. Costa, J.F., Nechtman and T. A. Stoming, “Association of hb santa ANA [α2β288(f4)LEU→PRO] and hb Porto alegre [α2β29(a6)SER→CYS] in a brazilian female,” *Hemoglobin*, vol. 18, no. 3, pp. 235–239, Jan 1994.

[177] G. Martinez , F. Lima, M. Wade, M. Estrada, B. Colombo, L. Heredero and H. Granda, “Haemoglobin Porto Alegre in a Cuban family,” *Journal of Medical Genetics*, vol. 14, no. 6, pp. 422–425, Dec 1977.

[178] J. J. Malcorra-Azpiazu, J. B. Wilson, T. P. Molchanova, D. D. Pobedimskaya, and T. H. J. Huisman, “HB Porto alegre or α2β29(a6) SER→CYS in unrelated families of the canary islands,” *Hemoglobin*, vol. 17, no.

5, pp. 457–461, Jan 1993.

[179] P. Faustino, A. Miranda, I. Picanço, E. M. Kimura, F. Ferreira Costa, and M. D. F. Sonati, "Genetic studies suggest a novel Portuguese origin for hemoglobin Porto Alegre," *Haematologica*, vol. 89, no. 8, pp. 1009–1010, Jan 2004.

[180] V. Baudin-Creuza, C. Fablet, F. Zal, B. N. Green, D. Promé, M. C. Marden, J. Pagnier and H. Wajcman, "Hemoglobin Porto Alegre forms a tetramer of tetramers superstructure," *Protein Science*, vol. 11, no. 1, pp. 129–136, Apr 2009.

[181] L. Lojo, P. Santiago-Borrero, E. Rivera, J. Renta, and C. L. Cadilla, "Asymptomatic child heterozygous for hemoglobin S and hemoglobin Pôrto Alegre," *Pediatric Blood Cancer*, vol. 56, no. 3, pp. 458–459, Mar 2011.

[182] C. V. Tondo, "Increased erythrocyte glutathion reductase activity in a hemoglobin Porto Alegre (β9 Ser → Cys) carrier," *Biochemical and Biophysical Research Communications.*, vol. 105, no. 4, pp. 1381–1388, Apr 1982.

[183] C. Tondo, J. Bonaventura, C. Bonaventura, M. Brunori, G. Amiconi, and E. Antonini, "Functional properties of hemoglobin pôrto alegre (α2Aβ29 Ser→Cys) and the reactivity of its extra cysteinyl residue," *BBA - Protein Structure*, vol. 342, no. 1, pp. 15–20, Mar 1974.

[184] J. Bonaventura and A. Riggs, "Polymerization of hemoglobins of mouse and man: Structural basis," *Science*, vol. 158, no. 3802, pp. 800–802, Nov 1967.

[185] M. S. Figueiredo, "The importance of hemoglobin A2 determination," *Revista Brasileira de Hematologia e Hemoterapia*, vol. 37, no. 5, pp. 287–289, Sep 2015.

[186] D. Edoh, C. Antwi-Boasiako, and D. Amuzu, "Fetal hemoglobin during infancy and in sickle cell adults," *African Health Science.*, vol. 6, no. 1, pp. 51–54, Mar 2006.

[187] M. Creary, D. Williamson, and R. Kulkarni, "Sickle cell disease: Current activities, public health implications, and future directions," *Journal of Women's Health*, vol. 16, no. 5, pp. 575–582, Jun 2007.

[188] J. B. Clegg, M. A. Naughton, and D. J. Weatherall, "An improved method for the characterization of human haemoglobin mutants: Identification of α2β295GLU, haemoglobin N (Baltimore)," *Nature*, vol. 207, no. 5000, pp. 945–947, Aug 1965.

[189] S. K. Ballas and D. K. Park, "Biosynthetic evidence for stability of hb n-Baltimore," *Hemoglobin*, vol. 9, no. 5, pp. 489–494, Jan 1985.

[190] T. S. Millimono, K. M. Loua, S. L. Rath, L. Relvas, C. Bento, M. Diakite, M. Jarvis, N. Daries, L. M. Ribeiro, L. Manco and J. S. Kaeda, "High prevalence of hemoglobin disorders and glucose-6-phosphate dehydrogenase (G6PD) deficiency in the Republic of Guinea (West Africa)," *Hemoglobin*, vol. 36, no. 1, pp. 25–37, Feb 2012.

[191] M. H. Steinberg, J. G. Adams, and B. J. Dreiling, "Alpha Thalassaemia in Adults with Sickle-Cell Trait," *British Journal of Haematology*, vol. 30, no. 1, pp. 31–37, May 1975.

[192] I. Akinsheye, A. Alsultan, N. Solovieff, D. Ngo, C. T. Baldwin, P. Sebastiani, D. H. Chui and M. H.Steinberg, "Fetal hemoglobin in sickle cell anemia," *Blood*, vol. 118, no. 1. The American Society of Hematology, pp. 19–27, Jul 2011.

[193] D. C. Rees, T. N. Williams, and M. T. Gladwin, "Sickle-cell disease," *The Lancet*, vol. 376, no. 9757, pp. 2018–2031, Dec 2010.

[194] E. A. Nyenwe and J. N. Fisher, "A mistaken diagnosis of type 2 diabetes due to hemoglobin N-baltimore," *American Journal of Medical Science*, vol. 336, no. 6, pp. 524–526, Dec 2008.

[195] M. B. Rumaney, V. J. N. Bitoungui, A. A. Vorster, R. Ramesar, A. P. Kengne, J. Ngogangand A. Wonkam, "The co-inheritance of alpha-thalassemia and sickle cell anemia is associated with better hematological indices and lower consultations rate in Cameroonian patients and could improve their survival," *PLoS One*, vol. 9, no. 6, Jun 2014.

[196] J. G. Adams, W. T. Morrison, R. L. Barlow, and M. H. Steinberg, "HB mississippi [β44(CD3)serâg]: A new variant with anomalous properties," *Hemoglobin*, vol. 11, no. 5, pp. 435–452, Jan 1987.

[197] G. R. Honig, M. Shamsuddin, R. G. Mason, L. N. Vida, L. M. Tremaine, G. E. Tarr and N. T. Shahidi, "Hemoglobin Nigeria (α-81 Ser→Cys): A new variant associated with α-thalassemia," *Blood*, vol. 55, no. 1, pp. 131–137, 1980.

[198] D. Rund, D. Filon, N. Bloch Isenberg, and N. Goldschmidt, "KLF1 Mutations Are Not Common in Israel but Can Explain Occasional Cases of Elevated HbA2 or Very Elevated Fetal Hemoglobin," *Blood*, vol. 132, no. Supplement 1, pp. 3640–3640, Nov 2018.

[199] D. Damgaard, P. H. Nissen, L. G. Jensen, G. G. Nielsen, A. Stenderup, M. L. Larsen and o. Faergeman, "Detection of large deletions in the LDL receptor gene with quantitative PCR methods," *BMC Medical*

Genetics, vol. 6, no. 1, p. 15 Apr 2005.

[200] S. Bandyopadhyay, B. C. Mondal, P. Sarkar, S. Chandra, M. K. Das, and U. B. Dasgupta, "Two β-globin cluster-linked polymorphic loci in thalassemia patients of variable levels of fetal hemoglobin," *European Journal of Haematology.*, vol. 75, no. 1, pp. 47–53, Jul 2005.

[201] I. Moumni, M. B. Mustapha, I. B. Mansour, A. Zoraï, K. Douzi, S. Sassi, D. Chaouachi, F. Mellouli, M. Bejaoui and S. Abbes, "Fetal Hemoglobin in Tunisian Sickle Cell Disease Patient: Relationship with Polymorphic Sequences Cis to the β-Globin Gene," *Indian Jounal of Hematology and Blood Transfusion*, vol. 32, no. 1, pp. 114–119, Mar 2016.

[202] T. K. T. Nguyen, P. Joly, C. Bardel, M. Moulsma, N. Bonello-Palot, and A. Francina Alain, "The XmnI Gγ polymorphism influences hemoglobin F synthesis contrary to BCL11A and HBS1L-MYB SNPs in a cohort of 57 β-thalassemia intermedia patients," *Blood Cells, Molecules and Diseases.*, vol. 45, no. 2, pp. 124–127, Aug 2010.

[203] A. A. Tantawy, N. G. Andrawes, A. Ismaeil, S. A. Kamel, and W. Emam, "Prevalence of Xmnl Gγ polymorphism in Egyptian patients with β-thalassemia major," *Annals of Saudi Medicine*, vol. 32, no. 5, pp. 487–491, Sep 2012.

[204] H. Nemati, Z. Rahimi, and G. Bahrami, "The Xmn1 polymorphic site 5′ to the Gγ gene and its correlation to the Gγ:Aγ ratio, age at first blood transfusion and clinical features in β-Thalassemia patients from Western Iran,", *Journal of Arab Child,l* vol. 37, no. 1, pp. 159–164, Jun 2010.

[205] M. Zavodna, A. Bagshaw, R. Brauning, and N. J. Gemmell, "The Accuracy, Feasibility and Challenges of Sequencing Short Tandem Repeats Using Next-Generation Sequencing Platforms," *PLoS One*, vol. 9, no. 12, p. e113862, Dec 2014.

[206] B. Modan, "Polycythemia. A review of epidemiological and clinical aspects," *Journal of Chronic Diseases*, vol. 18, no. 7. Elsevier, pp. 605–645, Jul 1965.

[207] M. C. Garel, Y. Blouquit, N. Arous, J. Rosa, "Hb Strasbourg α2β220 (B2) Val → Asp: A variant at the same locus as Hb olympia β20 Val → Met," *FEBS Letters*, vol. 72, no. 1, pp. 1–4, Dec 1976.

[208] F. G. Fernandez, A. Villegas, P. Ropero, M. D. Carreño, E. Anguita, M. Polo, A. Pascual and A. Henández, "Haemoglobinopathies with high oxygen affinity. Experience of Erythropathology Cooperative Spanish Group," *Annals of Hematology*, vol. 88, no. 3, pp. 235–238, Sep 2009.

[209] D. J. Taylor and T. Lind, "Red cell mass during and after normal pregnancy" *BJOG An International Journal of Obstetrics and Gynaecology*, vol. 86, no. 5, pp. 364–370, May 1979.

[210] S. Rahbar, D. Beale, W. A. Isaacs, and H. Lehmann, "Abnormal Haemoglobins in Iran. Observation of a New Variant—Haemoglobin J Iran (α2β2 77 His —> Asp)," *British Medical Journal*, vol. 1, no. 5541, pp. 674–677, Mar 1967.

[211] S. Rahbar and H. F. Bunn, "Association of hemoglobin H disease with Hb J-Iran (β77 His → Asp): Impact on subunit assembly," *Blood*, vol. 70, no. 6, pp. 1790–1791, 1987.

[212] J. Delanoe-Garin, M. D. Rhoda, C. T. Craescu, J. Bardakjian, Y. Blouquit, C. Lacombe, N. Arous, C. Poyart, D. Ganeval, R. Girot and J. Riou, "Hemoglobin j Iran α2 β2 77 (EF1) his à asp in: A russian-armenian family," *Hemoglobin*, vol. 10, no. 4, pp. 365–378, Jan 1986.

[213] W. R. Romero, M. Castillo, M. A. Chaves, G. F. Saenz, L. H. Gu, J. B. Wilson, E. Baysal, N. S. Smetanina, J. Y. Leonova and T. H. J. Huisman, "Hb Costa Rica or · 2 , 2 77(EF1)His→Arg: the first example of a somatic cell mutation in a globin gene," *Human Genetics*, vol. 97, no. 6, pp. 829–833, May 1996.

[214] I. Chanarin, I. R. McFadyen, and R. Kyle, "The physiological macrocytosis of pregnancy," *BJOG An International Journal of Obstetrics and Gynaecology*, vol. 84, no. 7, pp. 504–508, Jul 1977.

[215] C. Vernochet, S. M. Caucheteux, and C. Kanellopoulos-Langevin, "Bi-directional Cell Trafficking Between Mother and Fetus in Mouse Placenta," *Placenta*, vol. 28, no. 7, pp. 639–649, Jul 2007.

[216] R. L. Gross, J. V. O. Reid, P. M. Newberne, B. Burgess, R. Marston, and W. Hift, "Depressed cell mediated immunity in megaloblastic anemia due to folic acid deficiency," *The American Journal of Clinical Nutrition*, vol. 28, no. 3, pp. 225–232, Mar 1975.

[217] E. Peuchant, J. L. Brun, V. Rigalleau, L. Dubourg, M. J. Thomas, J. Y. Daniel, J. J. Leng and H. Gin, "Oxidative and antioxidative status in pregnant women with either gestational or type 1 diabetes," *Clinical Biochemistry*, vol. 37, no. 4, pp. 293–298, Apr 2004.

[218] A. U. Kurtoğlu, E. Eren, Ö. Erkal, E. Kurtoğlu, and N. Yilmaz, "Hemoglobin J-Iran (HBB c.232C > G): Interference with the assay of HbA1c," *Clinica chimica acta*, vol. 465. pp. 80–81, 2017.

[219] J. D. Bessman, "Microcytic Polycythemia: Frequency of Nonthalassemic Causes," *Journal of the American Medical Association*, vol. 238, no. 22, pp. 2391–2392, Nov 1977.

[220] N. F. Olivieri, Z. Pakbaz, and E. Vichinsky, “Hb E/beta-thalassaemia: A common & clinically diverse disorder,” *Indian Journal of Medical Research*, vol. 134, no. 10, pp. 522–531, Oct 2011.

[221] E. Vichinsky, “Hemoglobin e syndromes.,” *Hematology*, vol. 2007, no. 1, pp. 79–83, Jan 2007.

[222] J. Ha, R. Martinson, S. K. Iwamoto, and A. Nishi, “Hemoglobin E, malaria and natural selection,” *Evolution, Medicine, and Public Health*, vol. 2019, no. 1, pp. 232–241, 2019.

[223] V. F. Fairbanks, G. S. Gilchrist, B. Brimhall, J. A. Jereb, and E. C. Goldston, “Hemoglobin E trait reexamined: A cause of microcytosis and erythrocytosis,” *Blood*, vol. 53, no. 1, pp. 109–115, 1979.

[224] M. Nusrat, B. Moiz, A. Nasir, and M. Rasool Hashmi, “An insight into the suspected HbA2’ cases detected by high performance liquid chromatography in Pakistan,” *BMC Research Notes*, vol. 4, Dec 2011.

[225] L. A. Frohman, “Controversy about treatment of growth hormone-deficient adults: A commentary,” *Annals of Internal Medicine*, vol. 137, no. 3, pp. 202–204, Aug 2002.

[226] S. Mooney, “Bioinformatics for Geneticists,” *Briefings in Informatics*, vol. 5, no. 2, pp. 209–210, 2004.

[227] A. Anna and G. Monika, “Splicing mutations in human genetic disorders: examples, detection, and confirmation,” *Journal of Applied Genetics*, vol. 59, no. 3. pp. 253–268, Aug 2018.

[228] R. L. Batista, A. di Santi Rodrigues, M. Y. Nishi, N. L. Gomes, J. A. D. F. Junior, D. R. de Moraes, L. R. Carvalho, E. M. F. Costa, S. Domenice and B. B. Mendonca “A recurrent synonymous mutation in the human androgen receptor gene causing complete androgen insensitivity syndrome,” *Journal of Steroid Biochemistry and Molecular Miology*, vol. 174, pp. 14–16, Nov 2017.

[229] M. Hicks, I. Bartha, J. Di Iulio, J. Craig Venter, and A. Telenti, “Functional characterization of 3D protein structures informed by human genetic diversity,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 18, pp. 8960–8965, Apr 2019.

[230] I. Pichler, C. Minelli, S. Sanna, T. Tanaka, C. Schwienbacher, S. Naitza, E. Porcu, C. Pattaro, F. Busonero, A. Zanon and A. Maschio, “Identification of a common variant in the TFR2 gene implicated in the physiological regulation of serum iron levels,” *Human Molecular Genetics*, vol. 20, no. 6, pp. 1232–1240, Mar 2011.

[231] F. Guillem and B. Grandchamp, “Mutations du gène TMPRSS6 dans les anémies de type IRIDA et effets des variants fréquents sur les paramètres hématologiques,” *Hematologie*, vol. 17, no. 5. pp. 357–364, Nov 2011.

[232] S. P. Dion, F. Béliveau, L. P. Morency, A. Désilets, R. Najmanovich, and R. Leduc, “Functional diversity of TMPRSS6 isoforms and variants expressed in hepatocellular carcinoma cell lines,” *Scientific Reports*, vol. 8, no. 1, pp. 2–10, Aug 2018.

[233] T. Sato, S. Iyama, K. Murase, Y. Kamihara, K. Ono, S. Kikuchi, K. Takada, K. Miyanishi, Y. Sato, R. Takimoto and M. Kobune, “Novel missense mutation in the TMPRSS6 gene in a Japanese female with iron-refractory iron deficiency anemia,” *International Journal of Hematology*, vol. 94, no. 1, pp. 101–103, Jul 2011.

[234] P. L. Lee, J. C. Barton, P. L. Khaw, S. Y. Bhattacharjee, and J. C. Barton, “Common TMPRSS6 mutations and iron, erythrocyte, and pica phenotypes in 48 women with iron deficiency or depletion,” *Blood Cells, Molecules and Diseases*, vol. 48, no. 2, pp. 124–127, Feb 2012.

[235] E. Poggiali, F. Andreozzi, I. Nava, D. Consonni, G. Graziadei, and M. D. Cappellini, “The role of TMPRSS6 polymorphisms in iron deficiency anemia partially responsive to oral iron treatment,” *American Journal of Hematology*, vol. 90, no. 4, pp. 306–309, Apr 2015.

[236] B. Batar, I. Bavunoglu, Y. Hacioglu, M. Cengiz, T. Mutlu, S. Yavuzer, H. Yavuzer, D. C. Ercelebi, D. Erhan, S. Unal and A. Tunckale, “The role of TMPRSS6 gene variants in iron-related hematological parameters in Turkish patients with iron deficiency anemia,” *Gene*, vol. 673, pp. 201–205, Oct 2018.

[237] A. E. Donker, P. P. T. Brons, and D. W. Swinkels, “Microcytic anaemia with low transferrin saturation, increased serum hepcidin and non-synonymous TMPRSS6 variants: Not always iron-refractory iron deficiency anaemia,” *British Journal of Haematology*, vol. 169, no. 1, pp. 150–151, Apr 2015.

[238] S. K. Hanks, A. M. Quinn, and T. Hunter, “The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains,” *Science*, vol. 241, no. 4861, pp. 42–52, Jul 1988.

[239] S. J. Hubbard, K. H. Gross, and P. Argos, “Intramolecular cavities in globular proteins,” *Protein Engineering, Design and Selection*, vol. 7, no. 5, pp. 613–626, May 1994.

[240] I. S. Moreira, P. A. Fernandes, and M. J. Ramos, “Hot spots - A review of the protein-protein interface determinant amino-acid residues,” *Proteins: Structure, Function and Genetics*, vol. 68, no. 4. pp. 803–812, Sep 2007.

[241] R. A. Laskowski, J. D. Stephenson, I. Sillitoe, C. A. Orengo, and J. M. Thornton, “VarSite: Disease variants and protein structure,” *Protein Science.*, vol. 29, no. 1, pp. 111–119, Jan 2020.

[242] C. J. McDonald, L. Ostini, N. Bennett, N. Subramaniam, J. Hooper, G. Velasco, D. F. Wallace and V. N. Subramaniam, "Functional analysis of matriptase-2 mutations and domains: Insights into the molecular basis of iron-refractory iron deficiency anemia," *American Journal of Physiology-Cell Physiology*, vol. 308, no. 7, pp. C539–C547, Apr 2015.

[243] S. Chakrabarty, J. P. Kahler, M. A. T. van de Plassche, R. Vanhoutte, and S. H. L. Verhelst, "Recent advances in activity-based protein profiling of proteases," in *Current Topics in Microbiology and Immunology*, vol. 420, 2019, pp. 253–281, Springer, Cham 2018.

[244] M. D. Oberst, C. A. Williams, R. B. Dickson, M. D. Johnson, and C. Y. Lin, "The activation of matriptase requires its noncatalytic domains, serine protease domain, and its cognate inhibitor," *Journal of Biological Bhemistry*, vol. 278, no. 29, pp. 26773–26779, Jul 2003.

[245] L. Silvestri, F. Guillem, A. Pagani, A. Nai, C. Oudin, M. Silva, F. Toutain, C. Kannengiesser, C. Beaumont, C. Camaschella and B. Grandchamp, "Molecular mechanisms of the defective hepcidin inhibition in TMPRSS6 mutations associated with iron-refractory iron deficiency anemia," *Blood*, vol. 113, no. 22, pp. 5605–5608, May 2009.

[246] C. Y. Lin, I. C Tseng, F. P. Chou, S. F. Su, Y. W. Chen, M. D. Johnson and R. B. Dickson, "Zymogen activation, inhibition, and ectodomain shedding of matriptase," *Frontiers in Bioscience*, vol. 13, no. 2. pp. 621–635, Jan 2008.

[247] D. Shinta, Asmarinah, C. Adhiyanto, M. K. Htet, and U. Fahmida, "The association of tmprss6 gene polymorphism and iron intake with iron status among under-two-year-old children in Lombok, Indonesia," *Nutrients*, vol. 11, no. 4, Apr 2019.

[248] C. E. McLaren, S. McLachlan, C. P. Garner, C.D Vulpe, V. R. Gordeuk, J. H. Eckfeldt, P.C Adams, R. T. Acton, J. A. Murray, C. Leiendecker-Foster and B. M. Snively, "Associations between single nucleotide polymorphisms in iron-related genes and iron status in multiethnic populations," *PLoS One*, vol. 7, no. 6, p. e38339, Jun 2012.

[249] T. Tanaka, C. N. Roy, W. Yao, A. Matteini, R. D. Semba, D. Arking, J. D. Walston, L. P. Fried, A. Singleton, , J. Guralnik and G. R. Abecasis, "Agenome-wide association analysis of serum iron concentrations," *Blood*, vol. 115, no. 1, pp. 94–96, Jan 2010.

[250] E. Zintzaras, "Impact of Hardy-Weinberg equilibrium deviation on allele-based risk effect of genetic association studies and meta-analysis," *European Journal of Epidemiology*, vol. 25, no. 8, pp. 553–560, Aug 2010.

[251] H. Hoorfar, S. Sadrarhami, A. H. Keshteli, S. K. Ardestani, M. Ataei, and A. Moafi, "Evaluation of Iron Status by Serum Ferritin Level in Iranian Carriers of Beta Thalassemia Minor," *International Journal for Vitamin and Nutrition Research*, vol. 78, no. 45, pp. 204–207, May 2008.

[252] T. K. Dolai, K. S. Nataraj, N. Sinha, S. Mishra, M. Bhattacharya, and M. K. Ghosh, "Prevalance of iron deficiency in thalassemia minor: A study from tertiary hospital," *Indian Journal of Hematology and Blood Transfusion*, vol. 28, no. 1, pp. 7–9, Mar 2012.

[253] G. Salanti, G. Amountza, E. E. Ntzani, and J. P. A. Ioannidis, "Hardy-Weinberg equilibrium in genetic association studies: An empirical evaluation of reporting, deviations, and power," *European Journal of Human Genetics*, vol. 13, no. 7, pp. 840–848, Apr 2005.

[254] H. W. Deng, W. M. Chen, D. J. Schaid, and S. Jacobsen, "Re: 'Biased tests of association: Comparisons of allele frequencies when departing from Hardy-Weinberg proportions'," *American Journal of Epidemiology*, vol. 151, no. 3. pp. 335–337, 2000.

[255] G. M. Clarke, C. A. Anderson, F. H. Pettersson, L. R. Cardon, A. P. Morris, and K. T. Zondervan, "Basic statistical analysis in genetic case-control studies," *Nature Protocols*, vol. 6, no. 2, pp. 121–133, Feb 2011.

[256] A. Townsend and H. Drakesmith, "Role of HFE in iron metabolism, hereditary haemochromatosis, anaemia of chronic disease, and secondary iron overload," *The Lancet*, vol. 359, no. 9308, pp. 786–790, Mar 2002.

[257] I. J. Kullo, K. Ding, H. Jouni, C. Y. Smith, and C. G. Chute, "A genome-wide association study of red blood cell traits using the Electronic Medical Record," *PLoS One*, vol. 5, no. 9, Sep 2010.

[258] J. C. Chambers, W. Zhang, Y. Li, J. Sehmi, M. N. Wass, D. Zabaneh, C. Hoggart, H. Bayele, M. I. McCarthy, L. Peltonen and N. B. Freimer, "Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels," *National Genetics*, vol. 41, no. 11, pp. 1170–1172, Oct2009.

[259] J. C. Barton, L. F. Bertoli, and B. E. Rothenberg, "Peripheral blood erythrocyte parameters in hemochromatosis: Evidence for increased erythrocyte hemoglobin content," *Journal of Laboratory and Clinical Medicine*, vol. 135, no. 1, pp. 96–104, Jan 2000.

[260] M. Politou, V. Kalotychou, M. Pissia, Y. Rombos, N. Sakellaropoulos, and G. Papanikolaou, "The impact of the mutations of the HFE gene and of the SLC11A3 gene on iron overload in Greek thalassemia intermedia and βs/ βthal anemia patients," *Haematologica*, vol. 89, no. 4, pp. 490–492, Apr 2004.

[261] M. P. Roth, P. Giraldo, G. Hariti, E. S. Poloni, A. Sanchez-Mazas, G. F. De Stefano, J. M. Dugoujon and H. Coppin, "Absence of the hemochromatosis gene Cys282Tyr mutation in three ethnic groups from Algeria (Mzab), Ethiopia, and Senegal," *Immunogenetics*, vol. 46, no. 3, pp. 222–225, July 1997.

[262] M. Sirdah, K. Al Mghari, A. H. Abuzaid, and R. M. Al Haddad, "Should sex differences be considered when applying mathematical indices and formulas for discriminating β- thalassemia minor from iron deficiency?," *Practical Laboratory Medicine*, vol. 11, pp. 1–9, Jul 2018.

[263] L. Hamel, "Model Assessment with ROC Curves," in *Encyclopedia of Data Warehousing and Mining, Second Edition*, pp. 1316–1323, 2011.

[264] J. F. Matos, L. M. S. A. Dusse, R. V. B. Stubbert, M. R. Ferreira, W. Coura-Vital, A. P. S. M. Fernandes, J. R. de Faria, K. B. G. Borges and M. D. G. Carvalho, "Comparison of discriminative indices for iron deficiency anemia and β thalassemia trait in a Brazilian population," *Hematology*, vol. 18, no. 3, pp. 169–174, May 2013.

[265] I. F. Estevão, J. Peitl, and C. R. Bonini-Domingos, "Serum ferritin and transferrin saturation levels in β0 and β + thalassemia patients," *Genetics and Molecular Research*, vol. 10, no. 2, pp. 632–639, May 2011.

[266] E. Urrechaga, "Red blood cell microcytosis and hypochromia in the differential diagnosis of iron deficiency and β-thalassaemia trait," *International Journal of Laboratory Hematology*, vol. 31, no. 5, pp. 528–534, Oct 2009.

[267] G. Ntaios, A. Chatzinikolaou, Z. Saouli, F. Girtovitis, M. Tsapanidou, G. Kaiafa, Z. Kontoninas, A. Nikolaidou, C. Savopoulos, I. Pidonia and S. Alexiou-Daniel, "Discrimination indices as screening tests for β-thalassemic trait," *Annals of Hematology*, vol. 86, no. 7, pp. 487–491, Jul 2007.

[268] C. Beyan, K. Kaptan, and A. Ifran, "Discrimination indices as screening tests for beta-thalassemia trait [1]," *Annals of Hematology*, vol. 87, no. 1. pp. 61–62, 2008.

[269] E. Urrechaga and J. J. M. L. Hoffmann, "Critical appraisal of discriminant formulas for distinguishing thalassemia from iron deficiency in patients with microcytic anemia," *Clinincal Chemistry and Laboratory Medicine*, vol. 55, no. 10, pp. 1582–1591, Oct 2017.

[270] A. Demir, N. Yarali, T. Fisgin, F. Duru, and A. Kara, "Most reliable indices in differentiation between thalassemia trait and iron deficiency anemia," *Pediatric International*, vol. 44, no. 6, pp. 612–616, Dec 2002

[271] E. Miri-Moghaddam and N. Sargolzaie, "Cut off determination of discrimination indices in differential diagnosis between iron deficiency anemia and β- thalassemia minor," *International Journal of Hematology-oncology and Stem Cell Research*, vol. 8, no. 2, pp. 27–32, Apr 2014.

[272] G. Ntaios, A. Chatzinikolaou, Z. Saouli, F. Girtovitis, M. Tsapanidou, G. Kaiafa, Z. Kontoninas, A. Nikolaidou, C. Savopoulos, I. Pidonia and S. Alexiou-Daniel, "Discrimination indices as screening tests for β-thalassemic trait," *Annals of Hematology.*, vol. 86, no. 7, pp. 487–491, Jul 2007.

[273] A. Janel, L. Roszyk, C. Rapatel, G. Mareynat, M. G. Berger, and A. F. Serre-Sapin, "Proposal of a score combining red blood cell indices for early differentiation of beta-thalassemia minor from iron deficiency anemia," *Hematology*, vol. 16, no. 2, pp. 123–127, Mar 2011.

[274] S. N. Qasem and A. Mosavi, "Novel Meta-Heuristic Model for Discrimination between Iron Deficiency Anemia and Β-Thalassemia with CBC Indices Based on Dynamic Harmony Search (DHS)," Jan 2020.

[275] C. S. Thom, C. F. Dickson, D. A. Gell, and M. J. Weiss, "Hemoglobin variants: Biochemical properties and clinical correlates," *Cold Spring Harbor Perspectives in Medicine*, vol. 3, no. 3, Mar 2013.

[276] I. G. Denisov and S. G. Sligar, "Nanodiscs for structural and functional studies of membrane proteins," *Nature Structural and Molecular Biology*, vol. 23, no. 6. Nature Publishing Group, pp. 481–486, Jun 2016.

[277] P. Bhatia, A. Singh, A. Hegde, R. Jain, and D. Bansal, "Systematic evaluation of paediatric cohort with iron refractory iron deficiency anaemia (IRIDA) phenotype reveals multiple TMPRSS6 gene variations," *British Jorunal of Haematology*, vol. 177, no. 2, pp. 311–318, Feb 2017.

[278] A. P. Capra, E. Ferro, L. Cannavò, M. A. La Rosa, and G. Zirilli, "A child with severe iron-deficiency anemia and a complex TMPRSS6 genotype," *Hematology*, vol. 22, no. 9, pp. 559–564, Apr 2017.

[279] A. E. Donker, C. C. Schaap, V. M. Novotny, R. Smeets, T. M. Peters, B. L. van den Heuvel, M. F. Raphael, A. W. Rijneveld, I. M. Appel, A. J. Vlot and A. B. Versluijs, "Iron refractory iron deficiency anemia: a heterogeneous disease that is not always iron refractory," *American Journal of Hematology*, vol. 91, no. 12, pp. E482–E490, Set 2016.

[280] S. Çakmaklı, Ç Kaplan, M. Uzunoğlu, M. Büyükbayram, E. Görgülü, N. Ö. Zarif and E. Y. Keskin, "A novel homozygous nonsense mutation (p . Y78 *) in TMPRSS6 gene causing iron-refractory iron deficiency anemia (IRIDA) in two siblings," *Turkish Journal Pediatric*, vol. 62, no. 1, pp. 103–108, Jan 2020.

[281] C. Kannengiesser, F. Guillem, L. Silvestri, C. Oudin, A. Marfaing, L. Chaiba-Berrouche, J. Donadieu, F. Toutain, M. Silva, B. Isidor and G. Marguerite, "allelic heterogeneity of TMPRSS6 mutations in IRIDA: 61," *American Journal of Hematology*, vol. 84, no. 4, June 2009.

[282] F. Guillem, S. Lawson, C. Kannengiesser, M. Westerman, C. Beaumont, and B. Grandchamp, “Two nonsense mutations in the TMPRSS6 gene in a patient with microcytic anemia and iron deficiency,” *Blood*, vol. 112, no. 5, pp. 2089–2091, Sep 2008.

[283] A. J. Ramsay, V. Quesada, M. Sanchez, C. Garabaya, M. P. Sardà, M. Baiget, A. Remacha, G. Velasco and C. López-Otín, “Matriptase-2 mutations in iron-refractory iron deficiency anemia patients provide new insights into protease activation mechanisms,” *Human Molecular Genetics*, vol. 18, no. 19, pp. 3673–3683, July 2009.

[284] L. De Falco, F. Totaro, A. Nai, A. Pagani, D. Girelli, L. Silvestri, C. Piscopo, N. Campostrini, C. Dufour, F. A. Manjomi and M. Minkov, “Novel TMPRSS6 mutations associated with Iron-refractory Iron Deficiency Anemia (IRIDA),” *Human Mutation.*, vol. 31, no. 5, Mar 2010.

[285] S. Altamura, F. D’Alessio, B. Selle, and M. U. Muckenthaler, “A novel TMPRSS6 mutation that prevents protease auto-activation causes IRIDA,” *Biochemical Journal*, vol. 431, no. 3, pp. 363–371, Oct 2010.

[286] J. Allison, L. Drury, and J. B. Ford, “Two Novel TMPRSS6 Variants in a Compound Heterozygous Child with Iron Refractory Iron Deficiency Anemia (IRIDA),” *Journal Pediatric Hematology and Oncology*, vol. 42, no. 4, pp. e238–e239, May 2019.

[287] E. Beutler, C. Van Geet, D. M. W. M. te Loo, T. Gelbart, K. Crain, J. Truksa and P. L. Lee, “Polymorphisms and mutations of human TMPRSS6 in iron deficiency anemia,” *Blood Cells, Molecules and Diseases*, vol. 44, no. 1, pp. 16–21, Jan 2010.

[288] M. A. Melis, M. Cau, R. Congiu, G. Sole, S. Barella, A. Cao, M. Westerman, M. Cazzola and R. Galanello, “A mutation in the TMPRSS6 gene, encoding a transmembrane serine protease that suppresses hepcidin production, in familial iron deficiency anemia refractory to oral iron,” *Haematologica*, vol. 93, no. 10, pp. 1473–1479, Oct 2008.

[289] M. Cau, R. Galanello, N. Giagu, and M. A. Melis, “Responsiveness to oral iron and ascorbic acid in a patient with IRIDA,” *Blood cells, Molecules Diseases*, vol. 48, no. 2, pp. 121–123, Feb 2012.

[290] H. S. Choi, H. R. Yang, S. H. Song, J. Y. Seo, K. O. Lee and H. J. Kim, “A Novel Mutation Gly603Arg of TMPRSS6 in a Korean Female With Iron-Refractory Iron Deficiency,” *Pediatric Blood Cancer*, vol. 50, no. 5, pp. 1018–1025, May 2012.

[291] H. M. Yaish, C. P Farrell, R. D. Christensen, B. C. MacQueen, L. K. Jackson, J. Trochez-Enciso, J. Kaplan, D. M. Ward, W. K. Salah and J. D. Phillips, “Two novel mutations in TMPRSS6 associated with iron-refractory iron deficiency anemia in a mother and child,” *Blood Cells, Molecules and Diseases*, vol. 65, pp. 38–40, Apr 2017.

[292] N. Nie, J. Shi, Y. Shao, X. Li, M. Ge, J. Huang, J. Zhang, Z. Huang, D. Li and Y. Zheng, “A novel tri-allelic mutation of TMPRSS6 in iron-refractory iron deficiency anaemia with response to glucocorticoid,” *British Journal of Haematology*, vol. 166, no. 2, pp. 300–303, Mar 2014.

[293] I. Tchou, M. Diepold, P. A. Pilotto, D. Swinkels, M. Neerman-Arbez, and P. Beris, “Haematologic data, iron parameters and molecular findings in two new cases of iron-refractory iron deficiency anaemia,” *European Journal of Haematology*, vol. 83, no. 6, pp. 595–602, Nov 2009.

[294] A. Jaspers, J. Caers, G. Le Gac, C. Ferec, Y. Beguin, and G. Fillet, “A novel mutation in the CUB sequence of matriptase-2 (TMPRSS6) is implicated in iron-resistant iron deficiency anaemia (IRIDA),” *British Journal of Haematology*, vol. 160, no. 4, pp. 564–565, Dec 2013.

[295] R. Pellegrino, M. Coutinho, D. D'Ascola, A. M. Lopes, A. Palmieri, F. Carnuccio, M. Costa, G. Zecchina, G. Saglio, E. Costa and J. Barbot, “Two novel mutations in the tmprss6 gene associated with iron-refractory iron-deficiency anaemia (irida) and partial expression in the heterozygous form,” *British Journal of Haematology*, vol. 158, no. 5, pp. 666–668, Jun 2012.

[296] K. Kodama, A. Noguchi, H. Adachi, M. Hebiguchi, M. Yano, and T. Takahashi, “Novel mutation in the TMPRSS6 gene with iron-refractory iron deficiency anemia,” *Pediatrics International*, vol. 56, no. 4, pp. e41–e44, Sep 2014.

[297] K. E Finberg, M. M Heeney, D. R. Campagna, Y. Aydınok, H. A. Pearson, K. R. Hartman, M. M. Mayo, S. M. Samuel, J. J. Strouse, K. Markianos and N. C. Andrew, “Mutations in TMPRSS6 cause iron-refractory iron deficiency anemia (IRIDA),” *Nature Genetics*, vol. 40, no. 5, pp. 569–571, Apr 2008.

[298] Y. Xiong, Z. Wu, W. Yang, X. Zhao, G. Peng, K. Tang, Z. Tian, H. Xing, Q. Rao, M. Wang and J. Wang, “A novel splicing mutation of TMPRSS6 in a Chinese child with iron-refractory iron deficiency anaemia,” *British Journal of Haematology*, vol. 171, no. 4, pp. 647–649, Apr 2015.

[299] E. Yılmaz-Keskin, E. Sal, L. de Falco, M. Bruno, A. Iolascon, Ü. Koçak and İ Yenicesu, “Is the acronym IRIDA acceptable for slow responders to iron in the presence of TMPRSS6 mutations?,” *Turkish Journal Pediatrics*, vol. 55, no. 5, pp. 479–484, Sep 2013.

[300] F. Guillem, C. Kannengiesser, C. Oudin, A. Lenoir, P. Matak, J. Donadieu, B. Isidor, F. Méchinaud,P. Aguilar-Martinez, C. Beaumont and S. Vaulont, "Inactive matriptase-2 mutants found in IRIDA patients still repress hepcidin in a transfection assay despite having lost their serine protease activity," *Human Mutation*, vol. 33, no. 9, pp. 1388–1396, May 2012.

[301] L. Silvestri, F. Guillem, A. Pagani, A. Nai, C. Oudin, M. Silva, F. Toutain, C. Kannengiesser, C. Beaumont, C. Camaschella, and B. Grandchamp, "Molecular mechanisms of the defective hepcidin inhibition in TMPRSS6 mutations associated with iron-refractory iron deficiency anemia," *Blood*, vol. 113, no. 22, pp. 5605–5608, May 2009.

[302] J. Pinto, G. Nobre de Jesus, M. Palma Anselmo, L. Gonçalves, D. Brás, J. Madeira Lopes, J. Meneses, R. Victorino and P. Faustino, "Iron refractory iron deficiency anemia in dizygotic twins due to a novel TMPRSS6 gene mutation in addition to polymorphisms associated with high susceptibility to develop ferropenic anemia," *Journal of Investigative Medicine High Impact Case Reports*, vol. 5, no. 2, pp. 0–3, Apr 2017.

[303] S. Çakmakli, C. Acipayam, M. N. Yenmiş Inan, and H. Doğan, "Iron Refractory Iron Deficiency Anemia Due to 374 Base Pairs Deletion in the TMPRSS6 Gene," *Journal of Pediatric Hematology/Oncology*, vol. 41, no. 5, pp. e333–e335, July 2019.

[304] K. Lehmberg, R. Grosse, M. U. Muckenthaler, S. Altamura, P. Nielsen, H. Schmid, U. Graubner, F. Oyen, W. Zeller, R. Schneppenheim and G. E. Janka, "Administration of recombinant erythropoietin alone does not improve the phenotype in iron refractory iron deficiency anemia patients," *Annals of Hematology*, vol. 92, no. 3, pp. 387–394, Nov 2013.

[305] M. Parlare, A. Ferrao, and L. Relvast, "TMPRSS6 gene—Two new nonsense mutations associated with IRIDA," *Haematologica vol. 95,* Suppl. 2, *pp.704-704*, Jun 2010.

[306] I. Tchou, N. A. Marguerite, and B. Photis, "Cryptic splice site usage leading to truncated TMPRSS6 is responsible for iron refractory iron deficiency anaemia in an Italian Family," *European Journal of Haematology*, vol. 86, no. 2, pp. 178–179, Feb 2010.

[307] M. L. H. Cuijpers, E. T. G. Wiegerinck, R. Brouwer, T. J. M. De Witte, and D. W. Swinkels, "IJzergebreksanemie door een matriptase 2-mutatie," *Ned Tijdschr Geneeskd*, vol. 154, no. 32, pp. 1509–1513, Jun 2010.

[308] M. Traglia, D. Toniolo, C. Camaschella, A. Nai, A. Pagani, L. Silvestri, N. Campostrini, M. Corbella and D. Girelli, "Brief report TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum *TMPRSS6*," *Blood* vol. 118, no. 16, pp. 4459–4463, Aug 2011.

[309] E. S. Edison, R. Athiyarath, T. Rajasekar, M. Westerman, A. Srivastava and M. Chandy, "A novel splice site mutation c.2278 (-1) G>C in the TMPRSS6 gene causes deletion of the substrate binding site of the serine protease resulting in refractory iron deficiency anaemia," *British Journal of Haematology*, vol. 147, no. 5, pp. 763–766, Aug 2009.

[310] P. R. Finley, and N. W. Tietz. *Clinical guide to laboratory tests*,4th ed, Jun 2006.

[311] "Prescrição e determinação do hemograma – Normas de Orientação Clínica,", 063/2011, DGS Lisboa, Sep 2013. [Online]. Available: https://nocs.pt/prescricao-determinacao-hemograma/.

[312] SEBIA, Manual on "Capillarys hemoglobin (E) using the CAPILLARYS 2 flex-piercing instrument," Accessed: May, 23, 2018. pp. 92–103 [Online]. Available: http://www.ilexmedical.com/files/Sebia%20inserts/CAPILLARYS_HEMOGLOBIN(E).pdf

[313] D. A. Colantonio, L. Kyriakopoulou, M. K. Chan, C. H. Daly, D. Brinc, A. A. Venner, M. D. Pasic, D. Armbruster, and K. Adeli, "Closing the gaps in pediatric laboratory reference intervals: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children," *Clinical Chemistry*, vol. 58, no. 5, pp. 854-868, May 2012.

[314] O. P. Soldin, L. H. Bierbower, J. J. Choi, J. J. Choi, S. Thompson-Hoffman, and S. J. Soldin, "Serum iron, ferritin, transferrin, total iron binding capacity, hs-CRP, LDL cholesterol and magnesium in children; new reference intervals using the Dade Dimension Clinical Chemistry System," *Clinica Chimica Acta*, vol. 342, no. 1–2, pp. 211–217, Apr 2004.

[315] M. Aldrimer, P. Ridefelt, P. Rödöö, F. Niklasson, J. Gustafsson, and D. Hellberg, "Population-based pediatric reference intervals for hematology, iron and transferrin," *Scandinavian Journal of Clinical and Laboratory investigation*, vol. 73, no. 3, pp. 253–261, Apr 2013.

[316] Karbasy, K., Lin, D.C., Stoianov, A., Chan, M.K., Bevilacqua, V., Chen, Y. and Adeli, K., "Pediatric reference value distributions and covariate-stratified reference intervals for 29 endocrine and special chemistry biomarkers on the Beckman Coulter Immunoassay Systems: A CALIPER study of healthy community children," *Clinical Chemistry and Laboratory Medicine*, vol. 54, no. 4, pp. 643–657, Apr 2016.

Supplementary material

Table S.1 – Identified mutations in *TMPRSS6* gene associated with IRIDA phenotype in literature and ClinVar database.

ID	VT	rs_ID	NG_ 02856.2	NM_153609.3	NP_705837.1	AltN	MC	Region	PD	Clin_sig	Ref
1	SNV	rs5756516	g.10905G>A	c.-120G>A	-	-	UPS	UPS	-	B	[277], [278]
2	SNV	rs5756515	g.10912T>C	c.-113T>C	-	-	5' UTR	5'UTR	-	B	[278]
3	SNV	rs11704654	g.11218G>A	c.99G>A	p.Pro33=	P33P	Syn	Ex2	N-tail	B/LB	[115], [232]
4	SNV	rs750274321	g.11273C>T	c.154C>T	p.Arg52Cys	R52C	Miss	Ex2	N-tail	US	ClinVar
5	SNV	rs745388319	g.112289T>A	c.170T>A	p.Phe57Tyr	F57Y	Miss	Ex2	TM	US	ClinVar
6	SNV	-	g.16009G>A	c.230-6G>A	-	IVS2-6G>A	Splice	Int2	-	P	[279]
7	SNV	rs1569024289	g.16019C>G	c.234C>G	p.Tyr78Ter	Y78X	Non	Ex3	TM-SEA	P	[280]
8	SNV	rs147397866	g.16055C>A	c.270C>A	p.Gly90=	G90G	Syn	Ex3	SEA	US	ClinVar
9	SNV	rs369875437	g.16120G>A	c.335G>A	p.Arg112Leu	R112L	Miss	Ex3	SEA	P	[281]
10	SNV	rs199474803	g.16125G>A	c.340G>A	p.Glu114Lys	E114K	Miss	Ex3	SEA	P	[282]
11	SNV	rs267607121	g.16138C>A	c.353C>A	p.Ala118Asp	A118D	Miss	Ex3	SEA	P	[283]
12	SNV	rs1430692214	g.17907A>G	c.422A>G	p.Tyr141Cys	Y141C	Miss	Ex4	SEA	P	[284], [285]
13	SNV	rs375681801	g.18480A>G	c.438A>G	p.Gly146=	G146G	Syn	Ex5	SEA	US	ClinVar
14	SNV	-	g.18487A>C	c.445A>C	p.Thr149Pro	T149P	Miss	Ex5	SEA	P	[286]
15	SNV	rs72163489	g.18520G>A	c.478G>A	p.Glu160Lys	E160K	Miss	Ex5	SEA	US	ClinVar
16	INDEL	-	g.18653delC	c.611delC	p.Leu166fs	L166fs	FS	Ex5	SEA	P	[284], [287]
17	SNV	rs377665035	g.18667T>C	c.616+9T>C	-	IVS5+9T>C	Splice	Int5	-	US	ClinVar
18	SNV	rs776877803	g.18990T>C	c.635T>A	p.Ile212Thr	I212T	Miss	Ex6	SEA-CUB	P	[284]
19	SNV	-	g.19014G>C	c.658+1G>C	-	IVS6+1G>C	Splice	In6	-	P	[288], [289]
20	SNV	-	g.19014G>T	c.658+1G>T	-	IVS6+1G>T	Splice	In6	-	P	[290]
21	SNV	rs375234781	g.24774C>T	c.659-8C>T	-	IVS6-8C>T	Splice	Int6	-	US	ClinVar
22	INDEL	rs1296123434	g.24782delG	c.659delG	p.GLyfs	Gfs	Splice	Ex7	-	P	[291]
23	INDEL	-	g.24913delG	c.790delG	p.Gln229fs	Q229fs	FS	Ex7	CUB1	P	[284]

Continued

ID	VT	rs_ID	NG_ 02856.2	NM_153609.3	NP_705837.1	AltN	MC	Region	PD	Clin_sig	Ref
24	SNV	rs199474802	g.24827T>C	c.704T>C	p.Leu235Pro	L235P	Miss	Ex7	CUB1	P	[282]
25	SNV	rs769301726	g.24843C>T	c.720C>T	p.His240=	H240H	Syn	Ex7	CUB1	US	ClinVar
26	INDEL	-	g.24978delG	c.855delG	p.Trp247fs	W247fs	FS	Ex7	CUB1	P	[284]
27	SNV	rs769301726	g.24880A>G	c.757A>G	p.Lys253Glu	K253E	Miss	Ex7	CUB1	B	[233], [234], [292]
28	SNV	rs776180387	g.24935G>A	c.812G>A	p.Arg271Gln	R271Q	Miss	Ex7	CUB1	P	[284]
29	SNV	rs5995378	g.24986C>T	c.863C>T	p.Ser288Leu	S288L	Miss	Ex7	CUB1	US	[289]
30	SNV	rs757356137	g.24987G>T	c.863-1G>T	-	IVS7+1G>T	Splice	Ex7	-	P	[279]
31	SNV	rs201148397	g.28146G>T	c.865G>T	p.Val289Leu	V289L	Miss	Ex8	CUB1	US	ClinVar
32	SNV	rs370192027	g.28190G>A	c.909G>A	p.Ala303=	A303A	Syn	Ex8	CUB1	US	ClinVar
33	SNV	rs1373272804	g.28192C>T	c.911C>T	p.Ser304Leu	S304L	Miss	Ex8	CUB1	P	[284], [293]–[295]
34	INDEL	rs780218999	g.28292delA	c.1000+11delA	-	IVS8+11delA	Splice	Int8	-	US	ClinVar
35	INDEL	-	g.29139_30192del1054	c.1001-586_1146del1054	-	-	Splice	In8	-	P	[293]
36	INDEL	-	g.29714_29723del9	c.1001-11_1001-3del9	-	-	Splice	In8	-	P	[295]
37	SNV	rs185482276	g.29743C>T	c.1019C>T	p.Thr340Met	T340M	Miss	Ex9	CUB2	US	ClinVar
38	SNV	rs762921295	g.29785C>T	c.1061C>T	p.Pro354Leu	P354L	Miss	Ex9	CUB2	P	[296]
39	SNV	rs137853121	g.29789C>A	c.1065C>A	p.Tyr355Ter	Y355X	Non	Ex9	CUB2	P	[282], [297]
40	SNV	rs2111833	g.29807G>A	c.1083G>A	p.Ser361=	S361S	Syn	Ex9	CUB2	B	ClinVar
41	SNV	rs886057492	g.29820T>C	c.1096T>C	p.Cys366Arg	C366R	Miss	Ex9	CUB2	US	ClinVar
42	SNV	-	g.29828C>A	c.1104C>A	p.His369Asn	H369N	Miss	Ex9	CUB2	P	[294]
43	SNV	rs976867694	g.29837G>A	c.1113G>A	p.Thr371=	T371T	Splice	Ex9	CUB2	P	[298]
44	SNV	rs137853122	g.30225T>G	c.1179T>G	p.Tyr393Ter	Y393X	Non	Ex10	CUB2	P	[282]
45	SNV	-	g.430269G>A	c.1223+1G>A	-	IVS10+1G>A	Splice	In10	-	P	[299]
46	SNV	rs529716198	g.39280C>G	c.1224-4C>G	-	IVS10-4C>G	Splice	Int10	-	US	ClinVar
47	SNV	rs146266448	g.39293C>T	c.1233C>T	p.Gly411=	G411G	Syn	Ex11	CUB2	US	ClinVar
48	SNV	rs881144	g.39314C>T	c.1253C>T	p.Tyr418=	Y418Y	Syn	Ex11	CUB2	B/LB	ClinVar
49	SNV	rs199474804	g.39313A>G	c.1253A>G	p.Tyr418Cys	Y418C	Miss	Ex11	CUB2	P	[300]

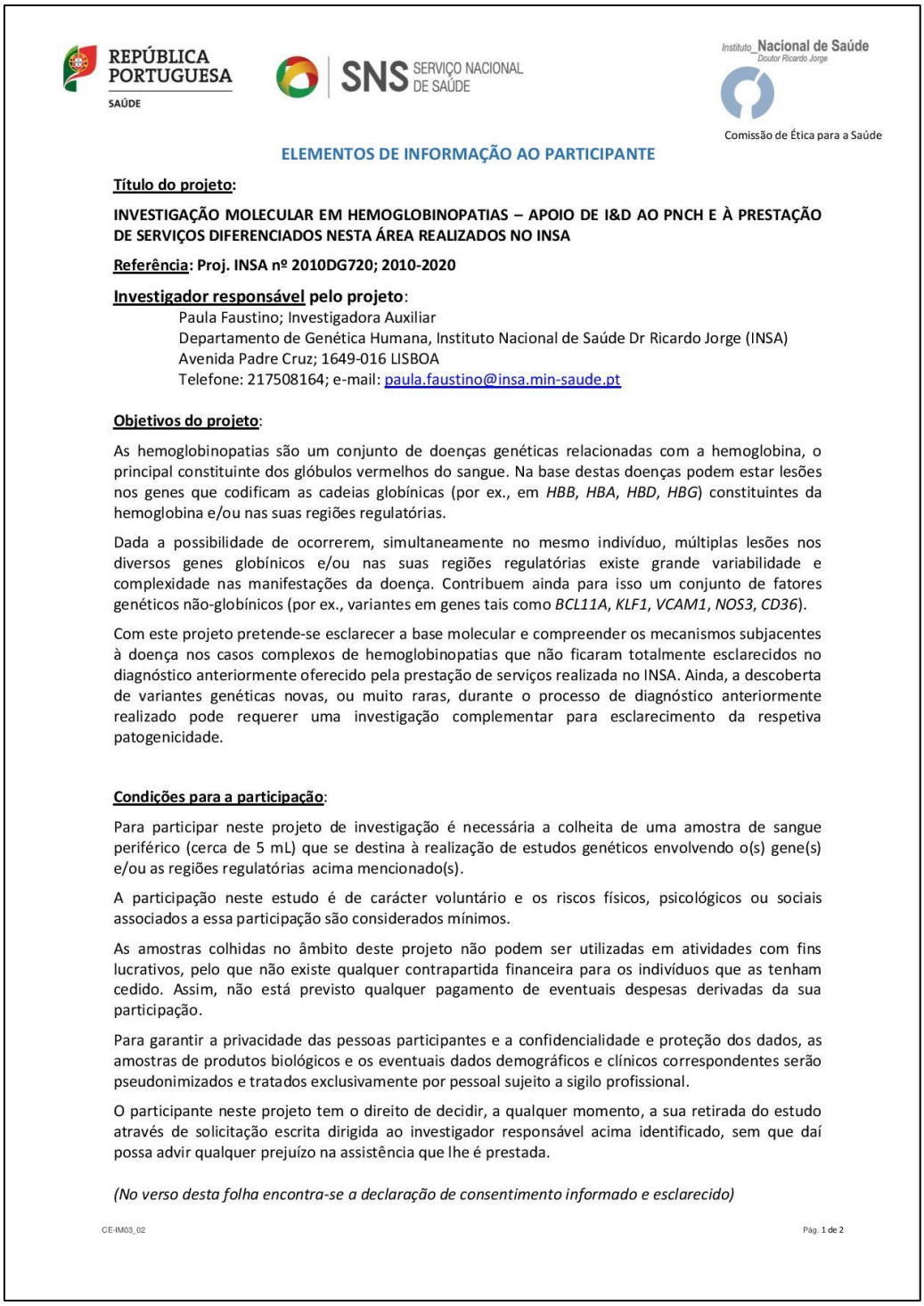
Continued

ID	VT	rs_ID	NG_ 02856.2	NM_153609.3	NP_705837.1	AltN	MC	Region	PD	Clin_sig	Ref
50	SNV	rs137853119	g.39384G>A	c.1324G>A	p.Gly442Arg	G442R	Miss	Ex11	CUB2	P	[56], [297], [301]
51	SNV	rs117576908	g.39396C>T	c.1336C>T	p.Arg446Trp	R446W	Miss	Ex11	CUB2	B/LB	[287]
52	SNV	rs1015906640	g.39433A>T	c.1369+4A>T	-	IVS11+4A>T	Splice	Int11	-	P	[281], [302]
53	INDEL	-	g.39443delA	c.1383delA	p.Glu461fs	E461fs	FS	Ex12	LDLR_A1	P	[297]
54	SNV	rs79816125	g.39964C>T	c.1468+10C>T	-	IVS12+10C>T	Splice	Int12	-	CI	ClinVar
55	SNV	rs138915369	g.40970C>T	c.1520C>T	p.Pro507Leu	Pp507l	Miss	Ex13	LDLR_A2	US	ClinVar
56	SNV	-	g.43591C>A	c.1642C>A	p.Cys510Ser	C510S	Miss	Ex13	LDLR_A2	P	[284], [299]
57	SNV	rs137853120	g.41011G>A	c.1561G>A	p.Asp521Asn	D521N	Miss	Ex13	LDLR_A2	P	[297], [301]
58	SNV	rs4820268	g.41013C>T	c.1563C>T	p.Asp521=	D521D	Syn	Ex13	LDLR_A2	B	ClinVar
59	SNV	rs387907018	g.41014G>A	c.1564G>A	p.Glu522Lys	E522K	Miss	Ex13	LDLR_A2	P	[56], [301]
60	SNV	rs786205058	g.41033G>A	c.1582+1G>A	-	IVS13+1G>A	Splice	Int13	-	P	[297]
61	INDEL	-	g.43774_44147del	c.1700-138_1869+67del	-	Exon15 del	Splice	-	-	P	[303]
62	SNV	rs76970337	g.43603G>A	c.1654G>A	p.Asp552Asn	D552N	Miss	Ex14	LDLR_A3	US	[279]
63	SNV	rs143441292	g.43631C>T	c.1682C>A	p.Ser561Ter	S561X	Non	Ex14	AS	P	[284]
64	INDEL	-	-	-	p.Ser570fs	S570fs	FS	Ex15	AS	P	[284]
65	SNV	-	g.43980T>C	c.1768T>C	p.Ser590Ter	S590X	Non	Ex15	SP	P	[304]
66	SNV	rs137853123	g.44007C>T	c.1795C>T	p.Arg599Ter	R599X	Non	Ex15	SP	P	[282], [299]
67	SNV	rs769083817	g.44019G>T	c.1807G>C	p.Gly603Arg	G603R	Miss	Ex15	SP	P	[290], [305]
68	Del	rs786205057	g.44025delG	c.1813delG	p.Ala605Profs	A605Pfs	FS	Ex15	SP	P	[281], [297]
69	SNV	-	g.45199C>G	c.1869-21C>G	-	IVS15-21C>G	Splice	Int15	-	P	[295]
70	SNV	-	g.45219C>G	c.1869-1C>G	-	IVS15-1C>G	Splice	Int15	-	P	[284], [293]
71	INDEL	-	g.45219C>G	c.1869-1C>G	p.Asp622fs	D622fs	FS	Ex16	SP	P	[297]
72	SNV	-	g.44080G>C	c.1868G>C	p.Ser623Thr	S623T	Splice	Ex16	SP	P	[306]
73	INDEL	rs869320724	g.45255_45256dupGC	c.1904_1905dupGC	p.Lys636fs	K636Afs	FS	Ex16	SP	P	[299], [304]
74	SNV	-	-	-	p.Leu674Phe	L674F	Miss	Ex16	SP	P	[287]
75	INDEL	rs786205060	g.45406_45409dupCCCC	c.2055_2058dupCCCC	p.Pro686fs	p.Pro686fs	FS	Ex16	SP	P	[283]

Continued

ID	VT	rs_ID	NG_ 02856.2	NM_153609.3	NP_705837.1	AltN	MC	Region	PD	Clin_sig	Ref
76	SNV	rs375189210	g.45436C>G	c.2085C>G	p.Phe695Leu	F695L	Miss	Ex16	SP	US	ClinVar
77	SNV	rs766931065	g.45456G>T	c.2105G>T	p.Cys702Phe	C702F	Miss	Ex16	SP	P	[279], [307]
78	SNV	rs115310908	g.45483G>T	c.2132G>T	p.Arg711Leu	R711L	Miss	Ex16	SP	CI	ClinVar
79	INDEL	-	g.45491G>C	C.2140+1G>C	p.Gly713fs	G713fs	FS	Ex16	SP	P	[297]
80	SNV	rs786205059	g.45492G>C	c.2140+1G>C	-	IVS16+1G>C	Splice	Int16	-	P	[297]
81	SNV	rs143878335	g.47615C>T	c.2154C>T	p.Asn718=	N718N	Syn	Ex17	SP	US	ClinVar
82	SNV	rs855791	g.47668T>C	c.2207T>C	p.Val736Ala	V736A	Miss	Ex17	SP	B	[57], [308]
83	SNV	rs2235321	g.47678C>T	c.2217C>T	p.Tyr739=	Y739Y	Syn	Ex17	SP	B	[234]
84	SNV	rs1363716662	g.48325G>C	c.2278-1G>C	-	IVS17-1G>C	Splice	Int17	-	P	[309]
85	SNV	rs199474805	g.48341C>G	c.2293C>G	p. Pro765Ala	P765A	Miss	Ex18	SP	P	[300]
86	SNV	rs776069764	g.48368C>T	c.2320C>T	p.Arg774Cys	R774C	Miss	Ex18	SP	P	[297]
87	SNV	rs73886915	g.48394C>T	c.2346C>T	p.Ser782=	S782S	Syn	Ex18	SP	LB	ClinVar
88	SNV	rs139105452	g.48431G>A	c.2383G>A	p.Val795Ile	V795I	Miss	Ex18	SP	LB	[307]
89	SNV	rs199957731	g.48519T>C	c.*35T>C	-	-	3'UTR	3'UTR	-	US	ClinVar
90	SNV	rs886057491	g.48624C>T	c.*140C>T	-	-	3'UTR	3'UTR	-	US	ClinVar
91	SNV	rs886057490	g.48787T>G	c.*303T>G	-	-	3'UTR	3'UTR	-	US	ClinVar
92	SNV	rs560949530	g.48983C>T	c.*499C>T	-	-	3'UTR	3'UTR	-	US	ClinVar
93	SNV	rs117575523	g.48987C>G	c.*503C>G	-	-	3'UTR	3'UTR	-	US	ClinVar
95	SNV	rs139014458	g.49060C>A	c.*576C>A	-	-	3'UTR	3'UTR	-	US	ClinVar

VT – Variant type; **SNV** – Single-nucleotide variant; **INDEL** – Insertion or deletion; **rs_ID** – reference single nucleotide polymorphism ID; **NG** –Reference sequence based on a genomic region; **NM** – Reference sequence based on a protein coding RNA (mRNA); **NP** – Reference sequence based on a protein (amino acid) sequence; **AltN** – Alternative name; **MC** – Molecular consequence; **UPS** – Upstream; **5' UTR** – Five prime untranslated region; **3' UTR –** Three prime untranslated region; **Miss** – Missense; **Syn** – Synonym; **Non** – Nonsense; **FS** – Frameshift; **Region** – Genomic Region ; **PD** – Protein Domain; **TM** – Transmembrane; **SEA** – Sperm protein, Enterokinase and Agrin ; **CUB** – C1r/C1s, Urchin embryonic growth factor and BMP1 (1 and 2) ; **LDLR** – Low-density lipoproteins receptor (classes A1, A2 and A3); **SP** – Serine protease; **Clin_sig** – Clinical significance; **B** – Benign; **LB** – Likely benign; P – Pathogenic; **CI** –Conflicting ideas; **US** – Uncertain significance; **-** – Not applicable or not found.



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; 2010-2020

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.

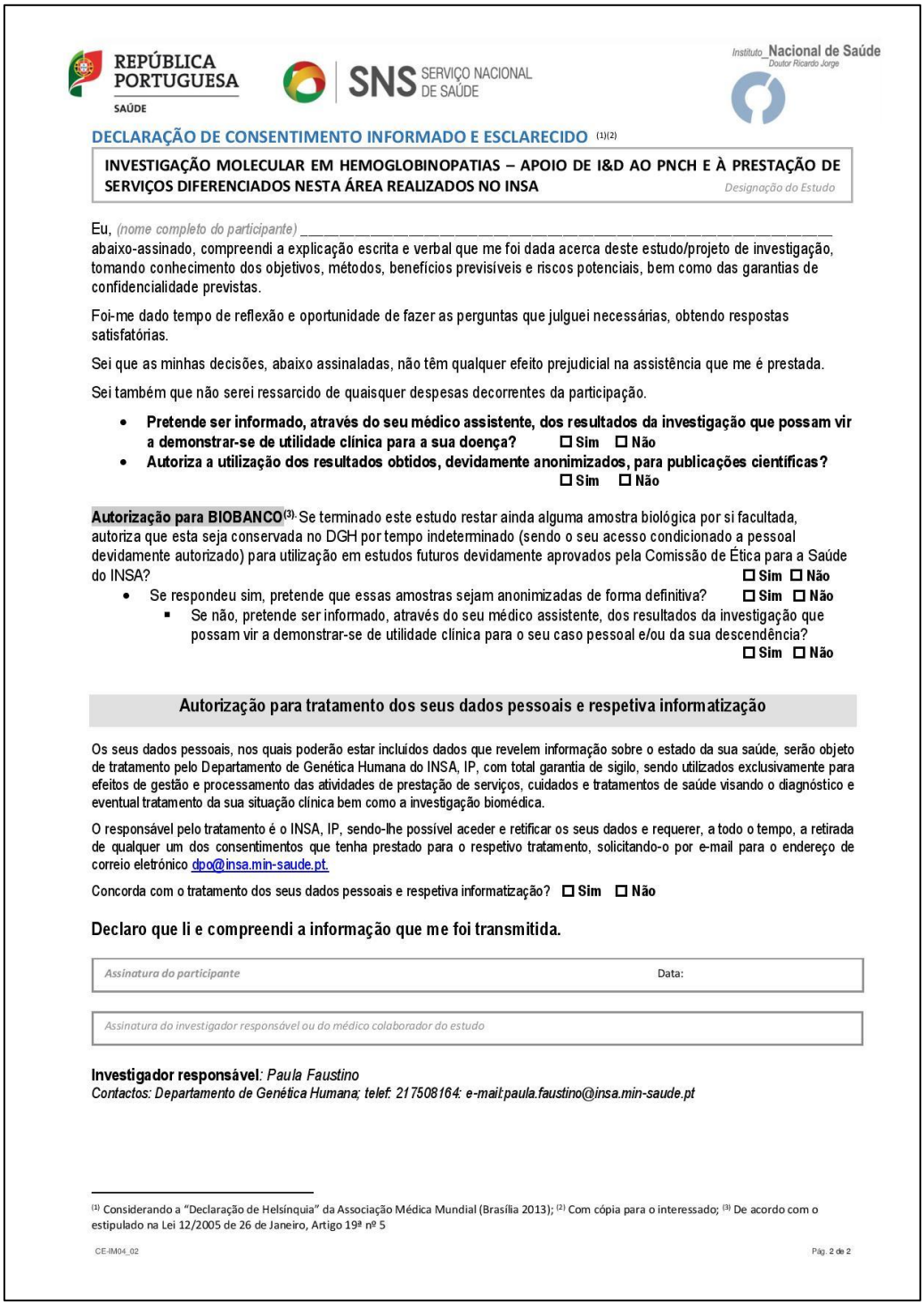
Para garantir a privacidade das pessoas participantes e a confidencialidade e proteção dos dados, as amostras de produtos biológicos e os eventuais dados demográficos e clínicos correspondentes serão pseudonimizados e tratados exclusivamente por pessoal sujeito a sigilo profissional.

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 2

Figure S.1 – Informed consent for molecular study of hemoglobinopathies (Part I).



REPÚBLICA PORTUGUESA
SAÚDE

SNS SERVIÇO NACIONAL DE SAÚDE

Instituto_Nacional de Saúde Doutor Ricardo Jorge

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**

Autorização para 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:

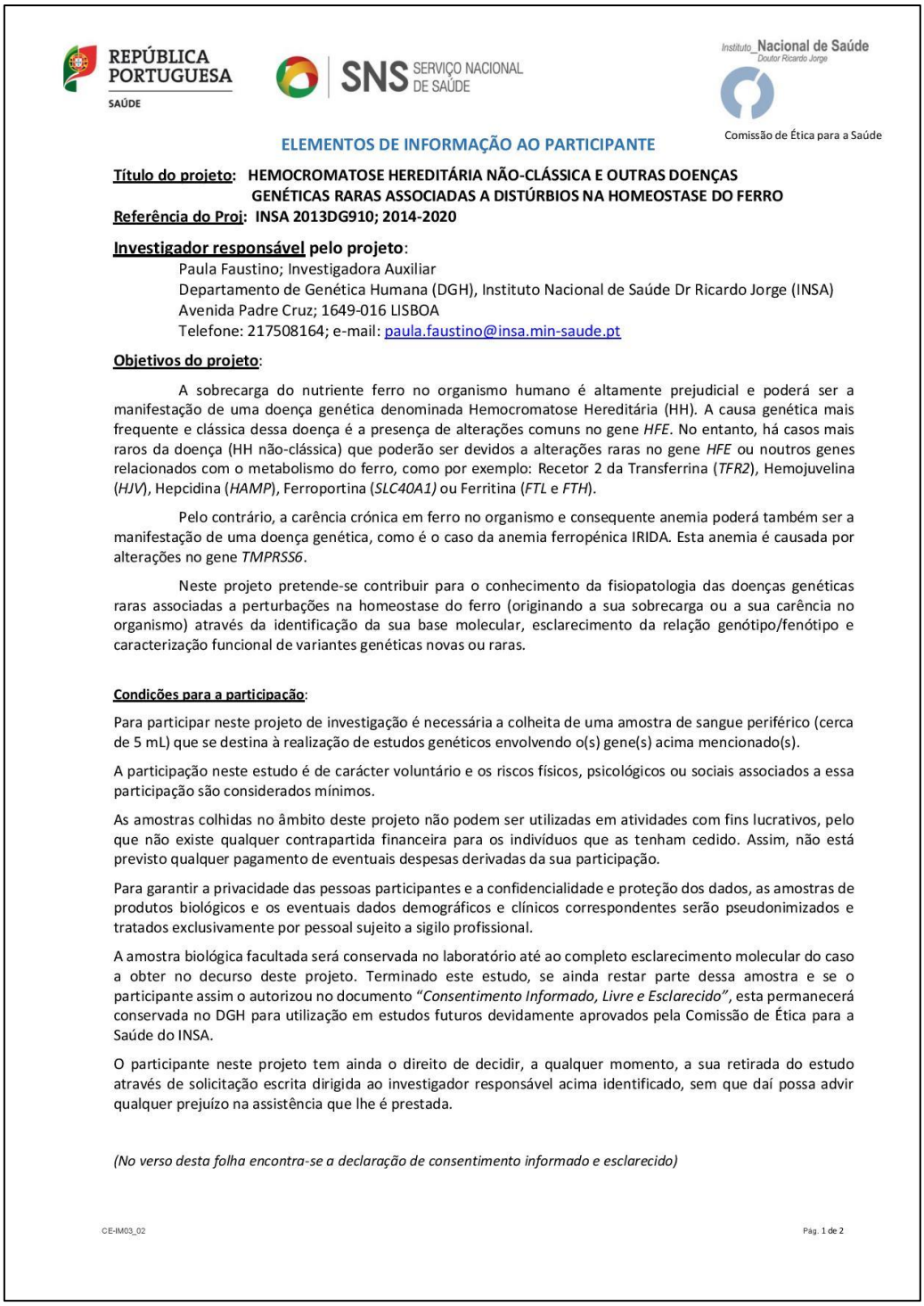
Assinatura do investigador responsável ou do médico colaborador do estudo

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 19ª nº 5

CE-IM04_02 Pág. 2 de 2

Figure S.2 – Informed consent for molecular study of hemoglobinopathies (Part II).



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: HEMOCROMATOSE HEREDITÁRIA NÃO-CLÁSSICA E OUTRAS DOENÇAS GENÉTICAS RARAS ASSOCIADAS A DISTÚRBIOS NA HOMEOSTASE DO FERRO

Referência do Proj: INSA 2013DG910; 2014-2020

Investigador responsável pelo projeto:

Paula Faustino; Investigadora Auxiliar
Departamento de Genética Humana (DGH), 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:

A sobrecarga do nutriente ferro no organismo humano é altamente prejudicial e poderá ser a manifestação de uma doença genética denominada Hemocromatose Hereditária (HH). A causa genética mais frequente e clássica dessa doença é a presença de alterações comuns no gene *HFE*. No entanto, há casos mais raros da doença (HH não-clássica) que poderão ser devidos a alterações raras no gene *HFE* ou noutros genes relacionados com o metabolismo do ferro, como por exemplo: Recetor 2 da Transferrina (*TFR2*), Hemojuvelina (*HJV*), Hepcidina (*HAMP*), Ferroportina (*SLC40A1)* ou Ferritina (*FTL* e *FTH*).

Pelo contrário, a carência crónica em ferro no organismo e consequente anemia poderá também ser a manifestação de uma doença genética, como é o caso da anemia ferropénica IRIDA. Esta anemia é causada por alterações no gene *TMPRSS6*.

Neste projeto pretende-se contribuir para o conhecimento da fisiopatologia das doenças genéticas raras associadas a perturbações na homeostase do ferro (originando a sua sobrecarga ou a sua carência no organismo) através da identificação da sua base molecular, esclarecimento da relação genótipo/fenótipo e caracterização funcional de variantes genéticas novas ou raras.

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) 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.

A amostra biológica facultada será conservada no laboratório até ao completo esclarecimento molecular do caso a obter no decurso deste projeto. Terminado este estudo, se ainda restar parte dessa amostra e se o participante assim o autorizou no documento "*Consentimento Informado, Livre e Esclarecido*", esta permanecerá conservada no DGH para utilização em estudos futuros devidamente aprovados pela Comissão de Ética para a Saúde do INSA.

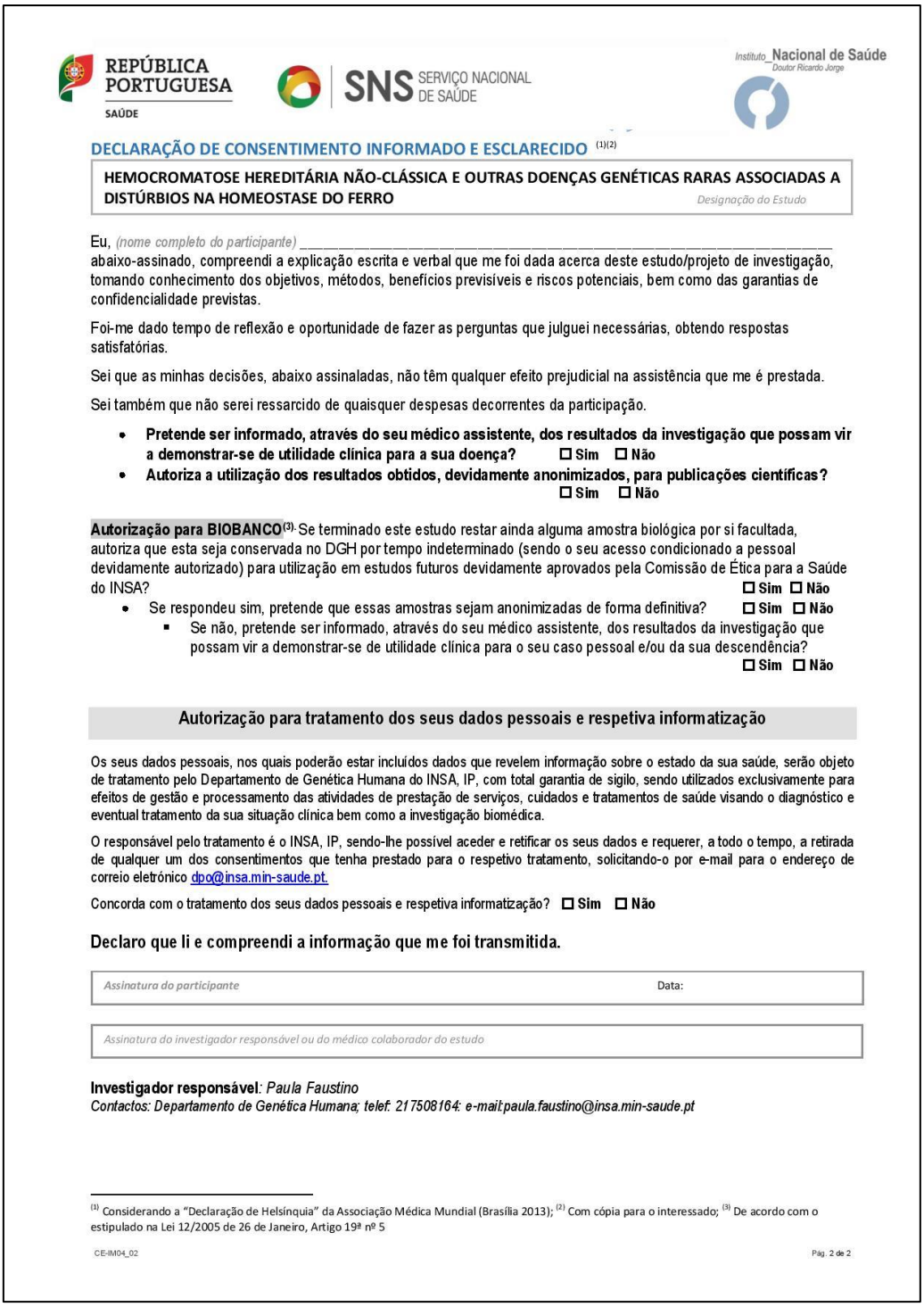
O participante neste projeto tem ainda o direito de decidir, a qualquer momento, a sua retirada do estudo através de solicitação escrita dirigida ao investigador responsável acima identificado, sem que daí possa advir qualquer prejuízo na assistência que lhe é prestada.

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

CE-IM03_02

Pág. 1 de 2

Figure S.3 – Informed consent for molecular study of hemochromatosis and other rare genetic diseases associated with disturbances in the iron homeostasis (Part I).



REPÚBLICA PORTUGUESA
SAÚDE

SNS SERVIÇO NACIONAL DE SAÚDE

Instituto Nacional de Saúde Doutor Ricardo Jorge

DECLARAÇÃO DE CONSENTIMENTO INFORMADO E ESCLARECIDO (1)(2)

HEMOCROMATOSE HEREDITÁRIA NÃO-CLÁSSICA E OUTRAS DOENÇAS GENÉTICAS RARAS ASSOCIADAS A DISTÚRBIOS NA HOMEOSTASE DO FERRO

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**

Autorização para 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 19ª nº 5

CE-IM04_02 Pág. 2 de 2

Figure S.4 – Informed consent for molecular study of hemochromatosis and other rare genetic diseases associated with disturbances in the iron homeostasis (Part II).

Table S.2 – Hematological and iron status parameters range for a normal adult, and indicators for Iron Deficiency Anemia, IRIDA, α-Thalassemia and β-Thalassemia Trait.

Parameters	Sex	Normal Adult	IDA	IRIDA	α-thal Trait	β-thal Trait
Red blood cells $x10^{12}$/L	M	4.32 – 5.66	↓	↓	↓	N
	F	3.88 – 4.99				
Hemoglobin (g/dL)	M	13.3 – 16.7	<13	↓	↓↓	<13
	F	11.8 – 14.8	<12	↓		<12
Hematocrit %	M	39.0 – 50.0	-	-	-	-
	F	36.0 – 44.0	-	-	-	-
Mean Cell Volume (fL)	M/F	82 – 98	<80	↓↓	<80	<80
Mean Cell Hemoglobin (pg)	M/F	27.3 – 32.6	<27	↓↓	<27	<27
Mean corpuscular hemoglobin concentration (g/dL)	M/F	31.6 – 34.9	<32	<32	-	N-↓
Red Cell Distribution Width (%)	M/F	9.9 – 15.5	↑↑	↑	N	N
Hemoglobin A2 (%)	M/F	2.2 – 3.5	N	N	N	>3,5
Serum Iron (µg/dL)	M	65 – 175	↓	↓↓	N	N-↑
	F	50 – 170				
Transferrin (mg/dL)	M	194 – 348	↓	↓↓	N	N
	F	181 – 416				
Total Binding Iron Capacity (µg/dL)	M/F	250 – 425	↓	↓↓	N	N
Transferrin Saturation (%)	M/F	16 – 45	<16	<10	N	N
Ferritin (µg/L)	M	20 – 250	<10	V	N	N-↑
	F	10 – 160				

Adapted from the literature [47], [147], [310] **IDA** – Iron Deficiency Anemia; **IRIDA** – Iron Refractory Iron Deficiency Anemia; **M** – Male; **F** – Female; **N** – Normal; **V** – Variable; ↓ – Low; ↓↓ – Very low; ↑ – High; ↑↑ –Very high

Table S.3 – Hematological and iron status parameters normal range for pediatric subjects.

Parameters	Sex	Age Range	Reference interval	Reference
Red blood cells $x10^{12}$/L	M/F	0 – 6 m	3.90 – 5.90	[311]
	M/F	6 m – 11 y	3.80 – 5.40	
	M	> 11 y	4.31 – 6.40	
	F	> 11 y	3.85 – 5.20	
Hemoglobin (g/dL)	M/F	0 – 6 m	14.0 – 18.0	[311]
	M/F	6 m –11 y	11.0 – 14.0	
	M	> 11 y	13.6 – 18.0	
	F	> 11 y	11.5 – 16.0	
Hematocrit %	M/F	0 – 2 w	42.0 – 68.0	[311]
	M/F	2 w – 2 m	35.0 – 50.0	
	M/F	2 m – 1 y	30.0 – 40.0	
	M/F	1 – 5 y	32.0 – 42.0	
	M	> 5 y	39.8 – 52.0	
	F	> 5 y	34.7 – 46.0	
Mean Cell Volume (fL)	M/F	0 – 2 w	88.0- 114.0	[311]
	M/F	2 w – 6 m	85.0 – 97.0	
	M/F	6 m – 11 y	72.0 – 86.6	
	M/F	> 11 y	80.0 – 97.0	
Mean Cell Hemoglobin (pg)	M/F	0 – 2 w	34.0 – 37.0	[311]
	M/F	2 w – 6 m	31.0 – 36.0	
	M/F	6 m – 5 y	25.0 – 31.0	
	M/F	> 5 y	26.0 – 34.0	
Mean corpuscular hemoglobin concentration (g/dL)	M/F	0 – 2 w	31.0 – 35.0	[311]
	M/F	2 w – 6 m	32.0 – 35.0	
	M/F	> 6 m	32.0 – 36.0	
Red Cell Distribution Width (%)	M/F	0 – 2 d	14.9 – 18.7	[311]
	M/F	> 2 d	11.5 – 15.0	
Hemoglobin A2 (%)	M/F	1 – 30 d	0.0 – 2.1	[312]
	M/F	1 – 2 m	0.0 – 2.6	
	M/F	3 – 5 m	1.3 – 3.1	
	M/F	≥ 6 m	2.0 – 3.3	
Serum Iron (µg/dL)	M/F	0-14 y	58.0 – 474.1	[313]
Transferrin (mg/dL)	M/F	0 – 9 w	1.04 –2.24	[313]
	M/F	9 w – 1 y	1.07 – 3.24	
	M/F	> 1 y	2.20 – 3.37	

Continued

Parameters	Sex	Age Range	Reference interval	Reference
Total Binding Iron Capacity (µg/dL)	M	0 – 90 d	155 – 330	[314]
	F	0 – 90 d	165 – 275	
	M	91 d – 12 m	150 – 380	
	F	91 d – 12 m	250 – 455	
	M	13 m – 3 y	215 – 420	
	F	13 m – 3 y	160 – 415	
	M	4 – 10 y	185 – 415	
	F	4 – 10 y	260 – 385	
	M	11 – 14 y	265 – 410	
	F	11 – 14 y	250 – 420	
Transferrin Saturation (%)	M/F	6 m – 11 y	6 – 41	[315]
	M/F	12 – 18 y	6 – 48	
Ferritin (µg/L)	M/F	0 – 15 d	39.8 – 539.9	[316]
	M/F	15 d – 6 m	15.3 – 374.6	
	M/F	6 m – 1 y	13.3 – 191.9	
	M/F	1 – 16 y	10.3 – 55.8	

M – Male; **F** – Female; **d** – Days; **w** – Weeks; **m** – Months; **y** – Years.

Table S.4 – Conventional PCR conditions for the amplification of the fragments of interest in the globin genes.

Gene	Region	Primers ID	Length (nt)	Sequence	CG %	Tm (˚C)	Master mix			PCR conditions			Fragment size (bp)
							Reagents	[conc.]	Vol (µL)	T (ºC)	Δt		
HBB	Promoter - Exon 2	L5'B (Fw)	19	5'-TAAGCCAGTGCCAGAAGAG-3'	52.6	56.8	ddH_2O	-	20.05	94	5'	x30	755
							Buffer David	10x	2.50	94	45"		
							BSA	10 mg/mL	0.35	60	45"		
							dNTPs	100 mM	0.50	72	45"		
							Primer L5'B (Fw)	25 µM	0.50	72	5'		
		R5'B (Rv)	20	5'-TCCCATTCTAAACTGTACCC-3'	45.0	53.3	Primer R5'B (Rv)	25 µM	0.50	4	pause		
							Go*Taq*®	5 U/µL	0.10				
							DNA	~80 ng/µL	0.50				
	Intron 2	PL38 (Fw)	19	5'-TGATGGCCTGGCTCACCTG-3'	63.2	61.0	ddH_2O	-	20.05	94	5'	x30	912
							Buffer David	10x	2.50	94	1'		
							BSA	10 mg/mL	0.35	67	1'		
							dNTPs	100 mM	0.50	72	1'30"		
							Primer PL38 (Fw)	25 µM	0.50	72	6'		
		PL39 (Rv)	19	5'-CCAAGCTAGGCCCTTTTGC-3'	57.9	58.8	Primer PL39 (Rv)	25 µM	0.50	4	pause		
							Go*Taq*®	5 U/µL	0.10				
							DNA	~80 ng/µL	0.50				
	Exon 3 - 3'UTR	15 (Fw)	24	5'- CAATGTATCATGCCTCTTTGCACC-3'	45.8	61.0	ddH_2O	-	20.05	94	5'	x30	862
							Buffer David	10x	2.50	94	45"		
							BSA	10 mg/mL	0.35	65	45"		
							dNTPs	100 mM	0.50	72	45"		
							Primer 15 (Fw)	25 µM	0.50	72	5'		
		14 (Rv)	24	5'-GAGTCAAGGCTGAGAGATGCAGGA-3'	54.2	64.4	Primer 14 (Rv)	25 µM	0.50	4	pause		
							Go*Taq*®	5 U/µL	0.10				
							DNA	~80 ng/µL	0.50				
HBD	Promoter - Exon 2	R117 (Fw)	18	5'-GGGCAAGTTAAGGGAATA-3'	30.8	56.9	ddH_2O	-	20.00	94	5'	x30	718
							Buffer David	10x	2.50	94	1'		
							BSA	10 mg/mL	0.35	57	1'		
							dNTPs	100 mM	0.50	72	1'30"		
							Primer R117 (Fw)	25 µM	0.50	72	6'		
		R119 (Rv)	19	5'-GGAGAAGAGCAGGTAGGT-3'	55.6	56.0	Primer R119 (Rv)	25 µM	0.50	4	pause		
							Ampli*Taq*®	5 U/µL	0.15				
							DNA	~80 ng/µL	0.50				

Continued

Gene	Region	Primers ID	Length (nt)	Sequence	CG %	Tm (˚C)	Master mix			PCR conditions			Fragment size (bp)
							Reagents	[conc.]	Vol (µL)	T (ºC)	Δt		
HBD	Exon 3-3'UTR	R109 (Fw)	19	5'-GAATAACCTGGGGATCAGT-3'	50.0	55.6	ddH_2O	-	20.00	94	5'	x30	612
							Buffer David	10x	2.50	94	1'		
							BSA	10 mg/mL	0.35	62	1'		
							dNTPs	100 mM	0.50	72	1'30"		
							Primer R109 (Fw)	25 µM	0.50	72	6'		
		R89 (Rv)	20	5'-TCCCTAGAACCTCTGCAGTG-3'	55.0	59.3	Primer R89 (Rv)	25 µM	0.50	4	pause		
							Ampli*Taq*®	5 U/µL	0.15				
							DNA	~80 ng/µL	0.50				
HBG2 (Gγ)	Promoter	R160 (Fw)	25	5'-CACTGAAACTGTTGCTTTATAGGAT-3'	30.8	56.9	ddH_2O	-	20.00	94	5'	x30	677
							Buffer David	10x	2.50	94	1'		
							BSA	10 mg/mL	0.35	55	1'		
							dNTPs	100 mM	0.50	72	1'		
							Primer R160 (Fw)	25 µM	0.50	72	10'		
		R161 (Rv)	24	5'-TGGCGTCTGGACTAGGAGCTTATT-3'	50.0	62.7	Primer R161 (Rv)	25 µM	0.50	4	pause		
							Ampli*Taq*®	5 U/µL	0.15				
							DNA	~80 ng/µL	0.50				
HBG1 (Aγ)	Promoter	Aγ159 (Fw)	26	5'-TGAAACTGTGGCTTTATAGAAATTGT-3'	30.8	56.9	ddH_2O	-	20.00	94	5'	x30	676
							Buffer David	10x	2.50	94	1'		
							BSA	10 mg/mL	0.35	62	1'		
							dNTPs	100 mM	0.50	72	1'		
							Primer Aγ159 (Fw)	25 µM	0.50	72	10'		
		R161 (Rv)	24	5'-TGGCGTCTGGACTAGGAGCTTATT-3'	50.0	62.7	Primer R161 (Rv)	25 µM	0.50	4	pause		
							Ampli*Taq*®	5 U/µL	0.15				
							DNA	~80 ng/µL	0.50				
	Exon 3	Aγ3 (Fw)	23	5'-GCCAGTGACTAGTGCTTGAAGGG-3'	56.5	64.2	ddH_2O	-	20.00	94	5'	x26	504
							Buffer David	10x	2.50	94	1'		
							BSA	10 mg/mL	0.35	65	1'		
							dNTPs	100 mM	0.50	72	1'		
							Primer Aγ3 (Fw)	25 µM	0.50	72	10'		
		Aγ4 (Rv)	23	5'-CAGGCATGCAGAAATACACATAC-3'	43.5	58.9	Primer Aγ4 (Rv)	25 µM	0.50	4	pause		
							Ampli*Taq*®	5 U/µL	0.15				
							DNA	~80 ng/µL	0.50				

Continued

Gene	Region	Primers ID	Length (nt)	Sequence	CG %	Tm (˚C)	Master mix			PCR conditions			Fragment size (bp)
							Reagents	[conc.]	Vol (µL)	T (ºC)	Δt		
HBA2	5'UTR - 3'UTR	A_5'UTR_ F1 (Fw)	20	5'- GGACTCCCCTGCGGTCCAGG-3'	55.0	59.3	ddH_2O	-	16.30	94	10'	x32	600
							Buffer α + βME	10x	2.50	94	1'		
							DMSO	10%	2.50	61	1'		
							BSA	10 mg/mL	0.40	72	1'		
							$MgCl_2$	0.1 M	0.50	72	10'		
							dNTPs	100 mM	0.50	4	pause		
							Primer A_F1 (Fw)	25 µM	0.50				
		A2_3'UTR _R1 (B_Rv)	23	5'- CTCCATTGTTGGCACATTCCGGG-3'	56.5	64.2	Primer B (Rv)	25 µM	0.50				
							Ampli*Taq**	5 U/µL	0.30				
							DNA	~80 ng/µL	1.00				

* The *Taq* DNA polymerases used were Go*Taq* (Promega, USA) and AmpliTaq (Applied Biosystems, USA). Primers melting temperature and CG% were calculated in Oligo Analysis Tool (Eurofins Genomics, Luxembourg; https://www.eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/). **5' UTR** – Five prime untranslated region; **3' UTR –** Three prime untranslated region; **Fw** – Forward; **Rv** – Reverse; **nt** –Nucleotide; **Tm** – Melting temperature; **ddH2O** – deionized and distilled water; **BSA** – Bovine serum albumin; **dNTPs** – Deoxynucleosides triphosphate; **βME** – 2-Mercaptoethanol; **DMSO** – Dimethyl sulfoxide; **MgCl2** – Magnesium chloride; **[conc.]** – Concentration; **T –** Temperature; **Δt –** Time variation; **bp** – base pairs.

Table S.5 – Gap-PCR conditions for detection of the alpha-thalassemia deletion -α$^{3.7kb}$.

Gene	Region	Primers ID	Length (nt)	Sequence	CG %	Tm (˚C)	Master mix			PCR conditions			Fragment size (bp)
								Reagents	[conc.]	Vol (µL)	T (ºC)	Δt	
HBA2/ HBA1	-α$^{3.7}$ deletion	A_5'UTR_F2 (A_Fw)	20	5'- GGGATGCACCCACTGGCACT-3'	65.0	63.5	Reaction 1	ddH2O	-	15.30	94	5'	1900
								Buffer α + βME	10x	2.50	94	1'	
								DMSO	10%	2.50	64	1'	
								BSA	10 mg/mL	0.40	72	1'	
								MgCl2	0.1 M	0.50	72	10'	
								dNTPs	100 mM	0.50	4	pause	
								Primer A_F2 (Fw)	25 µM	0.50			
								Primer B (Rv)	25 µM	0.50			
								Ampli*Taq**	5 U/µL	0.30			
								DNA	~40 ng/µL	2.00			
		A2_3'UTR_R1 (B_Rv)	23	5'- CTCCATTGTTGGCACATTCCGGG-3'	56.5	64.2						x34	
							Reaction 2	ddH2O	-	15.30			2100
								Buffer α + βME	10x	2.50			
								DMSO	10%	2.50			
								BSA	10 mg/mL	0.40			
								MgCl2	0.1 M	0.50			
								dNTPs	100 mM	0.50			
		A1_3'UTR_R1 (C_Rv)	20	5'-CTGCTGTCCACGCCCATGCC-3'	70.0	65.5		Primer A_F2 (Fw)	25 µM	0.50			
								Primer C (Rv)	25 µM	0.50			
								Ampli*Taq**	5 U/µL	0.30			
								DNA	~40 ng/µL	2.00			

* The *Taq* DNA polymerase used was AmpliTaq (Applied Biosystems, USA). Primers melting temperature and CG% were calculated in Oligo Analysis Tool (Eurofins Genomics, Luxembourg; https://www.eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/). **5' UTR** – Five prime untranslated region; **3' UTR –** Three prime untranslated region; **Fw** – Forward; **Rv** – Reverse; **nt** – Nucleotide; **Tm** – Melting temperature; **ddH2O** – deionized and distilled water; **βME** – 2-Mercaptoethanol; **DMSO** – Dimethyl sulfoxide; **BSA** – Bovine serum albumin; **MgCl2** – Magnesium chloride; **dNTPs** – Deoxynucleosides triphosphate; **[conc.]** – Concentration; **T –** Temperature; **Δt –** Time variation; **bp** – base pairs.

Table S.6 – Long-PCR conditions for the amplification of the gene *TMPRSS6.*

Gene	Region	Primers ID	Length (nt)	Sequence	CG %	Tm (˚C)	Master mix			PCR conditions			Fragment size (bp)
							Reagents	[conc.]	Vol (µL)	T (ºC)	Δt		
TMPRSS6	Exon 1- Exon 6	Ex1_Fw (Fw)	21	5'-CTGAGACCTCCGTCTGTCCTC-3'	61.9	63.7	ddH2O	-	15.90				
							BSA	10 mg/mL	0.35	98	4'		
							10xLA PCR Buffer II	25 mM	2.50	98	30"		
							dNTPs	100 mM	4.00	63	30"	x30	8266
							Primer Ex1_Fw	25 µM	0.50	72	8'		
		Ex6_Rv (Rv)	21	5'-CCCTGCACACACAACAGAAGC-3'	57.1	61.8	Primer Ex6_Rv	25 µM	0.50	72	10'		
							Takara LA *Taq* HS*	1.25 U/µL	0.25	4	pause		
							DNA	30 ng/µL	1.00				
	Exon 7 - Exon 10	Ex7_Fw (Fw)	21	5'-AGGCGTGAAGCTCAGTGTGTG- 3'	57.1	61.8	ddH2O	-	15.90				
							BSA	10 mg/mL	0.35	98	4'		
							10xLA PCR Buffer II	25 mM	2.50	98	30"		
							dNTPs	100 mM	4.00	64	30"	x30	5831
							Primer Ex1_Fw	25 µM	0.50	72	6'		
		Ex10_Rv (Rv)	21	5'-AGGCGTGAAGCTCAGTGTGTG- 3'	57.1	61.8	Primer Ex6_Rv	25 µM	0.50	72	10'		
							Takara LA *Taq* HS*	1.25 U/µL	0.25	4	pause		
							DNA	30 ng/µL	1.00				
	Exon 11 - Exon 18	Ex11_Fw (Fw)	21	5'- AGGGAGAAATCAGGGCAGAGG-3'	57.1	61.8	ddH2O	-	15.90				
							BSA	10 mg/mL	0.35	94	4'		
							10xLA PCR Buffer II	25 mM	2.50	94	30"		
							dNTPs	100 mM	4.00	65	30"	x30	9421
							Primer Ex1_Fw	25 µM	0.50	72	10'		
		Ex18_Rv (Rv)	21	5'-GAATACTTGTCCCCCTGCTTG-3'	52.4	59.8	Primer Ex6_Rv	25 µM	0.50	72	10'		
							Takara LA *Taq* HS*	1.25 U/µL	0.25	4	pause		
							DNA	20 ng/µL	1.00				

* The *Taq* DNA polymerase used was Takara LA *Taq* Hot Start version (Takara Bio, Japan) as well as the respective PCR buffer. Primers melting temperature and CG% were calculated in Oligo Analysis Tool (Eurofins Genomics, Luxembourg; https://www.eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/). **Fw** – Forward; **Rv** – Reverse; **nt** –Nucleotide; **Tm** – Melting temperature; **ddH2O** – deionized and distilled water; **BSA** – Bovine serum albumin; **dNTPs** – Deoxynucleosides triphosphate; **[conc.]** – Concentration; **T –** Temperature; **Δt –** Time variation; **bp** – base pairs.

Table S.7 – ARMS-PCR conditions for detection of the variant rs855791 (c.2207T>C - V736A) in the *TMPRSS6* gene.

Gene	Region/ Variant	Primers ID	Length (nt)	Sequence	CG%	Tm (˚C)	Master mix			PCR conditions			Fragment size (bp)
							Reagents	[conc.]	Volume (µL)	T (ºC)	Δt		
TMPRSS6	Exon 2	Ex2_Fw	21	5'-TGCCGCCTGATGTTGTTACTC-3'	52.4	59.8	Reaction 1: Buffer β	-	22.90	94	5'		395
							Ex2_Fw	25 µM	0.25	94	20"		
		Ex2_Rv	21	5'-GCCTGCTACAGTCACCCCAAG-3'	61.9	63.7	Ex2_Rv	25 µM	0.25	63	20"	x32	
							Primer Fw_T	25 µM	0.25	72	20"		
							Primer Ex17_Rv	25 µM	0.25	72	3'		
							Go*Taq**	5 U/µL	0.10	4	pause		
							DNA	~80 ng/µL	0.50				
	rs855791 V736A (T>C)	V736A_Fw_T	22	5'-CACAGGACCTGTGCAGCAAGG**T**-3'	59.1	64.0	Reaction 2: Buffer β	-	0.30				175
							Ex2_Fw	25 µM	2.00				
		V736A_Fw_C	22	5'-CACAGGACCTGTGCAGC**G**AGG**C**-3'	68.2	67.7	Ex2_Rv	25 µM	15.30				
							Primer Fw_C	25 µM	2.50				
							Primer Ex17_Rv	25 µM	2.50				
							Go*Taq**	5 U/µL	0.40				
		Ex17_Rv	21	5'-GATGTGAGCAAAGGGCCAGAC-3'	57.1	61.8	DNA	~80 ng/µL	0.50				

* The *Taq* DNA polymerase used was Go*Taq* (Promega, USA). Primers melting temperature and CG% were calculated in Oligo Analysis Tool (Eurofins Genomics, Luxembourg; https://www.eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/). **Fw** – Forward; **Rv** – Reverse; **nt** –Nucleotide; **Tm** – Melting temperature; **[conc.]** – Concentration; **T –** Temperature; **Δt –** Time variation; **bp** – base pairs.

Table S.8 – ARMS-PCR conditions for detection of the variants rs1799945 (c.187C>G - H63D) and rs1800562 (c.845G>A - C282Y) in the *HFE* gene.

Gene	Variant	Primers ID	Length (nt)	Sequence	CG%	Tm (˚C)		Master mix			PCR conditions			Fragment size (bp)
								Reagents	[conc.]	Vol. (µL)	T (ºC)	Δt		
HFE	rs1799945 H63D (C>G)	ARMS_ H63D_C_Fw	22	5'-ACATGGTTAACCGGCCTGTTGC-3'	54.5	62.1	Reaction 1	ddH2O	-	18.50	94 94 59 72 72 4	5' 30" 30" 30" 5' pause	x30	395
		ARMS_ H63D_N_Rv	30	5'-AGTTCGGGGCTCCACACG GCGACTCTCAA**G-3'**	63,3	73.6		Buffer David	10x	2.50				
		ARMS_ H63D_M_Rv	30	5'-AGTTCGGGGCTCCACACG GCGACTCTCAA**C**-3'	63.3	73.6		BSA	10 mg/mL	0.35				
								dNTPs	100 mM	0.50				
								Primer H63D_C_Fw	25 µM	0.50				
								Primer H63D_M_Rv	25 µM	0.50				
								Primer C282Y_C_Fw	25 µM	0.50				
								Primer C282Y_N_Rv	25 µM	0.50				
								Go*Taq**	5 U/µL	0.15				
								DNA	~80 ng/µL	1.00				
	rs1800562 C282Y (G>A)	ARMS_ C282Y_C_Fw	20	5'-TGGCAAGGGTAAACAGATCC-3'	50.0	57.3	Reaction 2	ddH2O	-	18.50				175
		ARMS_ C282Y_N_Rv	30	5'-GCTGATCCAGGCCTGGGTGC TCCACCTGC**C**-3'	70.0	76.3		Buffer David	10x	2.50				
		ARMS_C282 Y_M_Rv	30	5'-GCTGATCCAGGCCTGGGTG CTCCACCTGC**T**-3'	66.7	75.0		BSA	10 mg/mL	0.35				
								dNTPs	100 mM	0.50				
								Primer H63D_C_Fw	25 µM	0.50				
								Primer H63D_N_Rv	25 µM	0.50				
								Primer C282Y_C_Fw	25 µM	0.50				
								Primer C282Y_M_Rv	25 µM	0.50				
								Go*Taq**	5 U/µL	0.15				
								DNA	~80 ng/µL	1.00				

***** The *Taq* DNA polymerase used was Go*Taq* (Promega, USA). Primers melting temperature and CG% were calculated in Oligo Analysis Tool (Eurofins Genomics, Luxembourg; https://www.eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/). **Fw** – Forward; **Rv** – Reverse; **nt** –Nucleotide; **Tm** – Melting temperature; **ddH2O** – deionized and distilled water; **BSA** – Bovine serum albumin; **dNTPs** – Deoxynucleosides triphosphate; **[conc.]** – Concentration; **T –** Temperature; **Δt –** Time variation; **bp** – base pairs.

Table S.9 – Composition of buffer solutions used throughout the project.

PCR buffers		
Buffer β	**Buffer David (10x)**	**Buffer α (10x)**
KCl_2 50 mM Tris-HCl (pH = 8.8) 100 mM $MgCl_2$ 15 mM Gelatin – 0.01% (w/v)	$(NH_4)_2SO_4$ 166 mM Tris-HCl (pH = 8.8) 670 mM $MgCl_2$ 67 mM EDTA 0.067 mM β-Mercaptoetanol 100 mM	$(NH_4)_2SO_4$ 166 mM Tris-HCl (pH = 8.8) 670 mM $MgCl_2$ 15 mM EDTA 0.67 mM β-Mercaptoetanol 100 mM
Electrophoresis loading buffer	Water 8.75 µL Bromophenol Blue 0.22 g Glycerol 18.75 µL EDTA 75 µL NaOH (to turn blue)	

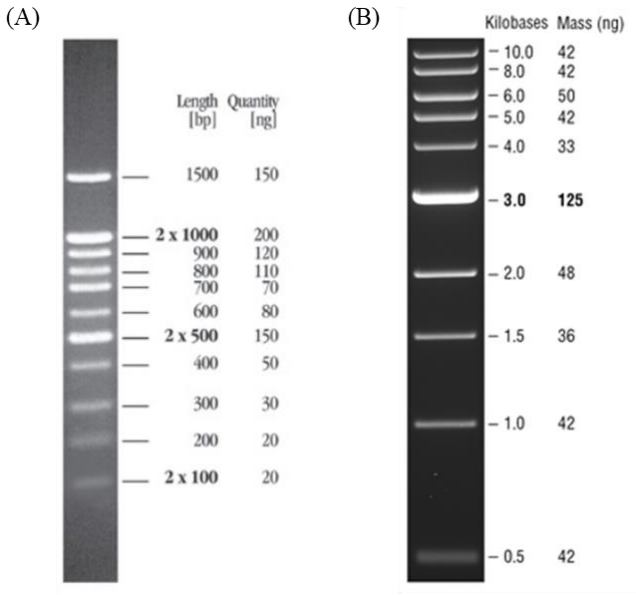


Figure S.5 – DNA ladders used in agarose gel electrophoresis. **(A)** DNA Ladder (L) 100 bp plus (AppliChem, USA) **(B)** 1Kb DNA Ladder (New England BioLabs, USA)

Table S.10 – Conventional PCR conditions for amplification of *TMPRSS6* gene's Exon 7.

Gene	Region	Primers ID	Length (nt)	Sequence	CG%	Tm (˚C)	Master mix			PCR conditions			Fragment size (bp)
							Reagents	[conc.]	Volume (µL)	T (ºC)	Δt		
TMPRSS6	Exon 7	Ex7_Fw	21	5'-AGGCGTGAAGCTCAGTGTGTG- 3'	57.1	61.8	Buffer β	-	22.90	94	5'	x32	584
							Primer Ex7_Fw	25 µM	0.25	94	20"		
							Primer Ex7_Rv	25 µM	0.25	63	20"		
							Taq DNA Polymerase*	5 U/µL	0.10	72	20"		
		Ex7_Rv	21	5'-CTAGCCGTCCTGTCTCCCAGA- 3'	61.9	63.7	DNA	~80 ng/µL	1.00	72	3"		
										4	pause		

* The *Taq* DNA polymerase use was *Taq* DNA Polymerase (Fermentas, USA). Primers melting temperature and CG% were calculated in Oligo Analysis Tool (Eurofins Genomics, Luxembourg; https://www.eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/). **Fw** – Forward; **Rv** – Reverse; **nt** –Nucleotide; **Tm** – Melting temperature; **[conc.]** – Concentration; **T –** Temperature; **Δt –** Time variation; **bp** – base pairs.

Table S.11 – Reaction mixture for RFLP detection of the variant rs2235324 (c.757A>G - K253E) in gene *TMPRSS6.*

Gene	Variant	Reaction Mix			Reaction conditions		Recognition sequence	Restriction product size (bp)	
		Reagents	[conc.]	Volume (µL)	T(ºC)	Δt		(-)	(+)
TMPRSS6	rs2235324 K253E (A>G)	ddH2O *Ava* I (NEB®) PCR product (Exon7)	- 10 U/µL -	11 1 8	37	3-16 h	5′...C▼Y C G R G...3′ 3′...G R G C Y▲C...5′	A – Wild-type 584	G - Mutation 336 + 248

* The restriction enzyme used was *Ava* I (New England BioLabs, USA). **ddH2O** – deionized and distilled water; **[conc.]** – Concentration; **T –** Temperature; **Δt –** Time variation; **bp** – base pairs; **Y** – pYrimidine; **R** – puRine.

Table S.12 – β-globin gene cluster MLPA probes arranged according to chromosomal location.

Length (nt)	SALSA MLPA probe	Gene/ Exon	Partial sequence (24 nt adjacent to ligation site)	Distance to next probe
Centromeric flanking probes				
253	18247-SP0630-L27000	OR51M1 gene	TGCTGGACCTGG-30 nt spanning oligo-TGCACACAGTAG	66.1 kb
298	18249-L24239	OR51B2 gene	TAGCAGTGTGGG-CTGCTTCCTACA	28.9 kb
HS5/ HS4/ HS3/HS2/ HS1				
261	05804-L27001	Upstream HBB-HS5	TCCATGAAGTAT-TACAGCATTTGG	6.8 kb
304	05806-L22542	HBB-HS4 region	CACTCAGCAGCT-ATGAGATGGCTT	0.4 kb
268	05807-L22540	HBB-HS4 region	CATTTCCATGTC-ATACTGAGAAAG	2.1 kb
328	05808-L05312	HBB-HS3 region	TTCCTGCCCATT-CAGGGCTCCAGC	1.1 kb
229	05810-L05314	HBB-HS3 region	TGGCCACCAGCT-ATCAGGGCCCAG	4.1 kb
292	06395-L05315	HBB-HS2 region	GTGCCCAGATGT-TCTCAGCCTAGA	4.8 kb
238	12189-L14345	HBB-HS1 region	TCCTGAGCTCTT-ATCTATATCCAC	6.0 kb
HBE1				
463	05813-L05317	HBE1 Exon 1	GCAGCAGCACAT-ATCTGCTTCCGA	13.8 kb
HBG2				
373	05815-L05319	Upstream HBG2	TGAGCAGATATA-AGCCTTACACAG	2.8 kb
427	14267-L15911	HBG2 Exon 3	GATGGTGACTGG-AGTGGCCAGTGC	0.2 kb
472	18109-SP0127-L27005	HBG2 Exon 3	CACATAAACACA-45 nt spanning oligo-TAAAAAAAGAAC	1.9 kb
HBG1				
409	05817-L05321	Upstream HBG1	CAATGAGCAGAT-ATAAGCTTTACA	2.5 kb
180	18245-L22954	HBG1 Intron 2	GACTAGTGCTTG-AAGGGGAACAAC	0.3 kb
436	14267-L27004	HBG1 Exon 3	GATGGTGACTGC-AGTGGCCAGTGC	2.8 kb
***HBBP1* (pseudogene)**				
283	18248-SP0631-L27002	Upstream HBBP1	GAAATGGGGAAC-33 nt spanning oligo-AAAGTGACTGCA	2.1 kb
445	06400-L05323	HBBP1 Exon 1	TGAAGCAAGGTT-AAGGTGAGAAGG	1.5 kb
382	05820-L05324	HBBP1 Exon 3	ATGGGGGAGGTT-GGGGAGAAGAGC	4.1 kb
HBD				
420	05821-L06327	Upstream HBD	CTCATCCTCCTT-ACATACATTTCC	3.4 kb
310	11886-L12686	HBD Exon 1	TGAACACTGTTA-TGTCAGAAGAAA	1.4 kb
399	06397-L05327	HBD Exon 3	ATTCACCCCACA-AATGCAGGCTGC	0.2 kb
391	18108-SP0126-L15076	HBD Exon 3	TTATTAGGCAGA-41 nt spanning oligo-AAATAGAATCTA	3.0 kb
HBB				
337	05824-L05328	Upstream HBB	GATTTTCATGGA-GGAAGTTAATAT	1.9 kb
365	11982-L24242	Upstream HBB	TACCCCTACTTT-CTAAGTCACAGA	0.8 kb
148	05827-L06319	HBB Promoter	ACAGGTACGGCT-GTCATCACTTAG	0.1 kb
189	05828-L05332	Exon 1	GTCAGAAGCAAA-TGTAAGCAATAG	0.1 kb
219	05830-L23703	HBB: c.20A>T wildtype-specific	TCTGACTCCTG**A**-GGAGAAGTCTGC	-
214	05830-L05307	HBB: c.20A>T mutation-specific	TCTGACTCCTG**T**-GGAGAAGTCTGC	0.1 kb
154	11883-L12683	Intron 1	AGGAGACCAATA-GAAACTGGGCAT	0.4 kb
196	05833-L05335	Intron 2	GGAAACAGACGA-ATGATTGCATCA	0.9 kb
166	13619-L15073	Exon 3	ATCCCCCAGTTT-AGTAGTTGGACT	0.3 kb
208	11885-L25666	0.2 kb after Exon 3	GAAAAGGATTCA-AGTAGAGGCTTG	0.3 kb
173	05836-L06321	0.5 kb after Exon 3	GTGAGCCCTTCT-TCCCTGCCTCCC	9.1 kb
Telomeric flanking probes				
274	11980-L12803	Downstream HBB	AGAGTTGAGCAA-GGCCTGAAATTT	15.8 kb
486	18253-L22962	OR51V1 gene	GAAAAGGATGAG-TATAGCATCCAA	48.3 kb
355	18251-L27003	OR52A1 gene	ATTGTGAAACTA-GCAGCAGCAAAT	45.3 kb
223	18246-L23704	Downstream OR52A5 gene	GGCACTTCCTTA-AGAGCACAGGAA	504.3 kb
160	18244-L22953	TRIM68 gene	AAAGAGAGGTCG-CAGAGGCCTGTC	-

nt – Nucleotide; **kb** – Kilobase

Table S.13 – MLPA assay conditions for detection copy number variations in β-globin gene cluster.

Reaction Mix			Reaction conditions		
Reagents	**Volume (µL)**	**Vol. Mix per tube**	**Steps**	**T(ºC)**	**Δt**
RNase	-	1 µL	RNase digestion	37	30''
MIX I (Hybridization Master Mix) *			DNA denaturation	98	5'
SALSA Probe mix **(black)**	1.5	3 µL		25	Pause*
MLPA Buffer **(yellow)**	1.5		Hybridisation of probes to sample DNA	95	1''
MIX II (Ligase Master Mix) **				60	16 h
ddH2O	25	32 µL	Ligation of hybridized probes	54	pause**
Ligase-65 Buffer A *(transparente)*	3.0			54	15'
Ligase-65 Buffer B **(white)**	3.0			98	5'
Ligase-65 **(green)**	1.0			20	pause***
MIX III (Polymerase Master Mix) ***			PCR amplification of ligated probes	95	30''
ddH2O		10 µL		60	30''
PCR Primer mix **(brown)**	7.5			72	60''
SALSA polymerase **(orange)**				72	20'
				15	pause

(35x)

* Add MIX I; ** Add MIX II; *** Add MIX III; Reagents are all from SALSA® MLPA® probemix P102-C1 HBB kit (MCR Holland, Netherlands). **ddH2O** – deionized and distilled water; **T –** Temperature; **Δt –** Time variation.

Table S.14 – Conventional PCR conditions for confirmation of Cape Verde deletion (~7.7 kb) in β-globin gene cluster.

Gene	Region	Primers ID	Length (nt)	Sequence	CG%	Tm (˚C)	Master mix			PCR conditions			Fragment size (bp)
							Reagents	**[conc.]**	**Volume (µL)**	**T (ºC)**	**Δt**		
HBB	~7.7 deletion *HBB* (Intron 2) - Downstream β-globin gene cluster	Primer CV_Fw	25	5-AACAGACGAATGATTGCATCAGTGT-3	40.0	59.7	ddH2O	-	20.00	94	5'		419
							Buffer David	10x	2.50	94	30"	x32	
							BSA	10 mg/mL	0.35	59	30"		
							dNTPs	100 mM	0.50	72	40"		
							Primer CV_Fw	25 µM	0.50	72	5'		
		Primer CV_Rv	28	5-TAATCATATGGTTTGTGTCTTTGGTTCT-3	40.0	59.7	Primer CV_Rv	25 µM	0.50	4	pause		
							Amplitaq*	5U/µL	0.15				
							DNA	~80 ng/µL	0.50				

* The *Taq* DNA polymerases used was AmpliTaq (Applied Biosystems, USA). Primers melting temperature and CG% were calculated in Oligo Analysis Tool (Eurofins Genomics, Luxembourg; https://www.eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/). **Fw** – Forward; **Rv** – Reverse; **nt** –Nucleotide; **Tm** – Melting temperature; **ddH2O** – deionized and distilled water; **BSA** – Bovine serum albumin; **dNTPs** – Deoxynucleosides triphosphate; **[conc.]** – Concentration; **T –** Temperature; **Δt –** Time variation; **bp** – base pairs.

Table S.15 – Automated Sanger sequencing reagents and conditions.

Reaction Mix			Reaction conditions		
Reagents	**[conc.]**	**Volume (µL)**	**T(ºC)**	**Δt**	
			96	4'	
BigDye Buffer	5x	1.75	96	10"	x34
BigDye	2x	0.50	55	5"	
Primer Fw or RV	2 pmol/µL	1.00	60	4'	
Purified PCR product	-	5.25	60	8'	
			4	pause	

Table S.16 – IUPAC code for nucleotides.

IUPAC Nucleotide code	Bases	Description
A	A	Adenine
C	C	Cytosine
G	G	Guanine
T (or U)	T (or U)	Thymine (or Uracil)
W	A or T	Weak
S	C or G	Strong
M	A or C	aMino
K	G or T	Keto
R	A or G	puRine
Y	C or T	pYrimidine
B	C or G or T	not A (B comes after A)
D	A or G or T	not C (D comes after C)
H	A or C or T	not G (H comes after G)
V	A or C or G	not T (V comes after T and U)
N	any base	any nucleotide (not a gap)
. or -	-	Gap

Table S.17 – IUPAC code for amino acid.

IUPAC Amino acid code	3-Letter code	Description
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid or aspartate
E	Glu	Glutamic acid or glutamate
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

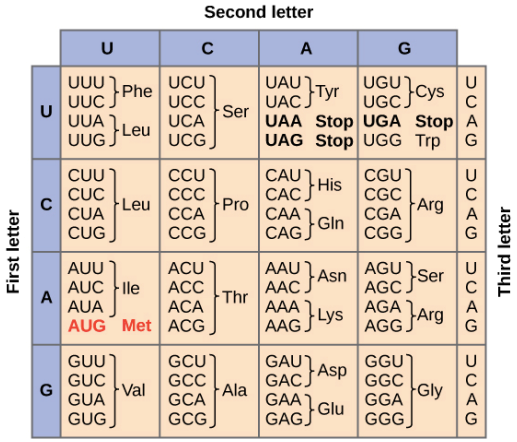


Figure S.6 – Standard genetic code. The coding sequence of a gene specifies the order in which amino acids are linked together in the encoded protein, with each unique triplet of nucleotide bases (codon) specifying a certain amino acid or a punctuation mark (start and stop). From OpenStax College, Biology (https://cnx.org/contents/GFy_h8cu@9.87:QEibhJMi@8/The-Genetic-Code).

Table S.18 – Conventional PCR conditions for amplification of *TMPRSS6* gene for variant confirmation.

Gene	Region	Primers ID	Length (nt)	Sequence	CG%	Tm (°C)	Master mix			PCR conditions			Fragment size (bp)
							Reagents	[conc.]	Volume (µL)	T(ºC)	Δt		
TMPRSS6	Exon 2	Ex2_Fw	21	5'-TGCCGCCTGATGTTGTTACTC-3'	52.4	59.8	Buffer β	-	22.90	94	5'		395
										94	20"	x32	
							Primer Ex2_Fw	25 µM	0.25	63	20"		
							Primer Ex2_Rv	25 µM	0.25	72	20"		
		Ex2_Rv	21	5'-GCCTGCTACAGTCACCCCAAG-3	61.9	63.7	*Taq* DNA Polymerase*	5 U/µL	0.10	72	3"		
							DNA	~80 ng/µL	0.50	4	pause		
	Exon 6	Ex6_Fw	20	5'-AGACAAGGCTGGCTCCAAGG-3'	60	61.4	Buffer β	-	22.90	94	5'		255
										94	20"	x32	
							Primer Ex6_Fw	25 µM	0.25	63	20"		
							Primer Ex6_Rv	25 µM	0.25	72	20"		
		Ex6_Rv	21	5'-CCCTGCACACACAACAGAAGC- 3'	57.1	61.8	*Taq* DNA Polymerase*	5 U/µL	0.10	72	3"		
							DNA	~80 ng/µL	0.50	4	pause		

Continued

Gene	Region	Primers ID	Length (nt)	Sequence	CG%	Tm (˚C)	Master mix			PCR conditions			Fragment size (bp)
							Reagents	[conc.]	Volume (µL)	T(ºC)	Δt		
TMPRSS6	Exon 8	Ex8_Fw	21	5'-GATGTCCAGACTCCCGTCCAC- 3'	61.9	63.7	Buffer β	-	22.90	94	5'		364
							Primer Ex8_Fw	25 µM	0.25	94	20"		
							Primer Ex8_Rv	25 µM	0.25	63	20"	x32	
							Taq DNA Polymerase*	5 U/µL	0.10	72	20"		
		Ex8_Rv	21	5'-GAATCTTCCCTCTCCCCATCC- 3'	57.1	61.8	DNA	~80 ng/µL	0.50	72	3"		
										4	pause		
	Exon 9	Ex9_Fw	21	5'-ATTTGCTGGCAGAGGTGGTAG- 3'	52.4	59.8	Buffer β	-	22.90	94	5'		458
							Primer Ex9_Fw	25 µM	0.25	94	20"		
							Primer Ex9_Rv	25 µM	0.25	63	20"	x32	
							Taq DNA Polymerase*	5 U/µL	0.10	72	20"		
		Ex9_Rv	21	5'-GGAAACACAGAATCCCAGGTG- 3'	52.4	59.8	DNA	~80 ng/µL	0.50	72	3"		
										4	pause		
	Exon 11	Ex11_Fw	21	5'-AGGGAGAAATCAGGGCAGAGG- 3'	57.1	61.8	Buffer β	-	22.90	94	5'		356
							Primer Ex11_Fw	25 µM	0.25	94	20"		
							Primer Ex11_Rv	25 µM	0.25	63	20"	x32	
							Taq DNA Polymerase*	5 U/µL	0.10	72	20"		
		Ex11_Rv	20	5'-CCTTGGTGGTTCCAGGGATG- 3'	60	61.4	DNA	~80 ng/µL	0.50	72	3"		
										4	pause		
	Exon 13	Ex13_Fw	24	5'-GTGATTGGTAACGTGCAATACAGC- 3'	45.8	61	Buffer β	-	22.90	94	5'		285
							Primer Ex13_Fw	25 µM	0.25	94	20"		
							Primer Ex13_Rv	25 µM	0.25	63	20"	x32	
							Taq DNA Polymerase*	5 U/µL	0.10	72	20"		
		Ex13_Rv	22	5'-TGAAGCATGTAGCAGGCCTAGANA 3'	50	60.3	DNA	~80 ng/µL	0.50	72	3"		
										4	pause		
	Exon 17	Ex17_Fw	21	5'-GTGGGCAGAGCAGGAGAGAAG- 3'	61.9	63.7	Buffer β	-	22.90	94	5'		337
							Primer Ex17_Fw	25 µM	0.25	94	20"		
							Primer Ex17_Rv	25 µM	0.25	63	20"	x32	
							Taq DNA Polymerase*	5 U/µL	0.10	72	20"		
		Ex17_Rv	21	5'-GATGTGAGCAAAGGGCCAGAC- 3'	57.1	61.8	DNA	~80 ng/µL	0.50	72	3"		
										4	pause		

* The *Taq* DNA polymerase use was *Taq* DNA Polymerase (Fermentas, USA). Primers melting temperature and CG% were calculated in Oligo Analysis Tool (Eurofins Genomics, Luxembourg; https://www.eurofinsgenomics.eu/en/ecom/tools/oligo-Analysis/). **Fw** – Forward; **Rv** – Reverse; **nt** –Nucleotide; **Tm** – Melting temperature; **[conc.]** – Concentration; **T –** Temperature; **Δt –** Time variation; **bp** – base pairs.

Table S.19 – Population study database variable description.

Variable Group	Variable	Variable description
Sample identification	Group	Groups in study: BTT-Beta Thalassemia Trait; IDA - Iron Deficiency Anemia and Controls
	MA	Subjects with microcytic anemia from BTT and IDA
	Ref	Subject ID (project or DGH)
Demographic characteristics	Sex	Sex (Male or Female)
	Age	Age at sample – All adults (>16)
Hematological parameters	RBC	Red Blood Cells (^12/L) – Index cutoff – 5
	Hb	Hemoglobin (g/dL)
	A	Anemia according to WHO (L - Mild, M - Moderate and S - Severe)
	MCV	Mean Cell Volume (fL)
	MCH	Mean Cell Hemoglobin (pg)
	RDW	Red Cell Distribution Width (%) – Bessman Index, cutoff – 15
	HbA2	Hemoglobin A2 (%)
Hematological Indices	E&F	England and Fraser Index, cutoff – 0
	Mentzer	Mentzer Index, cutoff – 13
	Srivastava	Srivastava Index, cutoff - 3.8
	S&L	Shine and Lal index, cutoff – 1530
	Ricerca	Ricerca Index, cutoff - 4.4
	RDWI	Jayabose Index, cutoff – 220
	G&K	Green and King Index, cutoff – 65
	MDHL	Mean Density of Hemoglobin per Liter Index, cutoff - 1.63
	MCHD	Mean Cell Hemoglobin Density Index, cutoff - 0.3045
	Sirdah	Sirdah Index, cutoff – 27
	Ehsani	Ehsani Index, cutoff – 15
Iron status	SI	Serum Iron (µg/dL)
	Tf	Transferrin (mg/dl)
	TIBC	Total Iron Binding Capacity (µg/dL)
	TSAT	Transferrin Saturation (%)
	Ft	Ferritin (µg/L)
	IS	Iron Saturation according to WHO (ID - Iron deficiency, N - Normal and IO - Iron overload)
Genetic variants	HBB	Molecular cause for Beta thalassemia trait - [HET] in the gene *HBB*
	BTT	Type of Beta Thalassemia Trait (β0 or β+)
	HBD	Possible variants in the gene *HBD*
	H63D	Genotype for variant H63D in the gene *HFE*
	C282Y	Genotype for variant C282Y in the gene *HFE*
	V736A	Genotype for variant V736A in the gene *TMPRSS6*
	K253E	Genotype for variant K253E in the gene *TMPRSS6*

Table S.20 – Detected variants in the *TMPRSS6* gene by NGS analysis of 10 subjects with suspicion of IRIDA.

ID	rs_ID	Location	QUAL	MC	Region	NM_153609.3	NP_705837.1
1	rs2235321	22:37066886-37066886	28507.88	Syn	Ex17	c.2217C>T	p.Tyr739=
2	rs855791	22:37066896-37066896	32317.94	Miss	Ex17	c.2207T>C	p.Val736Ala
3	rs796430458	22:37067030-37067033	3263.39	Int	Int16	c.2141-71_2141-69del	-
4	rs6000550	22:37067410-37067410	3750.89	Int	Int16	c.2141-448G>A	-
5	rs877908	22:37068584-37068584	21117.25	Int	Int16	c.2140+489A>G	-
6	rs148550682	22:37070116-37070116	1071.39	Int	Int15	c.1868+368G>A	-
7	rs5756504	22:37071230-37071230	3978.90	Int	Int13	c.1583-198G>A	-
8	rs5756505	22:37071314-37071314	12217.90	Int	Int13	c.1583-282C>G	-
9	rs5756506	22:37071352-37071352	13876.82	Int	Int13	c.1583-320C>G	-
10	rs4820268	22:37073551-37073551	23514.90	Syn	Ex13	c.1563C>T	p.Asp521=
11	rs13055107	22:37073781-37073781	1572.39	Int	Int12	c.1469-136C>T	-
12	rs2076085	22:37074001-37074001	22941.88	Int	Int12	c.1469-356G>T	-
13	rs2413450	22:37074184-37074184	20270.90	Int	Int12	c.1468+426A>G	-
14	rs2076086	22:37074496-37074496	15670.09	Int	Int12	c.1468+114G>A	-
15	rs2072860	22:37074564-37074564	20833.89	Int	Int12	c.1468+46C>T	-
16	rs763728125	22:37075000-37075004	4193.01	Int	Int11	c.1369+131_1369+134del	-
17	rs149728170	22:37075017-37075017	4336.22	Int	Int11	c.1369+118A>T	-
18	rs881144	22:37075250-37075250	6859.39	Syn	Ex11	c.1254C>T	p.Tyr418=
19	rs9610643	22:37075345-37075345	7098.87	Int	Int10	c.1224-65C>T	-
20	rs73160055	22:37084565-37084565	7797.10	Int	Int9	c.1114-161G>A	-
21	rs2111833	22:37084757-37084757	26320.92	Syn	Ex9	c.1083G>A	p.Ser361=
22	rs113040183	22:37084962-37084962	4388.39	Int	Int8	c.1001-123A>G	-
23	rs2179229	22:37085001-37085001	153009.04	Int	Int8	c.1001-162C>G	-
24	rs2743821	22:37085303-37085303	78352.79	Int	Int8	c.1001-464C>A	-
25	rs2235325	22:37085597-37085597	6931.53	Int	Int8	c.1000+686G>A	-
26	rs77254882	22:37085644-37085644	5280.10	Int	Int8	c.1000+639G>A	-
27	rs145053404	22:37086412-37086412	876.39	Miss	Ex8	c.871G>A	p.Gly291Ser
28	rs112165669	22:37086824-37086824	27473.41	Int	Int7	c.864-405C>T	-
29	rs73415576	22:37087191-37087191	37345.10	Int	Int7	c.864-772G>T	-
30	rs12171211	22:37087296-37087296	32575.10	Int	Int7	c.864-877T>C	-
31	rs147720102	22:37087440-37087440	3555.39	Int	Int7	c.864-1021G>A	-
32	rs5756507	22:37087441-37087441	59704.95	Int	Int7	c.864-1022T>C	-
33	rs11089821	22:37087651-37087651	10946.89	Int	Int7	c.864-1232T>C	-
34	rs113903305	22:37087719-37087719	1680.39	Int	Int7	c.864-1300G>A	-
35	rs9607412	22:37088297-37088297	98341.89	Int	Int7	c.863+1281A>G	-
36	rs9610646	22:37088334-37088334	60172.96	Int	Int7	c.863+1244A>C	-
37	rs9610647	22:37088904-37088904	24112.96	Int	Int7	c.863+674C>T	-
38	rs2235323	22:37089153-37089153	73266.00	Int	Int7	c.863+425A>G	-
39	rs2235324	22:37089684-37089684	35363.96	Miss	Ex7	c.757A>G	p.Lys253Glu
40	rs2543510	22:37095782-37095782	36253.96	Int	Int5	c.616+124C>T	-
41	rs2543509	22:37095799-37095799	39929.97	Int	Int5	c.616+107T>C	-
42	rs2743825	22:37096161-37096161	100591.97	Int	Int4	c.432-71A>C	-
43	rs73421618	22:37096398-37096398	1752.39	Int	Int4	c.431+250C>T	-

Continued

ID	rs_ID	Location	QUAL	MC	Region	NM_153609.3	NP_705837.1
44	rs5756512	22:37096545-37096545	88275.98	Int	Int4	c.431+103G>A	-
45	rs116623227	22:37096747-37096747	4977.39	Int	Int3	c.364-32G>A	-
46	rs34235814	22:37097097-37097097	170038.09	Int	Int3	c.364-384_364-383dup	-
47	rs2160906	22:37097138-37097138	38948.11	Int	Int3	c.364-423C>T	-
48	rs114477604	22:37097330-37097330	6840.39	Int	Int3	c.364-615C>G	-
49	rs529311198	22:37097438-37097438	1761.39	Int	Int3	c.364-723G>A	-
50	rs35771797	22:37098161-37098161	5379.53	Int	Int3	c.363+255A>G	-
51	rs732755	22:37098278-37098278	8076.53	Int	Int3	c.363+138A>G	-
52	rs732756	22:37098380-37098380	4683.53	Int	Int3	c.363+36A>G	-
53	rs8139988	22:37098681-37098681	27468.53	Int	Int2	c.230-132A>G	-
54	rs228904	22:37098774-37098774	40707.53	Int	Int2	c.230-225T>C	-
55	rs2017764	22:37098875-37098875	83876.98	Int	Int2	c.230-326C>T	-
56	rs732756	22:37099011-37099011	46417.94	Int	Int2	c.230-462A>G	-
57	rs61382053	22:37099296-37099296	1242.39	Int	Int2	c.230-747A>G	-
58	rs2281090	22:37099330-37099330	29678.53	Int	Int2	c.230-781C>T	-
59	rs2281091	22:37099510-37099510	31248.53	Int	Int2	c.230-961G>A	-
60	rs116189950	22:37099586-37099586	6077.39	Int	Int2	c.230-1037T>G	-
61	rs2281092	22:37099595-37099595	31708.53	Int	Int2	c.230-1046C>T	-
62	rs115610891	22:37099752-37099752	5544.39	Int	Int2	c.230-1203T>G	-
63	rs228906	22:37099782-37099782	97379.05	Int	Int2	c.230-1233T>C	-
64	rs140612996	22:37099877-37099877	19300.57	Int	Int2	c.230-1353_230-1329dup	-
65	rs9610649	22:37100087-37100087	46706.96	Int	Int2	c.230-1538G>A	-
66	rs4560232	22:37100243-37100243	69020.96	Int	Int2	c.230-1694C>T	-
67	rs7289616	22:37100370-37100370	55206.96	Int	Int2	c.230-1821C>T	-
68	rs570500627	22:37100414-37100414	86302.00	Int	Int2	c.230-1866dup	-
69	rs55945931	22:37100420-37100420	93703.96	Int	Int2	c.230-1871T>A	-
70	rs9619658	22:37100807-37100807	62704.96	Int	Int2	c.230-2258G>A	-
71	rs228907	22:37101553-37101553	36373.09	Int	Int2	c.229+1663C>T	-
72	rs74408154	22:37101811-37101811	9187.39	Int	Int2	c.229+1405G>A	-
73	rs17750152	22:37101986-37101986	25137.53	Int	Int2	c.229+1230G>A	-
74	rs112390782	22:37102038-37102039	42432.53	Int	Int2	c.229+1177del	-
75	rs57009010	22:37102284-37102284	3210.39	Int	Int2	c.229+932C>T	-
76	rs4140590	22:37102691-37102691	60012.92	Int	Int2	c.229+525C>T	-
77	rs4140589	22:37102979-37102979	58856.92	Int	Int2	c.229+237C>T	-
78	rs5995380	22:37103006-37103006	48367.09	Int	Int2	c.229+210G>A	-
79	rs11704654	22:37103346-37103346	35720.09	Syn	Ex2	c.99G>A	p.Pro33=
80	rs5756515	22:37103652-37103652	23560.92	5'UTR	5'UTR	c.-113T>C	-
81	rs5756516	22:37103659-37103659	18899.92	UPS	-	-	-

rs_ID – reference single nucleotide polymorphism ID; **QUAL** – Variant quality score; **NM** – Reference sequence based on a protein coding RNA (mRNA); **NP** – Reference sequence based on a protein (amino acid) sequence; **MC** – Molecular consequence; **UPS** – Upstream; **5' UTR** – Five prime untranslated region; **Miss** – Missense; **Syn** – Synonym; **Int** – Intronic variant; **Region** – Genomic Region.

Table S.21 – Hematological and biochemical data and comparison of case and control groups as well as sex groups.

	Controls			MW	BTT			MW	IDA			MW
Parameters	Total (N=134)	Male (n=42)	Female (n=92)	*p*[a]	Total (N=116)	Male (n=63)	Female (n=53)	*p*[b]	Total (N=67)	Male (n=10)	Female (n=57)	*p*[c]
Age (y)	39±0.42	43±19	38±14	0.341	45±18	44±18	45±18	0.929	44±16	59±16	41±15	**0.028**
RBC (x10^12/L)	4.34±1.24	4.71±0.4	4.18±0.31	**< 0.001**	5.7±0.61	6.05±0.46	5.28±0.48	**< 0.001**	4.45±0.43	4.68±0.46	4.41±0.41	0.458
Hb (g/dL)	13.74±4.52	15.05±1.08	13.14±0.74	**< 0.001**	11.82±1.2	12.48±1.01	11.05±0.91	**< 0.001**	10.23±1.45	10.39±1.67	10.2±1.43	0.916
MCV (fL)	91.45±1.93	92.6±4.67	90.93±4.38	0.079	64.8±4.21	64.06±4.03	65.66±4.29	0.103	72.4±7.67	71.22±6.34	72.6±7.91	0.684
MCH (pg)	31.71±1.56	32.06±1.82	31.54±1.96	0.205	20.79±1.39	20.62±1.18	21±1.58	0.449	23.11±3.35	22.23±2.45	23.26±3.48	0.684
RDW (%)	12.67±41.44	12.66±0.77	12.68±1.81	0.343	15.55±4.33	16.32±5.61	14.61±1.4	0.195	24.11±12.47	31.64±16.13	22.79±11.38	0.458
HbA2 (%)	-	-	-	-	4.13±0.43	4.09±0.36	4.17±0.48	0.425	-	-	-	-
SI (µg/dL)	113.53±42.41	120.9±39.76	110.16±41.97	0.111	108.06±31.05	109.46±28.5	106.51±33.88	0.664	33.35±17.68	27.89±10.45	34.38±18.63	0.684
Tf (mg/dL)	285.57±53	276.33±37.45	289.78±44.03	0.205	260.18±41.05	251.49±33.37	270.04±46.71	**0.029**	376.07±70.29	407±94.92	370.02±64.02	0.458
TIBC (µg/dL)	356.96±12.7	345.42±46.81	362.23±55.04	0.205	326.97±54.9	314.36±41.71	341.27±64.27	**0.029**	470.04±87.81	508.75±118.66	462.46±79.94	0.458
TSAT (%)	32.48±73.24	35.46±11.72	31.12±12.96	**0.048**	33.88±10.73	35.2±9.55	32.39±11.86	0.304	7.41±4.54	5.96±2.98	7.69±4.75	0.645
Ft (µg/L)	81.03±12.7	139.7±89.29	53.34±42.32	**< 0.001**	169.36±205.21	206.56±184.81	129.55±220.22	**< 0.001**	8.61±10.82	14.73±20.37	7.45±7.88	0.684

[a] – Comparison between the parameters of male and female subjects in the control group; **[b] –** Comparison between the parameters of male and female subjects in the BTT group; **[c] –** Comparison between the parameters of male and female subjects in the IDA group; **BTT** – Beta thalassemia trait; **IDA** – Iron deficiency anemia; **SD** – Standard deviation; ***p* – *p*-value (significant < 0.05)**; **MW –** Mann-Whitney test; **y** – Years; **m** – Month; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.22 – Hematological and biochemical data in control subjects, clustered for gender and V736A status.

Sex	Controls													
	Male (n = 37)							Female (n=86)						
Genotype	TT	TC	CC	KW	Mann-Whitney test			TT	TC	CC	KW	Mann-Whitney test		
	WT	HET	HOM					WT	HET	HOM				
Parameters	(n=2)	(n=18)	(n=17)	p^a	p^b	p^c	p^d	(n=6)	(n=43)	(n=37)	p^a	p^b	p^c	p^d
RBC (x10^12/L)	4.45±0.21	4.67±0.40	4.69±0.43	0.751	0.610	0.633	1.000	4.43±0.19	4.13±0.25	4.16±0.37	0.633	0.053	0.115	0.665
Hb (g/dL)	14.00±0.42	15.06±1.22	15.12±1.04	0.693	0.561	0.633	1.000	13.07±0.50	12.98±0.73	13.22±0.75	0.679	0.891	0.699	0.665
MCV (fL)	90.75±2.47	92.59±3.91	92.94±5.82	0.693	0.610	0.633	1.000	86.32±3.19	90.19±4.34	92.03±4.01	0.633	0.086	**0.048**	0.632
MCH (pg)	31.40±0.57	32.34±1.46	32.38±2.16	0.693	0.610	0.633	1.000	29.55±1.52	31.51±2.08	31.82±1.76	0.633	0.086	**0.049**	0.665
RDW (%)	12.10±0.28	12.89±0.82	12.61±0.73	0.693	0.561	0.633	1.000	13.13±0.72	12.92±2.46	12.52±0.88	0.679	0.199	0.226	0.794
SI (µg/dL)	96.00±26.87	120.33±32.20	121.82±50.32	0.693	0.610	0.633	1.000	76.83±27.77	105.26±34.00	116.65±47.21	0.633	0.125	0.054	0.665
Tf (mg/dL)	279.50±0.71	277.39±43.77	281.18±36.43	0.693	0.674	0.894	1.000	304.67±36.87	285.44±44.56	296.68±45.75	0.633	0.285	0.600	0.665
TIBC (µg/dL)	349.38±0.88	346.74±54.71	351.47±45.54	0.693	0.674	0.894	1.000	380.83±46.09	356.80±55.70	370.84±57.19	0.633	0.285	0.600	0.665
TSAT (%)	27.49±7.76	35.52±10.73	34.86±13.82	0.693	0.610	0.633	1.000	20.15±6.64	30.38±11.53	32.02±14.02	0.633	0.086	**0.049**	0.665
Ft (µg/L)	55.45±7.28	140.97±97.11	133.93±86.23	0.693	0.561	0.633	1.000	32.00±13.99	49.34±35.01	60.31±52.53	0.633	0.456	0.400	0.665

Mean and standard deviation are presented. **KW –** Kruskal-Wallis test; **WT –** Wild-type ; **HET** – Heterozygous ; **HOM –** Homozygous; ***p* – *p*-value (significant < 0.05)**; **[a] –** TT *vs* TC *vs* CC; **[b] –** TT *vs* TC; **[c] –** TT *vs* CC; **[c] –** TT *vs* CC; **[d] –** TC *vs* CC; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.23 – Hematological and biochemical data in beta-thalassemia trait subjects, clustered for gender and V736A status.

	BTT													
Sex	Male (n=49)							Female (n=40)						
Genotype	TT	TC	CC	KW	Mann-Whitney test			TT	TC	CC	KW	Mann-Whitney test		
Parameters	WT (n=5)	HET (n=15)	HOM (n=29)	p^a	p^b	p^c	p^d	WT (n=5)	HET (n=15)	HOM (n=20)	p^a	p^b	p^c	p^d
RBC (x10^12/L)	6.00±0.53	5.44±0.60	5.97±0.51	0.531	0.921	0.721	0.721	5.44±0.60	5.20±0.61	5.32±0.42	0.755	0.950	0.709	0.973
Hb (g/dL)	11.82±0.98	11.40±0.48	12.56±1.03	0.697	0.231	0.499	0.499	11.40±0.48	11.03±1.08	11.14±0.92	0.863	0.964	0.709	0.973
MCV (fL)	61.50±2.89	66.34±5.72	64.90±3.47	0.653	0.199	0.327	0.327	66.34±5.72	66.36±5.39	65.28±3.67	0.755	0.964	0.709	0.973
MCH (pg)	19.64±0.80	21.24±2.50	21.05±0.87	0.653	**0.045**	0.205	0.205	21.24±2.50	21.37±1.96	20.98±1.20	0.863	0.950	0.709	0.973
RDW (%)	13.64±1.20	14.36±1.04	17.11±6.61	0.653	0.114	0.205	0.205	14.36±1.04	14.69±1.14	14.63±1.49	0.863	0.950	0.709	0.973
SI (µg/dL)	95.60±26.80	101.60±28.50	104.19±29.39	0.653	0.921	0.655	0.655	101.60±28.50	112.93±39.03	112.45±33.07	0.863	0.950	0.709	0.973
Tf (mg/dL)	242.40±52.29	287.80±29.63	240.06±32.04	0.653	0.921	0.655	0.655	287.80±29.63	288.86±55.15	256.45±45.13	0.755	0.964	0.709	0.636
TIBC (µg/dL)	303.00±65.36	359.75±37.04	300.08±40.05	0.653	0.921	0.655	0.655	359.75±37.04	361.07±68.94	330.25±73.52	0.755	0.964	0.709	0.661
TSAT (%)	32.07±8.01	28.16±6.97	34.87±9.30	0.653	0.921	0.725	0.725	28.16±6.97	32.77±13.08	35.45±12.49	0.755	0.950	0.709	0.973
Ft (µg/L)	232.60±202.73	100.88±33.59	284.53±263.12	0.653	0.921	0.655	0.655	100.88±33.59	93.18±89.86	160.25±332.61	0.755	0.950	0.709	0.973

Mean and standard deviation are presented. **BTT –** Beta-thalassemia trait; **KW –** Kruskal-Wallis test; **WT –** Wild-type ; **HET** – Heterozygous ; **HOM –** Homozygous; ***p* – *p*-value (significant < 0.05)**; **[a] –** TT *vs* TC *vs* CC; **[b] –** TT *vs* TC; **[c] –** TT *vs* CC; **[c] –** TT *vs* CC; **[d] –** TC *vs* CC; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.24 – Hematological and biochemical data in iron deficiency anemia subjects, clustered for gender and V736A status.

Sex	IDA									
	Male (n=9)			Female (n=29)						
Genotype	TC	CC	Mann-Whitney test	TT	TC	CC	KW	Mann-Whitney test		
	HET	HOM		WT	HET	HOM				
Parameters	(n=5)	(n=4)	*p*	(n=4)	(n=23)	(n=22)	p^a	p^b	p^c	p^d
RBC (x10^12/L)	4.68±0.29	4.43±0.31	0.905	4.88±0.47	4.39±0.39	4.38±0.40	0.636	0.388	0.538	0.979
Hb (g/dL)	10.50±1.45	9.43±0.83	0.905	10.75±1.34	10.71±1.13	9.60±1.51	0.636	0.900	0.773	0.078
MCV (fL)	72.24±5.44	69.40±8.58	0.905	69.10±2.28	75.80±7.69	69.23±8.47	0.636	0.388	0.960	0.078
MCH (pg)	22.42±2.23	21.53±3.09	0.905	22.08±1.21	24.58±3.48	22.01±3.65	0.636	0.388	0.960	0.084
RDW (%)	33.24±16.67	27.28±18.73	0.905	23.98±11.3	20.09±9.93	19.91±4.98	0.636	0.615	0.960	0.306
SI (µg/dL)	26.60±9.07	29.50±13.23	0.905	27.00±8.89	35.06±20.21	35.10±20.13	0.699	0.900	0.960	0.979
Tf (mg/dL)	433.00±102.38	374.50±86.70	0.905	390.50±62.93	377.65±62.76	375.35±62.85	0.636	0.900	0.960	0.979
TIBC (µg/dL)	541.25±127.98	468.13±108.38	0.905	488.13±78.67	472.10±78.42	469.00±78.41	0.636	0.900	0.960	0.979
TSAT (%)	5.20±2.12	6.93±3.94	0.905	5.71±1.54	7.90±5.22	7.39±4.92	0.636	0.900	0.960	0.979
Ft (µg/L)	10.00±8.72	18.28±27.20	0.905	5.00±1.84	7.10±6.54	5.49±2.88	0.699	0.900	1.000	0.979

Mean and standard deviation are presented. **IDA –** Iron deficiency anemia; **KW –** Kruskal-Wallis test; **WT –** Wild-type ; **HET** – Heterozygous ; **HOM –** Homozygous; ***p* – *p*-value (significant < 0.05)**; **[a] –** TT *vs* TC *vs* CC; **[b] –** TT *vs* TC; **[c] –** TT *vs* CC; **[c] –** TT *vs* CC; **[d] –** TC *vs* CC; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.25 – Hematological and biochemical data in control subjects, clustered for gender and K253E status.

Sex	Controls											
	Male (n=36)						Female (n=83)					
Genotype	AA	AG	GG	AG+GG	KW	MW	AA	AG	GG	AG+GG	KW	MW
	WT	HET	HOM	HET+HOM			WT	HET	HOM	HET+HOM		
Parameters	(n=24)	(n=5)	(n=7)	(n=12)	p^a	p^b	(n=49)	(n=23)	(n=11)	(n=34)	p^a	p^b
RBC (x10^12/L)	4.72±0.44	4.66±0.41	4.51±0.31	4.58±0.34	0.670	0.534	4.19±0.31	4.14±0.33	4.1±0.33	4.13±0.33	0.885	0.443
Hb (g/dL)	15.18±1.29	14.76±0.67	14.83±0.73	14.8±0.68	0.670	0.534	13.04±0.65	13.14±0.88	13.2±0.81	13.16±0.85	0.885	0.794
MCV (fL)	92.18±4.96	93.36±6.08	93.46±3.77	93.42±4.6	0.670	0.696	90.21±4.43	91.2±4.97	91.73±2.93	91.37±4.37	0.784	0.443
MCH (pg)	32.24±1.82	31.9±2.01	32.96±1.62	32.52±1.79	0.793	0.696	31.18±1.86	31.77±2.23	32.42±1.95	31.98±2.14	0.784	0.259
RDW (%)	12.85±0.87	12.78±0.37	12.36±0.53	12.53±0.5	0.670	0.534	12.66±0.95	13.12±3.26	12.65±0.77	12.97±2.7	0.784	0.973
SI (µg/dL)	118.96±47.82	111.6±28.92	128.71±23.04	121.58±25.91	0.670	0.534	108±39.62	105.87±43.44	114.82±43.23	108.76±42.93	0.839	0.973
Tf (mg/dL)	275.92±37.98	273±40.7	303.14±32.92	290.58±37.87	0.670	0.534	301.9±44.51	271.83±44.43	296.18±35.39	279.71±42.77	0.784	0.237
TIBC (µg/dL)	344.9±47.47	341.25±50.88	378.93±41.15	363.23±47.34	0.670	0.534	377.37±55.64	339.78±55.54	370.23±44.24	349.63±53.46	0.784	0.237
TSAT (%)	34.91±13.72	33.44±10.66	34.37±7.72	33.98±8.61	0.670	0.960	29.17±11.64	31.86±13.49	32.11±15.57	31.94±13.96	0.784	0.559
Ft (µg/L)	138.03±85.26	157.14±116.44	74.41±47.68	108.88±89.36	0.670	0.534	53.34±43.5	51.77±31.71	50±62.39	51.18±43.32	0.784	0.973

Mean and standard deviation are presented. **KW –** Kruskal-Wallis test; **MW** – Mann-Whitney test; **WT –** Wild-type ; **HET** – Heterozygous ; **HOM –** Homozygous; ***p* –** *p*-value (**significant < 0.05**); **[a] –** AA *vs* AG *vs* GG; **[b] –** Recessive model: AA *vs* AG+GG; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.26 – Hematological and biochemical data in beta-thalassemia trait subjects, clustered for gender and K253E status.

	BTT											
Sex	Male (N=45)						Female (N=36)					
Genotype	AA	AG	GG	AG+GG	KW	MW	AA	AG	GG	AG+GG	KW	MW
	WT	HET	HOM	HET+HOM			WT	HET	HOM	HET+HOM		
Parameters	(n=30)	(n=6)	(n=9)	(n=15)	p^a	p^b	(n=23)	(n=7)	(n=6)	(n=13)	p^a	p^b
RBC (x10^12/L)	6.08±0.51	5.87±0.45	6.06±0.44	5.98±0.44	0.584	0.697	5.27±0.51	5.34±0.49	5.65±0.29	5.48±0.42	0.734	0.961
Hb (g/dL)	12.4±1.01	12.47±0.48	12.76±1.01	12.64±0.83	0.584	0.645	11.04±0.84	11.6±0.87	11.42±0.51	11.52±0.7	0.734	0.961
MCV (fL)	62.91±3.19	65.85±4.06	65.28±2.94	65.51±3.3	0.584	0.163	65.47±4.47	66.17±5.12	65.05±4.99	65.65±4.88	0.734	0.961
MCH (pg)	20.4±0.93	21.35±1.63	21.09±0.85	21.19±1.18	0.677	0.174	21.1±1.66	21.89±2.03	20.27±0.92	21.14±1.76	0.734	0.961
RDW (%)	17.56±7.29	18.43±7.01	15.24±1.04	16.52±4.56	0.584	0.378	14.45±1.1	14.69±0.77	15.25±2.27	14.95±1.59	0.734	0.961
SI (µg/dL)	109.58±26.13	111.83±32.62	96.44±25.23	102.6±28.36	0.584	0.645	113.39±34.05	113.43±16.74	126.67±34.22	119.54±25.98	0.734	0.961
Tf (mg/dL)	256.31±31.69	226±13.91	252.44±26.87	241.87±25.72	0.584	0.350	264.73±50.39	265.29±44.66	274.33±22.9	269.46±35.18	0.734	0.961
TIBC (µg/dL)	320.38±39.61	282.5±17.39	315.56±33.59	302.33±32.15	0.584	0.350	330.91±62.99	331.61±55.83	375.21±81.58	351.73±69.59	0.734	0.961
TSAT (%)	34.72±9.61	39.71±11.95	30.58±7.42	34.23±10.19	0.584	1.000	35.49±12.44	34.66±5.56	35.18±12.28	34.9±8.85	0.734	0.961
Ft (µg/L)	198.19±200.65	268.6±279.96	199.71±123.63	228.42±195.18	0.584	0.645	94.59±89.32	132.63±96.79	327.68±581.87	240.99±428.22	0.734	0.961

Mean and standard deviation are presented. **BTT –** Beta-thalassemia trait; **KW –** Kruskal-Wallis test; **MW** – Mann-Whitney test; **WT –** Wild-type ; **HET** – Heterozygous ; **HOM –** Homozygous; ***p* – *p*-value (significant < 0.05)**; **[a] –** AA *vs* AG *vs* GG; **[b] –** Recessive model: AA *vs* AG+GG; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.27 – Hematological and biochemical data in iron deficiency anemia subjects, clustered for gender and K253E status.

Sex	IDA								
	Male (n=7)			Female (n=45)					
Genotype	AA	AG	MW	AA	AG	GG	AG+GG	KW	MW
	WT	HET		WT	HET	HOM	HET+HOM		
Parameters	(n=6)	(n=1)	*p*	(n=29)	(n=8)	(n=8)	(n=16)	p^a	p^b
RBC (x10^12/L)	4.45±0.32	4.70	0.642	4.35±0.42	4.38±0.5	4.58±0.31	4.48±0.42	0.595	1.000
Hb (g/dL)	9.8±0.81	8.30	0.556	10.09±1.49	10.34±1.72	10.51±1.11	10.43±1.4	0.595	1.000
MCV (fL)	71.52±5.14	59.90	0.556	72.73±8.5	73.06±10.46	72.14±6.28	72.6±8.35	0.595	1.000
MCH (pg)	22.1±1.67	17.70	0.556	23.33±3.78	23.76±4.37	23.14±2.86	23.45±3.58	0.595	1.000
RDW (%)	34.98±19.19	18.60	0.857	19.81±7.9	20.34±8.66	20.89±9.24	20.61±8.66	0.595	1.000
SI (µg/dL)	27.5±10.01	13.00	0.556	34.34±19.33	45.14±25.29	28±6.11	36.57±19.79	0.712	1.000
Tf (mg/dL)	393±113.64	467.00	0.857	361.48±53.56	402.17±85.61	413.83±57.6	408±69.83	0.595	0.262
TIBC (µg/dL)	491.25±142.05	583.75	0.857	451.74±66.77	502.71±107.01	517.29±72	510±87.29	0.595	0.262
TSAT (%)	6.23±3.27	2.23	0.556	7.93±5.11	8.49±6.06	5.65±1.59	7.07±4.47	0.595	1.000
Ft (µg/L)	15.98±24.09	3.20	0.556	5.93±3.78	8.73±10.16	5.18±1.69	7.31±7.85	0.595	1.000

Mean and standard deviation are presented. **IDA –** Iron deficiency anemia; **KW –** Kruskal-Wallis test; **MW** – Mann-Whitney test; **WT –** Wild-type ; **HET** – Heterozygous ; **HOM –** Homozygous; ***p* – *p*-value (significant < 0.05)**; **[a] –** AA *vs* AG *vs* GG; **[b] –** Recessive model: AA *vs* AG+GG; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.28 – Hematological and biochemical data in control subjects, clustered for gender and H63D status.

Sex	Controls											
	Male (n=42)						Female (n=91)					
Genotype	CC	CG	GG	CG+GG	KW	MW	CC	CG	GG	CG+GG	KW	MW
	WT	HET	HOM	HET+HOM			WT	HET	HOM	HET+HOM		
Parameters	(n=30)	(n=10)	(n=2)	(n=12)	p^a	p^b	(n=56)	(n=31)	(n=4)	(n=35)	p^a	p^b
RBC (x10^12/L)	4.69±0.44	4.73±0.33	4.95±0.21	4.77±0.32	0.577	0.880	4.17±0.28	4.20±0.36	3.98±0.34	4.17±0.36	0.579	0.974
Hb (g/dL)	15.04±1.17	15.04±0.84	15.30±1.13	15.08±0.84	0.577	0.978	13.11±0.68	13.23±0.87	12.73±0.63	13.17±0.85	0.579	0.830
MCV (fL)	92.76±4.93	92.45±4.10	90.80±5.66	92.18±4.14	0.577	0.880	90.69±4.36	91.25±4.53	92.45±4.48	91.39±4.48	0.758	0.575
MCH (pg)	32.19±1.85	31.89±1.54	31.05±3.46	31.75±1.77	0.577	0.880	31.50±1.86	31.62±2.15	31.98±2.29	31.66±2.13	0.579	0.575
RDW (%)	12.76±0.75	12.47±0.80	12.15±0.92	12.42±0.79	0.577	0.880	12.39±0.88	13.25±2.82	12.68±0.15	13.19±2.66	0.579	0.232
SI (µg/dL)	121.80±38.32	126.70±44.95	78.50±9.19	118.67±44.87	0.577	0.978	108.41±42.96	115.65±42.83	96.75±22.9	113.49±41.25	0.735	0.575
Tf (mg/dL)	278.90±37.63	271.80±40.54	260.50±27.58	269.92±37.85	0.577	0.880	293.20±43.97	287.10±45.22	276.25±37.62	285.86±44.06	0.579	0.575
TIBC (µg/dL)	348.63±47.03	339.78±50.65	325.63±34.47	337.42±47.30	0.577	0.880	366.50±54.97	358.87±56.52	345.31±47.02	357.32±55.08	0.579	0.575
TSAT (%)	35.38±11.03	37.94±14.00	24.40±5.41	35.68±13.81	0.577	0.978	30.26±12.93	32.99±13.76	28.68±8.71	32.50±13.25	0.579	0.575
Ft (µg/L)	128.71±89.31	167.50±95.93	165.50±9.19	167.17±86.82	0.577	0.880	49.44±41.52	58.08±44.49	75.73±37.54	60.10±43.63	0.579	0.575

Mean and standard deviation are presented. **KW –** Kruskal-Wallis test; **MW** – Mann-Whitney test; **WT –** Wild-type ; **HET** – Heterozygous ; **HOM –** Homozygous; ***p* –** *p*-value (**significant < 0.05**); **[a] –** CC *vs* CG *vs* GG; **[b] –** Dominate model: CC *vs* CG+GG; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.29 – Hematological and biochemical data in beta-thalassemia trait subjects, clustered for gender and H63D status.

	BTT					
Sex	Male (n=63)			Female (n=53)		
Genotype	CC	CG	MW	CC	CG	MW
	WT	HET		WT	HET	
Parameters	(n=42)	(n=21)	*p*	(n=37)	(n=16)	*p*
RBC (x10^12/L)	6.05±0.37	6.06±0.61	0.765	5.24±0.49	5.38±0.47	0.622
Hb (g/dL)	12.34±0.79	12.74±1.32	0.123	11.06±0.95	11.00±0.84	0.622
MCV (fL)	63.32±3.77	65.63±4.19	0.123	66.06±4.23	64.73±4.4	0.317
MCH (pg)	20.42±1.19	21.01±1.09	0.123	21.19±1.68	20.57±1.27	0.317
RDW (%)	16.78±6.15	15.43±4.38	0.271	14.69±1.51	14.43±1.15	0.633
SI (µg/dL)	105.36±24.80	117.45±33.89	0.230	100.70±37.13	119.94±19.84	0.132
Tf (mg/dL)	258.46±33.49	237.90±29.36	0.123	274.86±48.71	258.13±40.42	0.317
TIBC (µg/dL)	323.08±41.86	297.38±36.7	0.123	348.82±68.24	322.67±50.52	0.317
TSAT (%)	33.02±8.54	39.45±10.17	0.123	30.14±12.65	37.94±7.40	0.092
Ft (µg/L)	192.00±177.26	236.65±202.52	0.585	135.6±257.26	113.93±65.62	0.317

Mean and standard deviation are presented. **BTT –** Beta-thalassemia trait; **MW** – Mann-Whitney test; **WT –** Wild-type ; **HET** – Heterozygous ; **HOM –** Homozygous; ***p* – *p*-value (significant < 0.05)**; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.30 – Hematological and biochemical data in iron deficiency anemia subjects, clustered for gender and H63D status.

Sex	IDA								
	Male (N=10)			Female (N=53)					
Genotype	CC	CG	MW	CC	CG	GG	CG+GG	KW	MW
	WT	HET		WT	HET	HOM	HET+HOM		
Parameters	(n=9)	(n=1)	*p*	(n=35)	(n=15)	(n=3)	(n=18)	p^a	p^b
RBC (x10^12/L)	4.73±0.46	4.20	0.889	4.37±0.43	4.45±0.39	4.70±0.10	4.49±0.37	0.555	0.495
Hb (g/dL)	10.59±1.64	8.60	0.889	10.39±1.38	10.11±1.54	9.47±0.29	10.01±1.42	0.555	0.495
MCV (fL)	71.11±6.72	72.20	1.000	74.05±7.55	71.93±8.42	64.13±4.80	70.63±8.37	0.555	0.495
MCH (pg)	22.41±2.53	20.60	0.889	23.97±3.45	22.76±3.56	20.13±0.91	22.32±3.39	0.555	0.495
RDW (%)	29.64±15.75	49.60	1.000	20.62±10.15	24.47±12.81	24.17±9.74	24.42±12.09	0.555	0.495
SI (µg/dL)	28.75±10.82	21.00	1.000	36.77±20.37	32.31±15.21	15.10±5.80	30.01±15.40	0.555	0.495
Tf (mg/dL)	422.63±88.24	282.00	0.889	371.62±62.43	370.54±61.06	327.50±19.09	364.80±58.75	0.555	0.843
TIBC (µg/dL)	528.28±110.30	352.50	0.889	464.40±77.93	463.17±76.33	409.77±23.91	456.05±73.40	0.555	0.843
TSAT (%)	5.97±3.19	5.96	1.000	8.13±5.10	7.42±4.34	3.72±0.96	6.92±4.23	0.555	0.632
Ft (µg/L)	16.18±21.91	6.00	1.000	7.99±8.82	6.80±7.62	6.55±0.64	6.76±6.96	0.555	0.746

Mean and standard deviation are presented. **IDA –** Iron deficiency anemia; **KW –** Kruskal-Wallis test; **MW** – Mann-Whitney test; **WT –** Wild-type ; **HET** – Heterozygous ; **HOM –** Homozygous; ***p* – *p*-value (significant < 0.05)**; **[a] –** CC *vs* CG *vs* GG; **[b] –** Dominate model: CC *vs* CG+GG; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.31 – **Hematological and biochemical data in control subjects, clustered for gender and C282Y status.**

Sex	Controls								
	Male (N=42)			Female (N=92)					
Genotype	GG	GA	MW	GG	GA	AA	GA+AA	KW	MW
	WT	HET		WT	HET	HOM	HET+HOM		
Parameters	(n=38)	(n=4)	*p*	(n=86)	(n=5)	(n=1)	(n=6)	p^a	p^b
RBC (x10^12/L)	4.66±0.39	5.20±0.08	0.067	4.17±0.30	4.32±0.52	4.32	4.27±0.48	**0.003**	0.806
Hb (g/dL)	14.95±1.02	16.03±1.23	0.290	13.12±0.75	13.34±0.83	13.34	13.33±0.74	0.207	0.806
MCV (fL)	92.76±4.52	91.08±6.52	0.867	90.96±4.42	89.46±3.69	89.46	90.50±4.17	0.738	0.887
MCH (pg)	32.19±1.78	30.83±1.98	0.290	31.55±1.95	31.12±2.42	31.12	31.43±2.30	0.928	0.887
RDW (%)	12.69±0.78	12.35±0.65	0.733	12.72±1.86	12.12±0.70	12.12	12.07±0.64	0.928	0.333
SI (µg/dL)	115.63±31.33	171.00±76.19	0.282	107.80±40.62	152.60±51.14	152.60	144.00±50.36	0.920	0.170
Tf (mg/dL)	275.29±38.73	286.25±22.88	0.867	291.69±42.01	279.00±58.040	279.00	262.50±65.79	0.928	0.254
TIBC (µg/dL)	344.12±48.41	357.81±28.6	0.867	364.61±52.51	348.75±72.55	348.75	328.13±82.24	0.928	0.254
TSAT (%)	34.18±10.10	47.61±20.04	0.290	30.13±12.21	45.42±18.04	45.42	45.33±16.13	0.738	0.170
Ft (µg/L)	140.14±91.47	135.45±75.83	0.983	50.21±39.20	108.76±61.49	108.76	96.72±62.41	0.738	0.235

Mean and standard deviation are presented. **KW –** Kruskal-Wallis test; **MW** – Mann-Whitney test; **WT –** Wild-type ; **HET** – Heterozygous ; **HOM –** Homozygous; ***p* – *p*-value (significant < 0.05)**; **[a] –** GG *vs* GA *vs* AA; **[b] –** Dominate model: GG *vs* GA+GA; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.32 – **Hematological and biochemical data in beta-thalassemia trait subjects, clustered for gender and C282Y status.**

Sex	BTT					
	Male (n=63)			Female (n=53)		
Genotype	GG	GA	MW	GG	GA	MW
	WT	HET		WT	HET	
Parameters	(N=60)	(N=3)	*p*	(N=50)	(N=3)	*p*
RBC (x10^12/L)	6.07±0.45	5.83±0.59	0.680	5.30±0.47	4.97±0.64	0.418
Hb (g/dL)	12.50±1.01	12.03±1.21	0.680	11.06±0.90	10.83±1.23	0.686
MCV (fL)	64.13±4.08	62.87±2.96	0.692	65.29±3.98	71.80±5.50	0.136
MCH (pg)	20.62±1.21	20.60±0.10	0.987	20.95±1.59	21.93±1.40	0.237
RDW (%)	16.07±5.16	21.37±12.21	0.692	14.45±1.16	17.23±2.57	0.136
SI (µg/dL)	108.89±28.84	125.50±3.54	0.680	105.74±34.40	119.33±24.21	0.584
Tf (mg/dL)	252.42±32.88	225.00±50.91	0.680	272.18±46.02	216.50±38.89	0.160
TIBC (µg/dL)	315.53±41.10	281.25±63.64	0.680	344.10±63.55	270.63±48.61	0.160
TSAT (%)	34.82±9.36	45.94±11.65	0.680	31.86±11.77	45.69±3.56	0.160
Ft (µg/L)	200.90±185.46	331.00±159.81	0.680	97.69±109.33	554.40±703.06	0.136

Mean and standard deviation are presented. **BTT –** Beta-thalassemia trait; **MW** – Mann-Whitney test; **WT –** Wild-type; **HET** – Heterozygous; ***p* – *p*-value (significant < 0.05)**; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.33 – Hematological and biochemical data in iron deficiency anemia subjects, clustered for gender and C282Y status.

Sex	IDA			
	Male (n=10)	Female (n=54)		
Genotype	GG	GG	GA	MW
	WT	WT	HET	
Parameters	(n=10)	(n=50)	(n=4)	*p*[a]
RBC (x10^12/L)	4.68±0.46	4.40±0.41	4.60±0.56	0.980
Hb (g/dL)	10.39±1.67	10.25±1.40	10.37±1.50	1.000
MCV (fL)	71.22±6.34	73.10±8.11	69.43±1.98	0.962
MCH (pg)	22.23±2.45	23.46±3.59	22.67±0.87	0.980
RDW (%)	31.64±16.13	21.68±10.86	25.80±12.91	0.962
SI (µg/dL)	27.89±10.45	34.91±19.35	27.00±0.00	1.000
Tf (mg/dL)	407.00±94.92	370.62±61.62	341.50±24.75	0.962
TIBC (µg/dL)	508.75±118.66	463.20±76.93	426.88±30.94	0.962
TSAT (%)	5.96±2.98	7.79±4.92	6.35±0.46	1.000
Ft (µg/L)	14.73±20.37	7.67±8.17	2.80	0.962

Mean and standard deviation are presented. **IDA –** Iron deficiency anemia; **MW** – Mann-Whitney test; **WT –** Wild-type; **HET** – Heterozygous; ***p* – *p*-value (significant < 0.05)**; **[a] –** Female GG *vs* GA; **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.34 – Association of genetic variants with hematological and iron parameters in control group.

Locus, SNP (risk allele)	Controls		
Parameters	β (SE)	95% CI	*p*
TMPRSS6, V736A (T)	TC vs CC	HET-HOM	
RBC (x10^12/L)	-0.10 (0.62)	(-1.33,1.12)	0.869
Hb (g/dL)	-0.04 (0.26)	(-0.56,0.48)	0.886
MCV (fL)	0.06 (0.05)	(-0.04,0.16)	0.237
MCH (pg)	0.02 (0.11)	(-0.19,0.23)	0.861
RDW (%)	-0.18 (0.19)	(-0.64,0.1)	0.352
SI (µg/dL)	0.01 (0.01)	($-4.08x10^{-03}$,0.02)	0.267
Tf (mg/dL)	$4.43x10^{-03}$ ($4.94x10^{-03}$)	(-0.01,0.01)	0.371
TIBC (µg/dL)	$3.54x10^{-03}$ ($3.96x10^{-03}$)	($-4.15x10^{-03}$,0.01)	0.371
TSAT (%)	0.01 (0.02)	(-0.02,0.04)	0.613
Ft (µg/L)	$-9.78x10^{-04}$ ($3.84x10^{-03}$)	(-0.01,0.01)	0.799
TMPRSS6, K253E (G)			
RBC (x10^12/L)	-1.14 (0.66)	(-2.48,0.11)	0.083
Hb (g/dL)	-0.14 (0.26)	(-0.67,0.37)	0.583
MCV (fL)	0.05 (0.05)	(-0.05,0.14)	0.328
MCH (pg)	0.20 (0.11)	(-0.01,0.43)	0.068
RDW (%)	0.07 (0.12)	(-0.18,0.38)	0.591
SI (µg/dL)	2.77x10-04 ($4.90x10^{-03}$)	(-0.01,0.01)	0.955
Tf (mg/dL)	-0.01 (0.01)	(-0.02,$4.37x10^{-03}$)	0.281
TIBC (µg/dL)	-4.31x10-03 ($4.00x10^{-03}$)	(-0.01,$3.50x10^{-03}$)	0.281
TSAT (%)	0.01 (0.02)	(-0.02,0.04)	0.601
Ft (µg/L)	-0.01 (0.01)	(-0.02, $-7.70x10^{-04}$)	0.051
HFE, H63D (G)			
RBC (x10^12/L)	-0.05 (0.63)	(-1.31,1.19)	0.943
Hb (g/dL)	0.01 (0.27)	(-0.52,0.53)	0.981
MCV (fL)	0.02 (0.05)	(-0.07,0.11)	0.664
MCH (pg)	0.01 (0.11)	(-0.19,0.22)	0.889
RDW (%)	0.53 (0.26)	(0.09,1.08)	**0.042**
SI (µg/dL)	$1.21x10^{-03}$ ($4.84x10^{-03}$)	(-0.01,0.01)	0.803
Tf (mg/dL)	$5.29x10^{-04}$ (0.01)	(-0.01,0.01)	0.917
TIBC (µg/dL)	$4.23x10^{-04}$ ($4.04x10^{-03}$)	(-0.01,0.01)	0.917
TSAT (%)	$3.78x10^{-03}$ (0.02)	(-0.03,0.03)	0.810
Ft (µg/L)	$8.20x10^{-04}$ ($4.13x10^{-03}$)	(-0.01,0.01)	0.843

From logistic regression models adjusted for sex and age. **β –** Beta coefficient; **SE –** Standard error; **CI** – Confidence interval ***p* –** *p*-value (**significant < 0.05**); **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin.

Table S.35 – Association of genetic variants with hematological and iron parameters in beta-thalassemia group.

Locus, SNP (risk allele)	BTT		
Parameters	β (SE)	95% CI	*p*
TMPRSS6, V736A (T)	TC vs CC	HET-HOM	
RBC (x10^12/L)	0.49 (0.5)	(-0.48,1.51)	0.327
Hb (g/dL)	0.11 (0.25)	(-0.38,0.61)	0.659
MCV (fL)	-0.03 (0.07)	(-0.17,0.09)	0.598
MCH (pg)	-0.23 (0.21)	(-0.65,0.18)	0.275
RDW (%)	0.01 (0.05)	(-0.09,0.11)	0.902
SI (µg/dL)	$-1.62x10^{-03}$ (0.01)	(-0.02,0.01)	0.844
Tf (mg/dL)	$-2.27x10^{-03}$ (0.01)	(-0.02,0.01)	0.733
TIBC (µg/dL)	$5.00x10^{-04}$ ($4.76x10^{-03}$)	(-0.01,0.01)	0.916
TSAT (%)	$1.25x10^{-03}$ (0.02)	(-0.05,0.05)	0.958
Ft (µg/L)	$-9.05x10^{-04}$ ($1.28x10^{-03}$)	($-3.75x10^{-03}$,$1.60x10^{-03}$)	0.479
HFE, H63D (G)			
RBC (x10^12/L)	-0.61 (0.49)	(-1.60,0.32)	0.208
Hb (g/dL)	-0.20 (0.23)	(-0.66,0.25)	0.384
MCV (fL)	0.01 (0.06)	(-0.10,0.12)	0.810
MCH (pg)	0.13 (0.17)	(-0.19,0.47)	0.448
RDW (%)	0.07 (0.06)	(-0.04,0.22)	0.297
SI (µg/dL)	-0.02 (0.01)	(-0.04,-0.01)	**0.006**
Tf (mg/dL)	0.01 (0.01)	($-4.34x10^{-04}$,0.02)	0.066
TIBC (µg/dL)	0.01 ($4.89x10^{-03}$)	($7.78x10^{-04}$,0.02)	**0.042**
TSAT (%)	-0.08 (0.02)	(-0.13,-0.04)	**0.001**
Ft (µg/L)	$3.96x10^{-04}$ ($1.20x10^{-03}$)	($-1.88x10^{-03}$,$3.10x10^{-03}$)	0.742

From logistic regression models adjusted for sex and age. **BTT –** Beta-thalassemia trait; **β –** Beta coefficient; **SE –** Standard error; **CI** – Confidence interval ***p* –** *p*-value (**significant < 0.05**); **RBC** – Red blood cells; **Hb** – Hemoglobin; **MCV** – Mean Cell Volume; **MCH** – Mean Cell Hemoglobin; **RDW** – Red Cell Distribution Width; **SI** – Serum iron; **Tf** – Transferrin; **TIBC** – Total Iron Binding Capacity; **TSAT** – Transferrin saturation; **Ft** – Ferritin

Table S.36 – Diagnostic accuracy findings for all considered indices used in microcytic anemia subjects.

Indices	Total									
	BTT (N= 111)	IDA (N=61)	Correctly diagnosed	Accuracy (Efficiency)	SENS	SPEC	PPV	NPV	YI	AUC
RBC										
BTT > 5	100	5	156	0.91	0.90	0.92	0.95	0.84	0.82	0.947
IDA < 5	11	56			0.92	0.90	0.84	0.95		
England and Fraser (E&F)										
BTT < 0	76	1	136	0.79	0.68	0.98	0.99	0.63	0.67	0.953
IDA > 0	35	60			0.98	0.68	0.63	0.99		
Mentzer										
BTT < 13	96	9	148	0.86	0.86	0.85	0.91	0.78	0.72	0.941
IDA > 13	15	52			0.85	0.86	0.78	0.91		
Srivastava										
BTT < 3.8	70	2	129	0.75	0.63	0.97	0.97	0.59	0.60	0.917
IDA > 3.8	41	59			0.97	0.63	0.59	0.97		
Shine and Lal (S&L)										
BTT < 1530	111	50	122	0.71	1.00	0.18	0.69	1.00	0.18	0.749
IDA > 1530	0	11			0.18	1.00	1.00	0.69		
Bessman (RDW)										
BTT < 15	73	5	129	0.75	0.66	0.92	0.94	0.60	0.58	0.890
IDA > 15	38	56			0.92	0.66	0.60	0.94		
Ricerca										
BTT < 4.4	106	30	137	0.80	0.95	0.51	0.78	0.86	0.46	0.952
IDA > 4.4	5	31			0.51	0.95	0.86	0.78		
Jayabose (RDWI)										
BTT < 220	98	1	158	0.92	0.88	0.98	0.99	0.82	0.87	0.957
IDA > 220	13	60			0.98	0.88	0.82	0.99		
Green and King (G&K)										
BTT < 65	96	0	157	0.91	0.86	1.00	1.00	0.80	0.86	0.957
IDA > 65	15	61			1.00	0.86	0.80	1.00		
MDHL										
BTT > 1.63	92	2	151	0.88	0.83	0.97	0.98	0.76	0.80	0.951
IDA < 1.63	19	59			0.97	0.83	0.76	0.98		
MCHD										
BTT > 0.3045	94	49	106	0.62	0.85	0.20	0.66	0.41	0.04	0.609
IDA < 0.3045	17	12			0.20	0.85	0.41	0.66		
Sirdah										
BTT < 27	81	2	140	0.81	0.73	0.97	0.98	0.66	0.70	0.945
IDA > 27	30	59			0.97	0.73	0.66	0.98		
Ehsani										
BTT < 15	92	9	144	0.84	0.83	0.85	0.91	0.73	0.68	0.931
IDA > 15	19	52			0.85	0.83	0.73	0.91		

BTT – Beta thalassemia trait; **IDA** – Iron deficiency anemia; **RBC** – Red blood cells; **RDW –** Red cell distribution width**; RDWI –** Red cell distribution width index; **MDHL** – Mean Density of Hemoglobin per Liter; **MCHD** – Mean Cell Hemoglobin Density **SENS** – Sensitivity; **SPEC** – Specificity; **PPV** – Positive predictive values; **NPV** – Negative predictive values; **YI** – Youden's index; **AUC** – Area under the curve.

Table S.37 – Diagnostic accuracy findings for all considered indices used in microcytic anemia male subjects.

Indices	Male									
	BTT (n=60)	IDA (n=10)	Correctly diagnosed	Accuracy (Efficiency)	SENS	SPEC	PPV	NPV	YI	AUC
RBC										
BTT > 5	60	1	69	0.99	1.00	0.90	0.98	1.00	0.90	0.979
IDA < 5	0	9			0.90	1.00	1.00	0.98		
England and Fraser (E&F)										
BTT < 0	57	1	66	0.94	0.95	0.90	0.98	0.75	0.85	0.983
IDA > 0	3	9			0.90	0.95	0.75	0.98		
Mentzer										
BTT < 13	59	2	67	0.96	0.98	0.80	0.97	0.89	0.78	0.989
IDA > 13	1	8			0.80	0.98	0.89	0.97		
Srivastava										
BTT < 3.8	51	1	60	0.86	0.85	0.90	0.98	0.50	0.75	0.983
IDA > 3.8	9	9			0.90	0.85	0.50	0.98		
Shine and Lal (S&L)										
BTT < 1530	60	8	62	0.89	1.00	0.20	0.88	1.00	0.20	0.815
IDA > 1530	0	2			0.20	1.00	1.00	0.88		
Bessman (RDW)										
BTT < 15	37	0	47	0.67	0.62	1.00	1.00	0.30	0.62	0.914
IDA > 15	23	10			1.00	0.62	0.30	1.00		
Ricerca										
BTT < 4.4	55	4	61	0.87	0.92	0.60	0.93	0.55	0.52	0.952
IDA > 4.4	5	6			0.60	0.92	0.55	0.93		
Jayabose (RDWI)										
BTT < 220	54	0	64	0.91	0.90	1.00	1.00	0.63	0.90	0.958
IDA > 220	6	10			1.00	0.90	0.63	1.00		
Green and King (G&K)										
BTT < 65	54	0	64	0.91	0.90	1.00	1.00	0.63	0.90	0.960
IDA > 65	6	10			1.00	0.90	0.63	1.00		
MDHL										
BTT > 1.63	58	1	67	0.96	0.97	0.90	0.98	0.82	0.87	0.973
IDA < 1.63	2	9			0.90	0.97	0.82	0.98		
MCHD										
BTT > 0.3045	52	8	54	0.77	0.87	0.20	0.87	0.20	0.07	0.707
IDA < 0.3045	8	2			0.20	0.87	0.20	0.87		
Sirdah										
BTT < 27	55	1	64	0.91	0.92	0.90	0.98	0.64	0.82	0.987
IDA > 27	5	9			0.90	0.92	0.64	0.98		
Ehsani										
BTT < 15	56	1	65	0.93	0.93	0.90	0.98	0.69	0.83	0.988
IDA > 15	4	9			0.90	0.93	0.69	0.98		

BTT – Beta thalassemia trait; **IDA** – Iron deficiency anemia; **RBC** – Red blood cells; **RDW –** Red cell distribution width**; RDWI –** Red cell distribution width index; **MDHL** – Mean Density of Hemoglobin per Liter; **MCHD** – Mean Cell Hemoglobin Density **SENS** – Sensitivity; **SPEC** – Specificity; **PPV** – Positive predictive values; **NPV** – Negative predictive values; **YI** – Youden's index; **AUC** – Area under the curve.

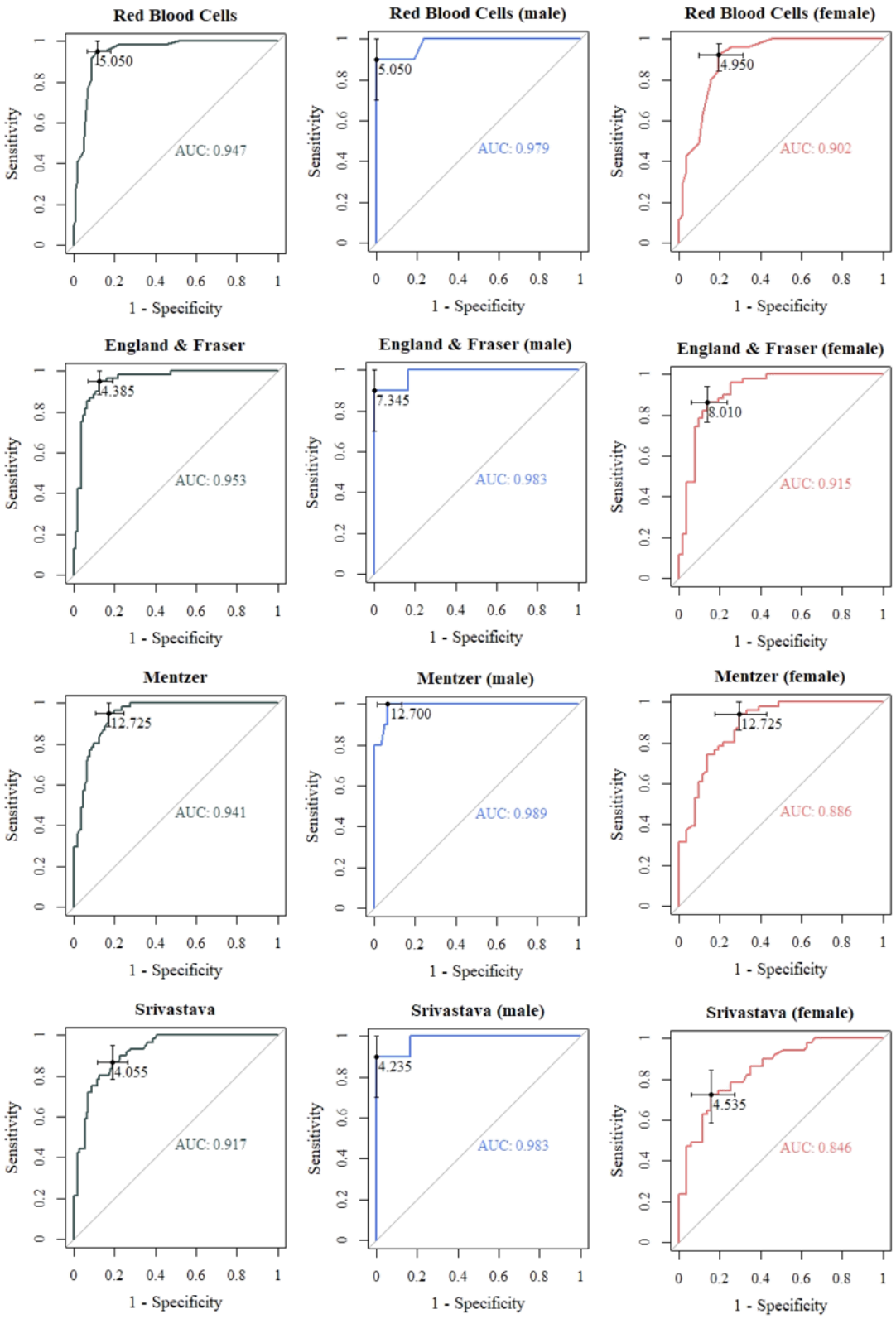


Figure S.7 – Receiver operating characteristics (ROC) curves constructed to visualize the area under the curve (AUC) of Red blood cells, England and Fraser, Mentzer and Srivastava hematologic indices to distinguish BTT cases from IDA. The plots represent analysis with the total population (left) and only the male (middle) and female (right) subjects. In the curve is marked the calculated best thresholds of each index and respective 95% confidence interval.

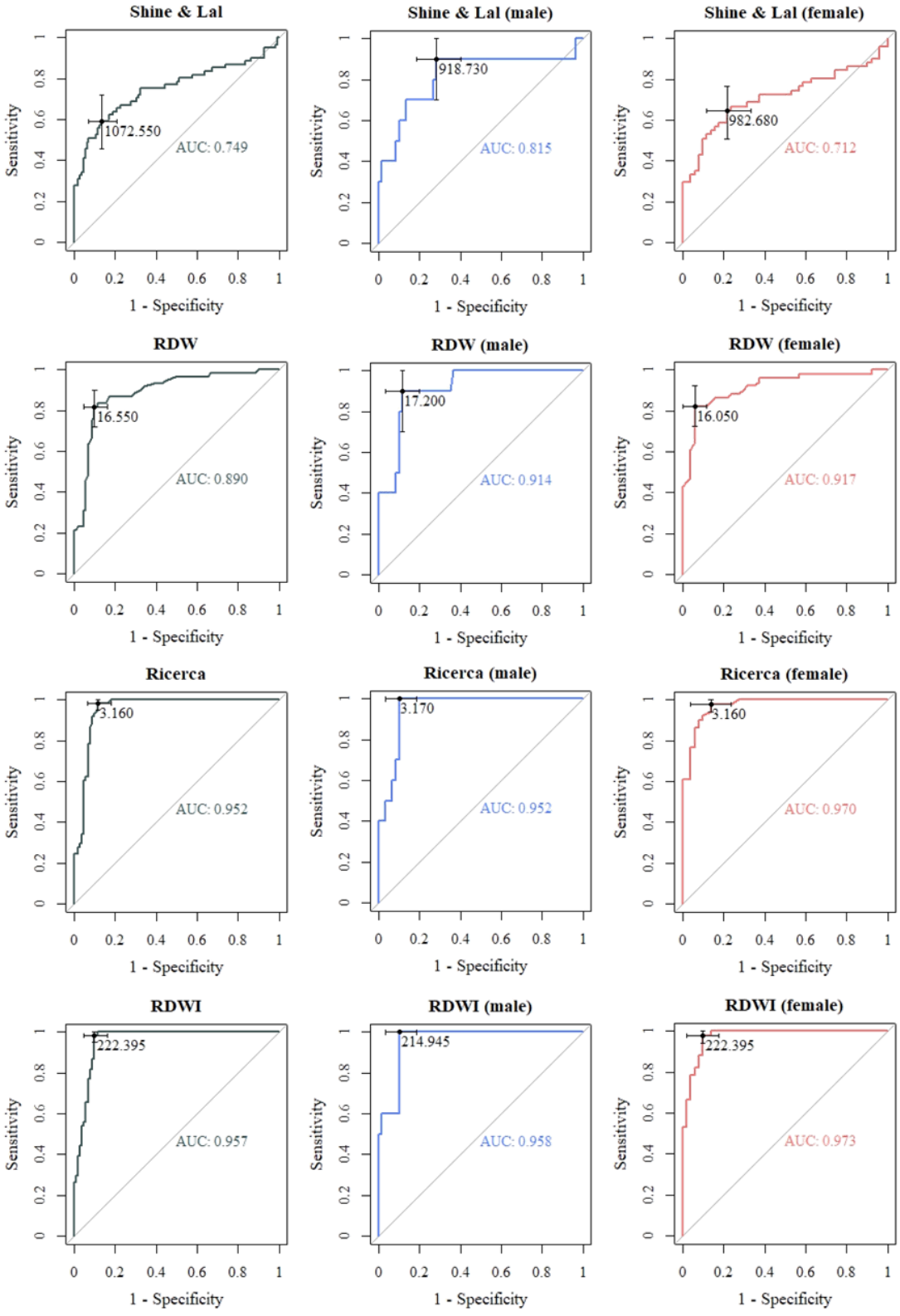


Figure S.8 – Receiver operating characteristics (ROC) curves constructed to visualize the area under the curve (AUC) of Shine and Lal, Bessman (RDW; Red cell distribution width), Ricerca and Jaybose (RDW; Red cell distribution width index) hematologic indices to distinguish BTT cases from IDA. The plots represent analysis with the total population (left) and only the male (middle) and female (right) subjects. In the curve is marked the calculated best thresholds of each index and respective 95% confidence interval.

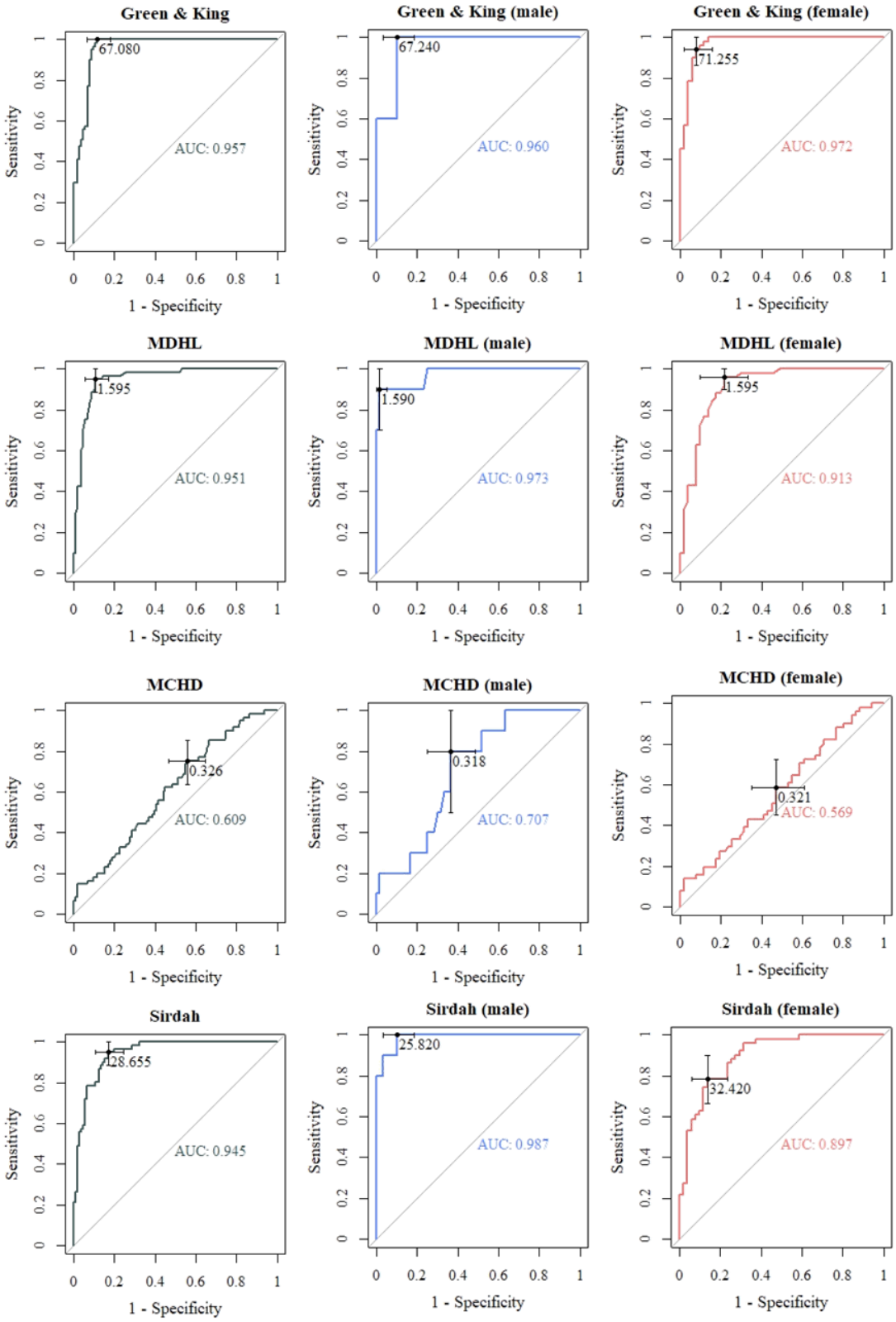


Figure S.9 - Receiver operating characteristics (ROC) curves constructed to visualize the area under the curve (AUC) of Green and King, Mean Density of Hemoglobin per Liter (MDHL), Mean Cell Hemoglobin Density (MCHD) and Sirdah hematologic indices to distinguish BTT cases from IDA. The plots represent analysis with the total population (left) and only the male (middle) and female (right) subjects. In the curve is marked the calculated best thresholds of each index and respective 95% confidence interval.

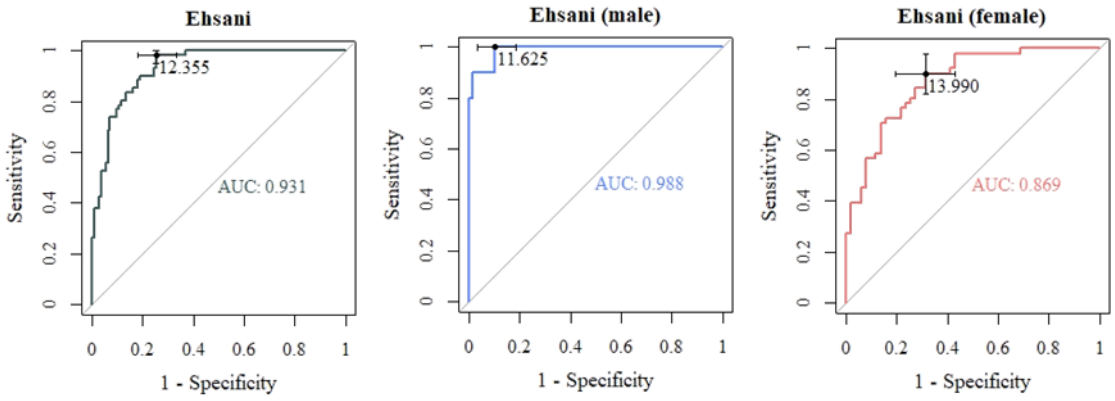


Figure S.10 – Receiver operating characteristics (ROC) curves constructed to visualize the area under the curve (AUC) of Ehsani hematologic index to distinguish BTT cases from IDA. The plots represent analysis with the total population (left) and only the male (middle) and female (right) subjects. In the curve is marked the calculated best thresholds of each and respective 95% confidence interval.

Table S.38 – Accuracy findings for all calculated new best thresholds.

Indices	Total					
	Cutoff	Indices	Cutoff	Indices	Cutoff	Indices
RBC	5	5.05	0.95	0.88	0.82	0.97
England and Fraser (E&F)	0	4.39	0.95	0.87	0.81	0.97
Mentzer	13	12.73	0.95	0.83	0.75	0.97
Srivastava	3.8	4.06	0.87	0.81	0.72	0.92
Shine and Lal (S&L)	1530	1072.55	0.59	0.86	0.71	0.79
Bessman (RDW)	15	16.55	0.82	0.90	0.82	0.90
Ricerca	4.4	3.16	0.98	0.88	0.82	0.99
Jayabose (RDWI)	220	222.40	0.98	0.90	0.85	0.99
Green and King (G&K)	65	67.08	1.00	0.88	0.82	1.00
MDHL	1.63	1.60	0.95	0.89	0.83	0.97
MCHD	0.3045	0.3260	0.75	0.44	0.43	0.77
Sirdah	27	28.66	0.95	0.83	0.75	0.97
Ehsani	15	12.36	0.98	0.75	0.68	0.99

RBC – Red blood cells; **RDW –** Red cell distribution width**; RDWI –** Red cell distribution width index; **MDHL** – Mean Density of Hemoglobin per Liter; **MCHD** – Mean Cell Hemoglobin Density **SENS** – Sensitivity; **SPEC** – Specificity; **PPV** – Positive predictive values; **NPV** – Negative predictive values; **YI** – Youden's index; **AUC** – Area under the curve.

Table S.39 – Accuracy findings for all calculated new best thresholds for male microcytic anemia subjects.

Indices	Male					
	Cutoff	Indices	Cutoff	Indices	Cutoff	Indices
RBC	5	5.05	0.90	1.00	1.00	0.98
England and Fraser (E&F)	0	7.35	0.90	1.00	1.00	0.98
Mentzer	13	12.70	1.00	0.93	0.71	1.00
Srivastava	3.8	4.24	0.90	1.00	1.00	0.98
Shine and Lal (S&L)	1530	918.73	0.90	0.72	0.35	0.98
Bessman (RDW)	15	17.20	0.90	0.88	0.56	0.98
Ricerca	4.4	3.17	1.00	0.90	0.63	1.00
Jayabose (RDWI)	220	214.95	1.00	0.90	0.63	1.00
Green and King (G&K)	65	67.24	1.00	0.90	0.63	1.00
MDHL	1.63	1.59	0.90	0.98	0.90	0.98
MCHD	0.3045	0.3179	0.80	0.63	0.27	0.95
Sirdah	27	25.82	1.00	0.90	0.63	1.00
Ehsani	15	11.63	1.00	0.90	0.63	1.00

RBC – Red blood cells; **RDW –** Red cell distribution width**; RDWI –** Red cell distribution width index; **MDHL** – Mean Density of Hemoglobin per Liter; **MCHD** – Mean Cell Hemoglobin Density **SENS** – Sensitivity; **SPEC** – Specificity; **PPV** – Positive predictive values; **NPV** – Negative predictive values; **YI** – Youden's index; **AUC** – Area under the curve.

Table S.40 – Accuracy findings for all calculated new best thresholds for female microcytic anemia subjects.

Indices	Female					
	Cutoff	Predicted Cutoff	SENS	SPEC	PPV	NPV
RBC	5	4.95	0.92	0.80	0.82	0.91
England and Fraser (E&F)	0	8.01	0.86	0.86	0.86	0.86
Mentzer	13	12.73	0.94	0.71	0.76	0.92
Srivastava	3.8	4.54	0.73	0.84	0.82	0.75
Shine and Lal (S&L)	1530	982.68	0.65	0.78	0.75	0.69
Bessman (RDW)	15	16.05	0.82	0.94	0.93	0.84
Ricerca	4.4	3.16	0.98	0.86	0.88	0.98
Jayabose (RDWI)	220	222.40	0.98	0.90	0.91	0.98
Green and King (G&K)	65	71.26	0.94	0.92	0.92	0.94
MDHL	1.63	1.60	0.96	0.78	0.82	0.95
MCHD	0.3045	0.3213	0.59	0.53	0.56	0.56
Sirdah	27	32.42	0.78	0.86	0.85	0.80
Ehsani	15	13.99	0.90	0.69	0.74	0.88

RBC – Red blood cells; **RDW –** Red cell distribution width**; RDWI –** Red cell distribution width index; **MDHL** – Mean Density of Hemoglobin per Liter; **MCHD** – Mean Cell Hemoglobin Density **SENS** – Sensitivity; **SPEC** – Specificity; **PPV** – Positive predictive values; **NPV** – Negative predictive values; **YI** – Youden's index; **AUC** – Area under the curve.