UNIVERSIDADE D
COIMBRA

Raquel da Silva Frias

Molecular characterization of Congenital Erythrocytosis and Idiopathic Erythrocytosis analysed by Next-Generation Sequencing

Dissertação no âmbito do Mestrado em Biologia Celular e Molecular, orientada pela Doutora Celeste Bento e pelo Doutor Licínio Manco e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Outubro de 2021

Faculdade de Ciências e Tecnologia da Universidade de Coimbra

MOLECULAR CHARACTERIZATION OF CONGENITAL ERYTHROCYTOSIS AND IDIOPATHIC ERYTHROCYTOSIS ANALYSED BY NEXT-GENERATION SEQUENCING

Raquel da Silva Frias

Dissertação no âmbito do Mestrado em Biologia Celular e Molecular, orientada pela Doutora Celeste Bento e pelo Doutor Licínio Manco e apresentada ao Departamento de Ciências da Vida da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Outubro de 2021

1 2 9 0

UNIVERSIDADE D COIMBRA

“You look at science (or at least talk of it) as some sort of demoralising invention of man, something apart from real life, and which must be cautiously guarded and kept separate from everyday existence. But science and everyday life cannot and should not be separated. Science, for me, gives a partial explanation for life. In so far as it goes, it is based on fact, experience and experiment.” Rosalind Franklin: “The Dark Lady of DNA".

In memoriam of victims of COVID-19 pandemic.

Agradecimentos

Agradeço à Doutora Celeste Bento, por me orientar e me introduzir ao mundo dos eritrócitos e das eritrocitoses. Obrigada pela oportunidade de desenvolver a minha tese de mestrado no Serviço de Hematologia Molecular do CHUC. Pela disponibilidade de orientação, toda paciência, todas as aulas dadas e confiança em meu trabalho. Obrigada, por me incentivar quando eu estava cansada e desacreditada; por todo carinho e amizade e a torcida que fez aos meus projetos pessoais. Sem dúvidas, construímos muito mais que uma tese.

Agradeço à Janete Pereira, que me conduziu e me orientou na parte prática do laboratório, me ensinando praticamente tudo que sei sobre as técnicas usadas neste trabalho. Gratidão por sua confiança, sua grande paciência, e por ser um grande exemplo para mim. Gratidão por sua amizade e carinho, me acolher no convívio de sua família e me ensinar sobre algumas tradições portuguesas e suas deliciosas comidas.

Agradeço ao Doutor Licínio Manco, que me recebeu no Laboratório de Populações e Doenças do CIAS, me ensinou sobre Epigenética e como analisar as metilações. Também por me orientar neste projeto e conduzir as correções, dedicando seu tempo e contribuindo com seu conhecimento científico para a finalização desta tese.

Agradeço a toda equipe do Laboratório de Hematologia Molecular, que de alguma forma contribuíram para este projeto. Em especial pelas amigas e companheiras de estágio: Jéssica Malva, Isabella Rocha, Ariana Mendes e Ema Neves. Gratidão por todo apoio, “helps” e toda partilha.

Agradeço à professora Ana Luísa Carvalho e ao professor Carlos Duarte de Almeida por me ajudarem com muitas questões neste mestrado. Sempre foram muito solícitos e queridos.

Ao meu amado Anderson, que embarcou de cabeça, corpo e alma neste projeto. Como ele mesmo diz: “O nosso sonho”. Atravessou o oceano junto com nossos dois gatinhos em meio ao caos da pandemia para estar ao meu lado neste último ano do mestrado. Todo seu amor, sua leveza de viver a vida, sua energia boa e seu apoio traduziram-se em força para que eu pudesse realizar esse trabalho. Mesmo quando estivemos separados fisicamente, o nosso amor nos manteve fortes e unidos. Amo você, meu amor! Gratidão por vocês (você, Bilu e Snow) existirem na minha vida.

Aos meus pais, Lacir e Mercedes, meus irmãos, Claiton, Meire, Roselaine e Gabriel, aos meus sobrinhos e familiares, amo muito vocês! Vocês me ensinam sobre valores e princípios, mostrando que com muita força de vontade podemos conquistar nossos sonhos e objetivos.

Em especial, agradeço à minha mãe, que é minha raiz, exemplo de força e meu eterno amor.

À minha irmã Meire e meu cunhado Sergio, por serem meus pilares, meus amores, meus companheiros de todos os momentos. Vocês são meus primeiros incentivadores, desde os meus primeiros anos na escola básica até aqui.

À minha sogra Carmen, meus cunhados João Fernando, Evelyn, César, Orávio, todos nossos sobrinhos, Josiane, tios e tias. Vocês são incríveis! Obrigada pelo acolhimento, orações, boas energias, mensagens de carinho e incentivo. Ou seja: Obrigada por tanto amor… Gratidão, amo vocês!

Ao meu primo Ariovaldo Veiga de Almeida, que me mostrou os caminhos de Portugal, através da sua admiração por este país e pela Universidade de Coimbra, plantou em nós o desejo de estar aqui. Com todo carinho e apoio, me incentivou na realização deste projeto: o tão sonhado "Mestrado em Biologia Celular e Molecular".

Ao Neimar, Fabiana Neiva, Théo e ao Pipoca, pela amizade de muitos anos, por estarem aqui em Portugal e ser nossa família do coração! Por sempre nos acolher tão bem e por me aguentar durante o primeiro lockdown: Meus dias de confinamento foram suportáveis por estar com vocês!

Um agradecimento especial à minha amiga Fabiana Neiva, pela parceria e por dedicar seu dom artístico no desenvolvimento da Figura 1 da introdução. Seu dom artístico é incrível!

Aos meus amigos que estão no Brasil, torcendo por mim, me incentivando e esperando para em breve estarmos juntos novamente. Em especial à Cynthia, minha amada amiga e sua família! Gratidão por todo incentivo e amor.

Às amigas que encontrei em Portugal, Ana, Gabi, Gi e Thaysa. Vocês estiveram ao meu lado nos momentos de alegria e de tristezas, me apoiando e me "suportando". Gratidão por trilharem comigo este caminho, por partilharmos muitos sentimentos, muitas risadas e algumas garrafas de vinho e sidra.

À toda malta do BCM: vocês foram fundamentais! Em Especial a Ana Filipa, Ana Rita, Carol, Francisca, Inês, Simão e Teresa.

Fazer um mestrado já é um grande desafio em qualquer lugar do mundo. Agora fazer um mestrado em outro país, que fala a mesma língua, mas ainda assim não perceber quase nada do que falam, além de ter aulas em inglês, foi ainda mais desafiador. Sentir saudades das pessoas que amo e nunca fiquei longe; tudo isso somado a viver os desafios de uma pandemia, as incertezas, os medos, a solidão… Foi mesmo um grande desafio! Por isso, faço aqui um agradecimento especial a mim, à Deus, à toda espiritualidade e as boas vibrações que me sustentam. Eu consegui chegar até aqui e me orgulho do trabalho que fiz. Aprendi muito sobre minhas limitações e capacidades nesta caminhada. Gratidão Raquel por "aguentar", por não perder a fé e manter viva a esperança de um mundo melhor!

Resumo

Eritrocitose, ou policitemia, é uma patologia caracterizada por um aumento significativo da massa eritrocitária (>125%), dos níveis de hemoglobina e hematócrito, em relação aos valores de referência, tendo em conta a idade e o sexo. A eritrocitose pode ser congénita (EC) ou adquirida. De acordo com os mecanismos fisiopatológicos e com base nos níveis de eritropoietina (Epo) é classificada como primária ou secundária. A eritrocitose primária é causada por um defeito intrínseco nos precursores dos eritrócitos e, portanto, independente dos níveis de Epo, sendo a forma congénita devida à presença de variantes patogénicas de ganho de função nos genes *EPOR* e *SH2B3* e a forma adquirida, Policitemia Vera, devida variantes patogénicas somáticas no gene *JAK2*. Estas variantes levam à ativação constante do recetor de eritropoietina após ser estimulado pela Epo. A EC secundária resulta de uma regulação positiva da transcrição do gene da Epo, *EPO*, a qual pode ser provocada por defeitos nos constituintes da via de deteção de oxigénio, molecularmente caracterizada por variantes patogénicas nos genes *VHL*, *EPO*, *EGLN1* e *EPAS1*, ou com origem em hemoglobinas com alta afinidade para o oxigénio, devido a variantes patogénicas nos genes globínicos (*HBB*, *HBA1*, *HBA2*) ou nos genes *BPGM* e *PKLR*. A eritrocitose secundária adquirida é devida a patologias ou fatores externos que induzem a hipoxia tecidual, fatores fisiológicos que induzem a diminuição no volume do plasma sanguíneo ou à presença de tumores que induzem um aumento de produção de Epo. Estão descritos onze genes onde variantes patogénicas de ganho ou perda de função foram associadas a EC (*EPOR, SH2B3, VHL*, *EPO*, *EGLN1*, *EPAS1, HBB*, *HBA1*, *HBA2, BPGM* e *PKLR)*. No entanto aproximadamente 60% dos pacientes continua sem uma etiologia molecular identificada, sendo designados como portadores de eritrocitose idiopática (EI). Estudos de Sequenciação de Nova Geração (NGS) são fundamentais para a identificação novas variantes nos genes já descritos ou em genes candidatos, que possam esclarecer a origem da patologia.

Métodos: Neste estudo foram analisadas 77 amostras de indivíduos com EI, acompanhados na Unidade de Eritropatologia e Metabolismo do Ferro, do Laboratório de Hematologia Molecular - Centro Hospitalar e Universitário de Coimbra. Os testes de laboratório foram orientados pela história clínica e familiar e pelos níveis de Epo, usando a sequenciação Sanger e NGS, com um painel de genes dedicados à EC, para encontrar variantes patogénicas que justifiquem o fenótipo apresentado**.**

Resultados e Discussão: O estudo realizado permitiu a identificação de variantes em 28 dos 77 indivíduos estudados. Em 5 pacientes, foram detetadas 4 variantes patogénicas que já estavam descritas como associadas a EC, nos genes *SH2B3*, *HBB* e *VHL,* Em 15 pacientes foram detetadas 13 variantes novas nos genes *EPOR*, *JAK2*, *SH2B3, EGLN1, EPAS1, EPO, PKLR* e *VHL*. Em 8 pacientes encontramos 6 novas variantes em genes candidatos, *EGLN2, EGLN3, HIF1α, HIF3α* e *PIEZO1*. O grau de patogenicidade das variantes encontradas foi avaliado pelo uso de ferramentas

in silico. Das 18 variantes analisadas: 6 foram classificadas como Patogénicas, 1 como Provável Patogénica e 4 como Variantes de Significado Indeterminado. Das restantes, 7 variantes foram classificadas como Benignas ou Prováveis Benignas.

Conclusão e Perspectivas Futuras: Este estudo permitiu a identificação de variantes patogénicas em 5 dos indivíduos estudados. Em 18 indivíduos foram detectadas variantes classificadas pelas ferramentas *in silico* como provavelmente patogénicas ou de significado indeterminado. Será necessário confirmar a patogenicidade destas variantes com estudos familiares e funcionais. A verificar-se serem variantes patogénicas, o estudo realizado terá contribuído para a identificação da causa molecular responsável pela EC em 11 do total de 18 amostras estudadas pelas ferramentas *in silico*. Estudos de sequenciação completa do exoma ou do genoma serão necessários para identificar novos genes que possam estar associados a EC e permitirem o diagnóstico dos restantes indivíduos que permanecem como IE.

Palavras-chave: eritrocitose congénita, eritrocitose idiopática, eritropoietina, hipóxia, Sanger, Sequenciação de Nova Geração (NGS).

Abstract

Erythrocytosis or polycythemia is a pathology characterized by a significant increase in erythrocyte mass (>125%) as well as an increase in hemoglobin and hematocrit levels to reference values according to age and sex. Erythrocytosis can be congenital (CE) or acquired. According to pathophysiological mechanisms and based on the levels of erythropoietin (Epo), they can be classified as primary or secondary. Primary erythrocytosis is caused by an intrinsic defect in the erythroid precursors and, therefore, independent of Epo levels, the congenital form is molecularly characterized by pathogenic variants in the *EPOR* and *SH2B3* genes and the acquired form, Polycythemia Vera, by somatic pathogenic variants in *JAK2* gene. Variants in these genes lead to constant activation of the erythropoietin receptor after being stimulated by Epo. Secondary CE results from up-regulation of EPO transcription, which can be caused by defects in the components of the oxygen-sensing pathway, molecularly characterized by pathogenic variants in the *VHL, EPO, EGLN1*, and *EPAS1*, or with origin in hemoglobin with high oxygen affinity, due to pathogenic variants in the globin genes (*HBB, HBA1, HBA2*), or in the *BPGM* and *PKLR*. Secondary acquired erythrocytosis results from external factors that induce tissue hypoxia, by a physiological factor that causes a decrease in blood plasma volume, or by tumors that induce higher Epo production. (*EPOR, SH2B3, VHL*, *EPO*, *EGLN1*, *EPAS1, HBB*, *HBA1*, *HBA2, BPGM* e *PKLR)*. However, approximately 60% of patients still do not have an identified molecular etiology, being designated as having idiopathic erythrocytosis (IE). Next-Generation Sequencing (NGS) studies are essential to identify new pathogenic variants in genes already described or in candidate genes that can clarify the origin of the pathology.

Design and Methodology: In this study, 77 samples of patients with IE, followed up at the Unidade de Eritropatologia e Metabolismo do Ferro, do Laboratório de Hematologia Molecular - Centro Hospitalar e Universitário de Coimbra, were analysed. Laboratory tests were guided by clinical and family history and Epo levels, using NGS and Sanger sequencing, with a NGS panel of genes dedicated to erythrocytosis to find pathogenic variations which justify the presented phenotype.

Results and Discussion: The study carried out allowed the identification of variants in 28 of the 77 individuals studied. In 5 patients, 4 pathogenic variants that were already described as associated with CE were detected in the *SH2B3, HBB,* and *VHL* genes. In 15 patients, 13 new variants in *EPOR, JAK2, SH2B3, EGLN1, EPAS1, EPO, PKLR,* and *VHL* genes were detected. In 8 patients we found 6 new variants in candidate genes, *EGLN2, EGLN3, HIF1α*, *HIF3α* and *PIEZO1*. The degree of pathogenicity of the variants found was evaluated using *in silico* tools. Of the 18 variants analysed: 6 were classified as Pathogenic, 1 as Likely Pathogenic and 4 as Variants of Uncertain Significance. The remaining 7 variants were classified as Benign or Likely Benign.

Conclusion: This study allowed the identification of pathogenic variants in 5 of the individuals studied. In 18 individuals, variants classified by the *in silico* tools as probably pathogenic or of

uncertain significance were detected. It will be necessary to confirm the pathogenicity of these variants with family and functional studies. When verifying that they are pathogenic variants, the study carried out allowed the identification of the molecular cause responsible for CE in 11 of the 18 samples studied by *in silico* tools. Complete exome or genome sequencing studies will be needed to identify new genes that may be associated with EC and allow diagnosis of the individuals who remain as IE.

Key Words: congenital erythrocytosis, idiopathic erythrocytosis, erythropoietin, hypoxia, Sanger, Next-Generation Sequencing (NGS)

Sumary

List of Abbreviations

ACMG- American College of Medical Genetics

BFU-E- Burst-forming unit erythroid

1,3-BPG- 1,3 bisphosphoglycerate

2,3-BPG- 2,3 bisphosphoglycerate

BPGM- Bisphosphoglycerate mutase

CE- Congenital Erythrocytosis

CFU-E- Colony-forming unit erythroid

CHA - Congenital hemolytic anemia

CO_2- Carbon dioxide

DNA- Deoxyribonucleic acid

ECYT1 - CE type 1

ECYT2 - CE type 2

ECYT3 - CE type 3

ECYT4 - CE type 4

ECYT5 - CE type 5

ECYT6 - CE type 6

ECYT7 - CE type 7

ECYT8 - CE type 8

EGLN1 - Egl-9 Family Hypoxia Inducible Factor 1

EGLN2 - Egl-9 Family Hypoxia Inducible Factor 2

EGLN3 - Egl-9 Family Hypoxia Inducible Factor 3

EPAS1 - Endothelial PAS Domain Protein 1

EPO - Erythropoietin gene

Epo - Erythropoietin protein

EPOR - Erythropoietin receptor

Hb - Hemoglobin

HBA1 - Hemoglobin Subunit Alpha 1

HBA2 - Hemoglobin Subunit Alpha 2

HBB - Hemoglobin Subunit Beta

Hb- Hemoglobin

Hct- Hematocrit

HFE - Homeostatic Iron Regulator

HIF - Hypoxia-Inducible Transcription Factors

HIF1α – Hypoxia-inducible factor 1

HIF2α – Hypoxia-inducible factor 2

HIF3α – Hypoxia-inducible factor 3

HIFβ - Hypoxia-inducible subunit β

HX - Hereditary xerocytosis

IE – Idiopathic erythrocytosis

JAK2- Janus kinase 2

JH1 - Jak homology domains **-** Catalytic Tyrosine kinase domain

JH2 - Jak homology domains **-** pseudo-kinase domain

LNK- Lymphocyte Adaptor Protein

MIM - Mendelian Inheritance in Man

MPN - Myeloproliferative neoplasms

NGS- Next-Generation Sequencing

O_2- Oxygen

OMIM - Online Mendelian Inheritance in Man

PCR- Polymerase Chain Reaction

PFCP- Primary Familial and Congenital Polycythemia

PIEZO1 - Piezo Type Mechanosensitive Ion Channel Component 1

PH domain- Pleckstrin homology domain

PHDs- Propyl hydroxylases

PHD2- Propyl hydroxylase domain-containing protein 2

PKLR - Pyruvate Kinase Liver/R

PI3K- Phosphatidylinositol 3-kinase

PV - Polycythemia Vera

RBC- Red blood cells

S1P- Sphingosine-1-Phosphate

SH2 domain- Src homology 2 domain

SH2B3 - SH2B Adaptor Protein 3

SHP-1- Src homology 2 domain-containing inositol polyphosphate 5-phosphatase-1

STAT5- Signal transducer activator of transcription 5

VHL - von Hippel Lindau protein

WGS - Whole Genome Sequencing

WHO- World Health Organization

List of Tables

Tables:

Supplementary Table:

List of Figures

1. Introduction

1.1. Background: Red Blood Cells – Functions, structure, and genesis.

Red blood cells (RBCs), also known as erythrocytes, constitute the most abundant cell type in the body, whose functions are to perform tissue oxygenation through blood flow, carbon dioxide (CO_2) transport to the elimination and blood buffering by acid-basic regulation [1]. These cells have a lipoprotein cell membrane, giving the cell a capacity for stability and structural breakdown, so they can easily deform and return to their usual shape, allowing greater mobility in the circulatory system and through the capillary network. Mature cells have a biconcave disk shape, without nucleus or organelles, which allows for a greater accumulation of hemoglobin (Hb). It is through Hb molecules that RBCs can transport oxygen (O_2) to tissues, allowing them to perform their main function [1,2].

During the first embryonic stage, the erythrocytes progenitor cells are produced in the yolk sac by a process called erythropoiesis. At six weeks of uterine life, erythrocytes start to be produced and matured mainly by the liver, and in smaller proportions by the spleen and lymph nodes. Furthermore, it is only at the third month of uterine life that this process gradually becomes carried out by the bone marrow, where they will start to be fully produced at the ninth month of gestation, after birth and throughout life. Until the age of five years, the erythrocytes are produced by all bones, from this age on the bone marrow of long bones, and by the age of twenty years they start to be produced only in marrow of membranous bones. An adult produces an average of approximately 2.4 million new RBCs every second [2,3].

The erythroid precursor can be divided into two strains termed burst-forming unit erythroid (BFU-E) and colony-forming unit erythroid (CFU-E). In a semi-solid culture medium, the human cell line BFU-E takes 14 days to form more mature CFU-E colonies which complete their cycle after 7 days giving rise to mature erythrocytes [2,4,3]. Erythropoiesis is effectively regulated by extracellular signals, such as erythropoietin (Epo) which is a hormone whose levels are the most significant regulator of RBCs production, other ones are interleukins and glucocorticoids [2,5].

Epo is a glycoprotein of the cytokine family, which production is performed mainly by kidney cells (90%), with a minor contribution of hepatocytes (10%). Its production is conducted by the oxygen-sensing pathway and regulated by a family of transcription factors called Hypoxia-Inducible Transcription Factors (HIF). HIF is a heterodimeric molecule composed by two subunits, HIFα subunit, which has three isoforms: HIF1α, HIF2α, HIF3α, and HIFβ subunit. Under hypoxia physiological conditions, HIF2α translocates to the cell nucleus, binding to HIFβ, inducing *EPO* transcription, and then Epo will be released into the blood circulation. Whereas in normoxia state,

HIF2α is hydroxylated by prolyl hydroxylase domain proteins (PHD2) and von Hippel-Lindau tumor suppressor (VHL) signals to the ubiquitin-proteasome complex promoting the degradation of the HIF2α molecule, performing a negative control for the Epo production [6,7,4,8].

Indeed, for proliferation and differentiation of the towards mature erythrocytes, Epo needs to bind to a specific receptor called Erythropoietin Receptor (Epor) located at the surface of the erythroblast. When this binding occurs, it triggers a cascade of phosphorylation by Janus kinase 2 (JAK2)/Signal transducer activator of transcription 5 (STAT5) pathway (JAK2/STAT5 pathway) [9,10,2,4,8]. There are precise homeostatic mechanisms to ensure sufficiently, but not excessively, RBC production (Figure 1). Any imbalance in those homeostatic mechanisms can lead to pathologies such as anemia or erythrocytosis [5,6].

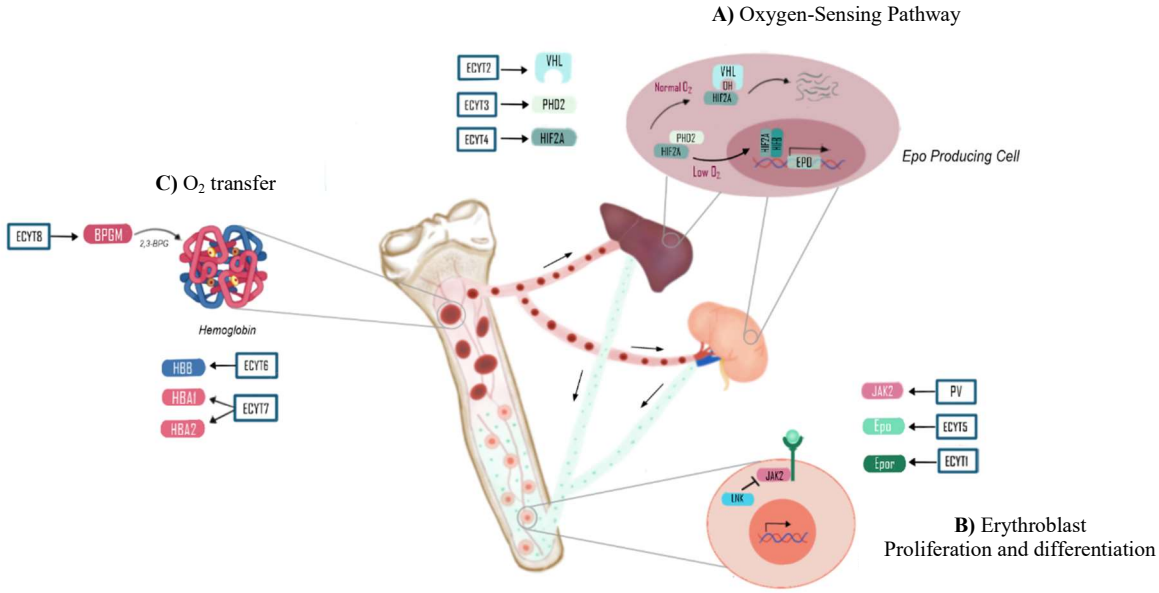

FIGURE 1: Schematic representation of the genes and mechanisms involved in Congenital Erythrocytosis (ECYT1-8) and Polycythemia Vera.
A) Oxygen sensing; **B)** Signaling for proliferation and differentiation of erythroblasts; **C)** Oxygen transfer **A)** Erythropoietin (Epo) production is performed by kidney cells and hepatocytes in response to hypoxia conditions. In normoxia conditions, the Hypoxia-Inducible Factor 2α (HIF-2α) is hydroxylated by Propyl Hydroxylase Domain-containing Protein 2 (PHD2) and binds the von Hippel Lindau protein (VHL) that induces its ubiquitination and degradation. In the absence of oxygen, hydroxylase activity of the PHD2 is reduced and HIF2α is stabilized, translocates into the nucleus and associates to HIF-β. The active HIF transcription factor induces expression of a number of target genes, including Epo. **B)** Epo is released in the circulation and is transported to the bone marrow, Epo binds its receptor, Epor, on the erythroblast cells. Epor phosphorylates Janus Kinase 2 (JAK2), which activates signaling cascade for proliferation and differentiation of erythroblasts to Red Blood Cells (LNK is an inhibitor of JAK2) and **C)** The RBC reestablish tissue oxygenation through of the blood flow. Hemoglobin alpha (HBA1 and HBA2) and beta (HBB) compound tetramer protein responsible for transporting O_2. The regulation of hemoglobin oxygen affinity system is controlled by endogenous heterotropic ligands, such as 2,3-bisphosphoglycerate (2,3-BGP) The BPGM is an enzyme that catalyzes 1,3-bisphosphoglycerate (1,3-BPG) to 2,3-BPG.

1.2. Erythrocytosis

Erythrocytosis, also called polycythemia, is a rare disease characterized by an increase in the erythrocytes mass. This condition was discovered 150 years ago by the French physician and physiologist Denis Jourdanet. He studied the effects of high altitude and low partial pressure of oxygen affecting people's health, comparing the blood of people living at high altitudes with the blood of people living at sea level. In this study, he found that the blood of people living at higher altitudes is more viscous than that of people living at lower altitudes [6,11]. Blood hyperviscosity, which is a consequence of increased red cell mass and elevated Hb, can result in a variety of nonspecific and moderate clinical features, including headache, fatigue, paraesthesia, blurred vision, vertigo, red suffused eyes, redness of the hands, and arterial hypertension. However, critical and potentially life-threatening situations can also occur, such as acute myocardial infarction, stroke, or peripheral thromboembolic phenomena [6,12]. In specific situations, routine phlebotomies should be prescribed to prevent complications and alleviate symptoms caused by blood hyperviscosity [13,12,14,15,5].

The World Health Organization (WHO) classifies erythrocytosis as a pathology characterized by an increase in the hematocrit (Hct) and the hemoglobin value above the reference values for sex, age, and living at high altitude, which associates with an increase in the erythrocytes mass in the peripheral blood. WHO's classification defines erythrocytosis as follows: elevated mass of RBCs (>125%) for both sex, males with Hb values greater than 18.5 g/dL or Hct greater than 52 percent, females with Hb values greater than 16.5 g/dL or Hct greater than 48 percent, observed in separate blood counts performed at different time points [5,16]. There are three different forms of erythrocytosis, namely relative, apparent and absolute. Relative erythrocytosis occurs due to a decrease in blood plasma associated with the use of diuretics, severe diarrhea and/or vomiting. Apparent erythrocytosis is related to arterial hypoxia associated with smoking, exposure to carbon monoxide and sleep apnea. Finally, absolute erythrocytosis is a pathology that can be classified as primary or secondary, with acquired or congenital origin (Figure 2) [17,18,16].

<table>
<tr><td rowspan="3">ABSOLUTE ERYTHROCYTOSIS</td><td rowspan="2">Primary
Intrinsic erythroblast defect
(Low or normal Epo)</td><td>Congenital</td><td>• Pathogenic variant genes:
EPOR
SH2B3</td></tr>
<tr><td>Acquired
Polycythemia Vera</td><td>• Pathogenic variant genes:
JAK2</td></tr>
<tr><td rowspan="2">Secondary
Extrinsic erythroblast defect
(Elevated or normal Epo)</td><td>Congenital</td><td>• Oxygen sensing pathway
Pathogenic variant genes:
VHL, EPAS1, EGLN1, EPO
• High oxygen affinity
Pathogenic variant genes:
HBA1, HBA2, HBB
BPMG, PKLR</td></tr>
<tr><td>Acquired</td><td>• Hypoxia driven
• Local hypoxia
• Epo secreting tumors
• Exogenous Epo</td></tr>
</table>

FIGURE 2: Classification of Absolute Erythrocytosis and associated genes.
Epo: erythropoietin.

Erythrocytosis can be congenital (CE) or acquired, according to pathophysiological mechanisms, and based on erythropoietin (Epo) levels can be classified as primary or secondary. The main feature of primary erythrocytosis is normal or low Epo level due to molecular defects in *JAK2*, *EPOR*, and *SH2B* Adaptor Protein 3 (*SH2B3*), leading to changes in EPOR functionality [9,19,20]. The acquired erythrocytosis or classical Polycythemia Vera (PV) is a myeloproliferative disorder characterized by somatic pathogenic variants in *JAK2* gene, with V617F and exon 12 variants being the most frequent and responsible for 96% of the cases described. PV can manifest with thrombotic events, the main cause of morbidity and mortality. In fact, about one third of the patients with PV have a thrombotic event during their lifetime, and younger patients have an increased risk of early death from cardiovascular disease in comparison with the general population, representing 45% of all deaths [21,12,6,22]. In order to diagnose the classical acquired form, it is important to perform a bone marrow biopsy, allowing to confirm hypercellularity and panmyelosis, and, additionally, a positive test for a *JAK2* variants [23–25]. Congenital Erythrocytosis Type 1 (ECYT1) or Primary Familial and Congenital Polycythemia (PFCP) is a rare proliferative disease of HSCs, with a dominant transmission pattern. Patients with ECYT1 have usually a strong family history of erythrocytosis and are negative for *JAK2* V617F and exon 12 variants. When suspecting

of acquired congenital erythrocytosis, *EPOR* should be the first gene to be studied. However, if no pathogenic variants is found in *EPOR*, other genes as *SH2B3* or *JAK2* should be analysed [12,5,20].

Secondary erythrocytosis differs from primary CE because it doesn't occur due to defects in erythroblasts, but rather in the Epo production pathway, also known as the oxygen sensing pathway, being characterized by higher levels of Epo or inadequate normal levels [6]. Secondary erythrocytosis can also be congenital or acquired. The acquired is form is the most frequent and is due to some pathologies, such as cardiac and pulmonary deficiencies, renal and hepatic diseases, erythropoietin-secreting tumors, androgen administration, or high-altitude disease. Congenital secondary erythrocytosis is rare, 40% of the cases result from inherited genetic variants with autosomal recessive or dominant transmission patterns [17,6,8].

Pathogenic variants in genes that participate in the oxygen-sensing pathway responsible for producing Epo in response to the state of hypoxia, such as *VHL*, Egl-9 Family Hypoxia Inducible Factor 1 (*EGLN1*), Endothelial PAS Domain Protein 1 (*EPAS1*), and *EPO* are well described for causing secondary CE. High oxygen affinity Hbs causes tissue hypoxia and promotes increased production of red blood cells, therefore Hb variants with high oxygen affinity, due to pathogenic variants in the beta (β) or alpha (α) globin (*HBB*, *HBA*) genes or due to a 2,3 bisphosphoglycerate (2,3-BPG) deficiency, which results from pathogenic variants in the *BPGM*, can also cause secondary CE [17,18,5,26].

Patients with CE caused by high affinity hemoglobin are usually asymptomatic. Because of this, it is crucial to distinguish it from the other forms of secondary erythrocytosis related to the oxygen-sensing pathway. In the first case, the non-delivery of oxygen to the tissues generates a state of hypoxia and increases the RBC production, therefore the problem is not the excess of RBC, but the non-delivery of oxygen to the tissues. The prescription of phlebotomies can be necessary but may be dangerous for the patient's health because, with frequent phlebotomies, the delivery of oxygen to the tissues will be even more deficient, which may lead to functional anemia, thereupon, needs to be taken into consideration before any suspicion [5,15,12].

1.3. Molecular Mechanisms – Erythrocytosis

1.3.1. Polycythemia Vera

The specific variants at *JAK2* (OMIM:147796 - Table1) are associated with PV. This gene is located on chromosome 9 at position p24.1, has 4 transcripts, and the transcript JAK2-201 (ENST00000381652.4 - NM_004972.4) has 7023 base pairs, 1132 amino acids, and 25 exons [27]. The JAK2_human (UniProt: O60674) has six important domains (Figure 3) [28]. Two N-terminal FERM domain which are located at FERM_F1 (PF18379) p.39 - p.134 and FERM_F2 (PF18377) p.142 - p.261. This domain is structurally similar to a Pleckstrin homology domain (Jak1_Phl domain - PF17887) and consequently is capable of binding to both peptides and phospholipids at different sites [28–30]. The Jak1_Phl, located at p.299–p.381 that is activated in response to cytokines and interferons. The SH2 domain (PF00017), located at p.400-p.481 functions as a regulatory module of intracellular signaling phosphorylation cascades [31]. The two Jak homology domains (JH1 and JH2 - PF07714) are important for downstream signaling. The pseudo-kinase domain (JH2) located at p.545–p.805, lacks catalytic activity, and it is involved in the auto-inhibition of JAK2 [32]. The C-terminal JH1 is structurally similar to JH2, it is located at p.849-p.1123, and is responsible for phosphorylation [28,30,33]. The JAK2 protein promotes the growth and division, thereby the proliferation of cells. This protein participates in the JAK/STAT5 signaling pathway, which transmits chemical signals from outside the cell to the cell's nucleus. *JAK2* pathogenic variants result in the production of a constitutively activated JAK2 protein, which seems to improve the survival of the cell and increases the production of RBCs [6,34].

The molecular basis of PV consists in somatic variants with a gain-of-function in *JAK2* from the erythroid progenitors. The most common is a missense variant (c.1849G>T) in exon 14, with the replacement of a valine amino acid with a phenylalanine at position 617 in the protein (p.V617F). There are also rare cases of pathogenic variants in exon 12. These variants lead to the inactivation of the pseudokinase domain, leading to constitutive enzymatic activity and intracellular signal transduction, and resulting in unlimited cellular proliferation [22,35,36].

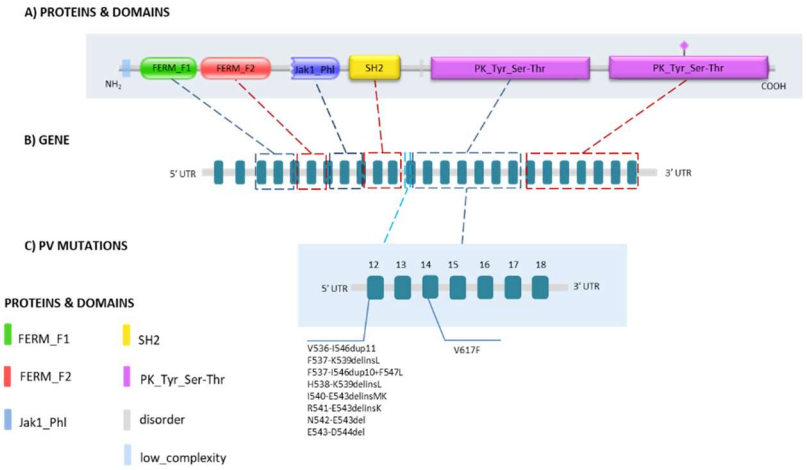

FIGURE 3: Schematic illustration of *JAK2* gene and protein structure.
A) Tyrosine-protein kinase JAK2 protein (JAK2_HUMAN) has different features and 6 domains. FERM1 domain; FERM2 domain; Jak1_Phl domain; SH2 domain; and two PK_Tyr_Ser-Thr domain [28,30]. Adapted from Pfam- EMBL-EBI, 2021 [22,37]. **B)** *JAK2* gene has 25 exons [27]. **C)** Location of pathogenic variants associated with PV: exons 12 and 14.

1.3.2. Congenital Erythrocytosis Type 1 (ECYT1)

ECYT1 is characterized by an autosomal dominant or *de novo* variants in *EPOR*. The *EPOR* (OMIM:133171 – Table1) is located on chromosome 19 at position p13.2, has 8 transcripts. The transcript EPOR-201 (ENST00000222139.11 - NM_000121.4) has 2411 base pairs, 508 amino acids, and 8 exon [38,39]. The EPOR_Human protein (UniProt: P19235) contains two functional domains: the Erythropoietin receptor - ligand binding domain (EpoR_lig-bind - PF09067), located at p.37–p.140, and the Fibronectin type III domain (fn3 - PF00041), located at p.146-p.232 (Figure 4) [28,30]. The main function of EPOR is activate the erythroblast to proliferation and differentiation pathway to originate mature erythrocytes, after EPO binding [40].

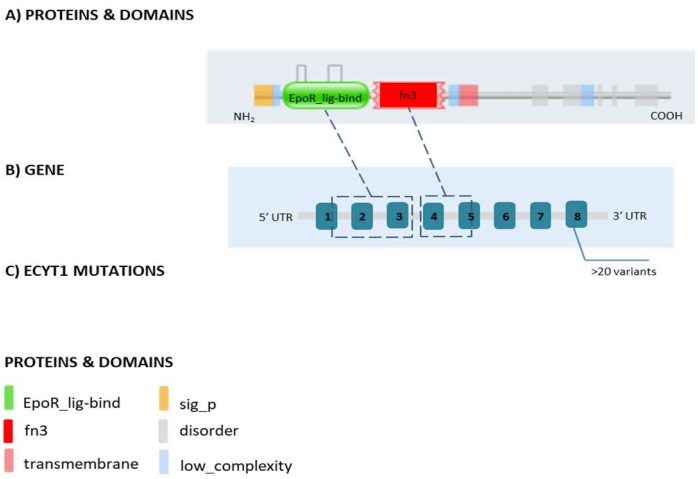

FIGURE 4: Schematic illustration of *EPOR* gene and protein structure.
A) Erythropoietin Receptor protein (EPOR_HUMAN) has different features and 2 domains. erythropoietin receptor - ligand binding domain; Fibronectin type III domain. Adapted from Pfam-EMBL-EBI [30]. **B)** *EPOR* gene has 8 exons [39]. **C)** Location of variants associated with ECYT1: exon 8 [5].

More than 20 variants associated with ECYT1 have been reported in this gene (Supplementary Table 1). These variants lead to a truncated receptor and lack of an important negative regulatory region in the cytoplasmic domain of EPOR, responsible for recruiting the Src homology 2 domain-containing inositol polyphosphate 5-phosphatase-1 (SHP-1), which is involved in cell signaling and is a negative regulator of the JAK/STAT pathway. *EPOR* pathogenic variants lead to premature stop codon with loss of the SHP-1 docking site. Once Epo binds to the receptor, it is turned on but not turned off again and therefore continues to drive red blood cell production without further Epo stimulation [9,10,41,42].

When *EPOR* or *JAK2* pathogenic variants are not found, and in accordance with clinical patterns associated with primary congenital erythrocytosis, *SH2B3* variants must be evaluated. This protein, also known as Lymphocyte Adaptor Protein (LNK), participates in the JAK/STAT pathway, and is encoded by the *SH2B3* (OMIM: 605093 – Table1), located on chromosome 12 at position q24.12. LNK contains three functional domains: Phenylalanine zipper domain, followed by central Pleckstrin homology (PH) and SH2 domains. The SH2 domain mediates most of the biological effects of LNK through interactions with its phosphorylated signaling partners, including EPOR and JAK2. The SH2 domain of LNK is essential for the inhibitory function of LNK, and the PH domain also plays a significant role in the inhibitory function, but is not essential [24,28,30,31,43–45].

LNK acts as a negative regulator of JAK2 via the EPOR/JAK2/STAT5 pathway. When the JAK/STAT pathway is active, LNK adaptor protein expression is induced as a biological response. Consequently, LNK inhibits JAK2/STAT signaling through a negative feedback loop [22,46]. *SH2B3* pathogenic variants are more commonly associated with acquired myeloproliferative neoplasms, but rarely are also associated with CE, where the variants result in a defective LNK that does not act as a negative regulator of the JAK/STAT pathway leading to increased erythropoiesis. Five variations in *SH2B3* gene are described associated with erythrocytosis (Supplementary Table 7), with an autosomal dominant transmission or occurring *de novo* [19,43,47].

1.3.3. Congenital Erythrocytosis Type 2 (ECYT2)

ECYT2 is molecularly characterized by homozygous, or compound heterozygous pathogenic variants in *VHL* (OMIM:608537 - Table 1), which is located on chromosome 3 at position p25.3, has three transcripts. The VHL_201 transcript (ENST00000256474.3 - NM_000551.4) has 4414 base pairs, 213 amino acids, and 3 exons [48]. It commonly encodes 2 proteins called pVHL213, pVHL160 and in an alternative splice VHLΔE2 encodes the pVHL172 protein [49–51,51]. The VHL_Human protein (UniProt: P40337) contains two functional domains: β-domain (PF01847), located at p.63–p.144, and VHL box domain or α-domain (PF17211), located

at p.156-p.204 (Figure 5) [28,30,52]. The main function of VHL is to act as a tumor suppressor, with the pVHL213 and pVHL160 proteins acting as signaling proteins in the ubiquitin-proteasome degradation process [49,53]. Its main target is the hydroxylated HIF2α molecule, thus performing upregulation in the production of erythropoietin and other genes which are transcribed by this pathway [54,55].

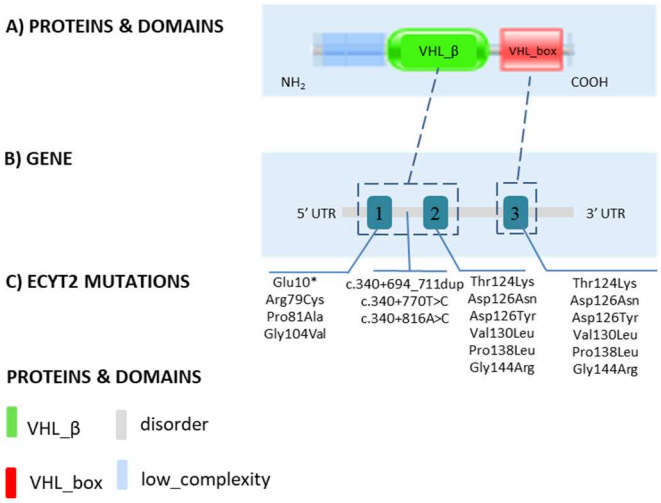

FIGURE 5: Schematic illustration of *VHL* gene and protein structure.
A) von Hippel–Lindau tumor suppressor (VHL_HUMAN) has different features and 2 domains. VHL beta domain and VHL box domain. Adapted from Pfam- EMBL-EBI [28,30]. **B)** *VHL* gene has 3 exons [30,48]. Location of pathogenic variants associated with ECYT2: distributed in between 3 exons, and in intronic region 1 [5,53]. *Stop Codon.

Pathogenic variants in the *VHL* are related to several disease phenotypes, such as different types of cancers, von Hippel-Lindau disease, and ECYT2 [54,56]. Within ECYT2 about 18 variants in *VHL* have been reported (Supplementary Table 2), with an autosomal recessive genetic pattern; however, there are some clinical cases where the family genetic study demonstrates unexplained heterozygous patterns [5]. The first variants found in *VHL* associated with CE, were described in the autonomous Republic of Russia and the disease was called Chuvash polycythemia, molecularly characterized by a homozygous missense, *VHL*: c.598C>T (p.Arg200Trp). The mutant protein has impaired interaction with HIF-2α and reduced ubiquitination under normoxic conditions, thereby promoting an increase in Epo levels. The other *VHL* variants lead to CE by the same mechanism [57,58].

In a recent study carried out by Lenglet and coworkers [53] a new exon was discovered, where a deep-intronic region was found between E1 and E2 a possible result of an alternative splice. The new cryptic exon was named E1', which combined with E1 encodes a 193aa protein called X1 (114aa from E1 and 79aa E1'). In this study, 12 independent families were studied and variants in E1' appeared in 8 of them. They found the variants in different patterns of genetic inheritance: compound heterozygosity (c.429C>T/c.340+770T>C, c.598C>T/c.340+770T>C, c.598C>T/c.340+694_711dup, c.430G>A/c.340+694_711dup), and in homozygosity (c.340+816A>C) [53,59]. The discovery of these variants in E1' allowed new genetic studies within

the group of people diagnosed with idiopathic erythrocytosis, where the possible identification of these pathogenic variants patterns could allow a definitive diagnosis.

1.3.4. Congenital Erythrocytosis Type 3 (ECYT3)

ECYT3 is a disease with an autosomal dominant inheritance pattern, which is molecularly characterized by heterozygous loss-of-function variants in the *EGLN1* gene (OMIM:606425 - Table 1). This gene is located on chromosome 1 at position q42.2 and, has 7 transcripts. The EGLN1_201 transcript (ENST00000366641.4 - NM_022051.3) has 4335 base pairs, 426 amino acids, and 5 exons, that encodes the PHD2 protein [60]. The PHD2 mainly acts as an oxygen sensor, catalyzing the prolyl hydroxylation of HIF-2α using oxygen as a substrate and directing HIF2α to degradation, allowing the subsequent VHL signaling to the ubiquitin-proteasome complex. However, when oxygen levels are below normal (hypoxia), the PHD2 enzyme becomes less active, HIF stabilizes and translocate to the cell nucleus, triggering the transcription of several genes involved in the production of red blood cells and blood vessels [5,7,56,61]. EGLN1_HUMAN protein or PHD2 (UniProt: Q9GZT9), contains two functional domains: MYND zinc finger domain (zf-MYND), located at p.21-p.58 and 2OG-Fe(II) oxygenase superfamily domain, located at p.298-p.391 (Figure 6) [30]. The PHD also has more two distinct isoforms: PHD1 and PHD3, which are encoded by *EGLN2* and *EGLN3* genes, respectively, all 3 isoforms efficiently hydroxylate HIFα subunits under normoxia [62].

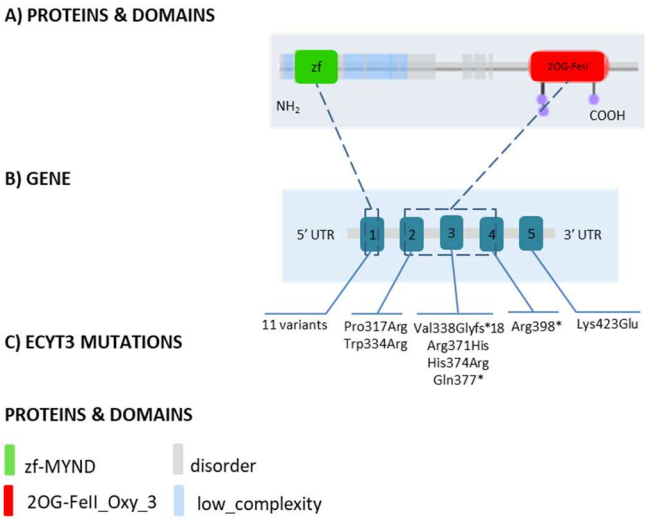

FIGURE 6: Schematic illustration of *EGLN1* gene and protein structure.
A) Egl nine homolog 1 (EGLN1_HUMAN) has different features and 2 domains. zf-MYND and 2OG-FeII_Oxy_3. Adapted from Pfam- EMBL-EBI [28,30]. **B)** *EGLN1* gene has 5 exons [48]. **C)** Location of pathogenic variants associated with ECYT3: Variants are distributed among all exons, but exon 1 has more variants than the other exons [5,53]. *Stop codon.

Currently, 19 pathogenic variants in *EGLN1* are described and related to ECYT3 (Supplementary Table 3), where functional studies have shown that the mutant protein has significantly reduced enzyme activity and consequently deregulated Epo production [5,63–65]. The first pathogenic variant in *EGLN1* associated with CE was described in 2006 by Percy *et al.* [63] being characterized as heterozygous conversion *EGLN1* : c.950C>G. This variant results in the replacement of p.Pro317Arg in a highly conserved region of the protein and leads to significantly reduced enzyme activity [5,63]. A new pathogenic variant was recently described by Bonnin *et al.* [66] in a patient who presented with moderate erythrocytosis with normal serum erythropoietin, reported as a heterozygous nonsense in exon 1 of *EGLN1*, which had never been described. This *EGLN1*: c.400C>T (p. Gln134*) is responsible for creating a premature stop codon in the mRNA thereby producing a truncated protein that has lost all of its catalytic domain [66].

1.3.5. Congenital Erythrocytosis Type 4 (ECYT4)

ECYT4 is molecularly characterized by gain-of-function variants in *EPAS1* (OMIM:603349 - Table 1). The *EPAS1* gene is located on chromosome 2 at position p21, has 10 transcripts. The EPAS1_201 transcript (ENST00000263734.5 - NM_001430.5) has 5155 base pairs, 870 amino acids, 16 exons, and encodes the transcription factor HIF2α [67]. HIF2α or EPAS1_Human (UniProt: Q99814) has four functional domains (Figure 7): PAS fold domain (PF00989), located at p.85–p.183; PAS_3 (PF08447), located at p.254–p.341; Hypoxia-inducible factors-1 (HIF-1- PF11413), located at p.517–p.549; HIF-1 alpha C terminal transactivation domain (HIF-1α_Ct - PF08778), located at p.833–p.869 (Figure 7) [28,30].

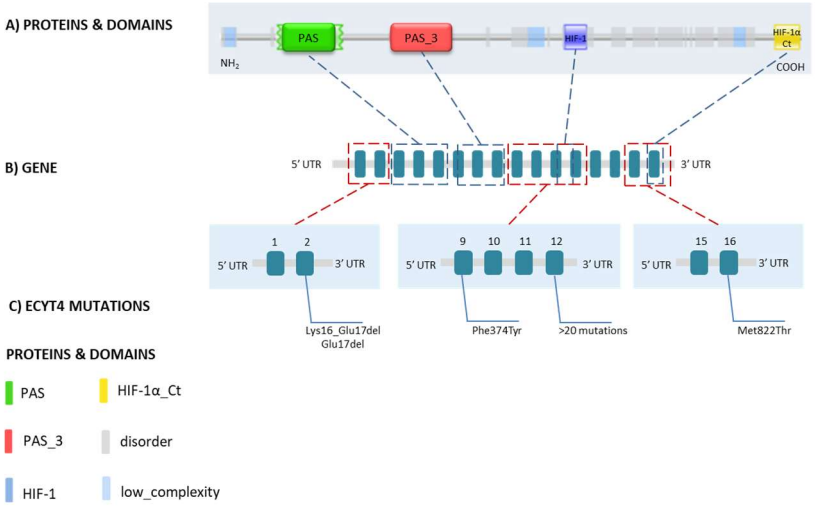

FIGURE 7: Schematic illustration of *EPAS1* gene and protein structure.
A) Endothelial PAS domain-containing protein 1 (EPAS1_HUMAN) has different features and 4 domains. PAS fold domain; PAS_3, Hypoxia-inducible factors (HIF-1); HIF-1 alpha C terminal transactivation domain (HIF-1α_Ct). Adapted from Pfam- EMBL-EBI [28,30]. B) *EPAS1* gene has 16 exons [48]. Location of pathogenic variants associated with ECYT4: distributed in exons 2, 9, 12 (majority), and 16 [5,53].

The family of transcription factors HIF, inducing *EPO* gene expression and transcription, also activates the transcription of other important genes that promote the synthesis of hemoglobin, heme and the absorption of iron [68,6]. It is widely expressed in mammals and performs a key role in hypoxia pathway allowing the adaptation of mammals to high altitude environments. Also, the low level of Hb in Tibetan populations is correlated with several single nucleotide polymorphisms (SNPs) found in *EPAS1* [69].

The first *EPAS1* pathogenic variants found in erythrocytosis patients were the pathogenic missense p.Gly537Trp, p.Gly537Arg, p.Met535Val, and p.Pro534Leu [5,70,71]. At least 19 pathogenic variants in the *EPAS1* with autosomal-dominant inheritance have been identified as responsible for causing ECYT4 (Supplementary Table 4). However, in a recent review, Kristan *et al.*[72] retrieved 20 variants in *EPAS1* correlated to CE. All variants are heterozygous, the majority of them located in exon 12, on the C-terminal oxygen dependent degradation domain (ODD). ODD is the target region for regulation at the post-translational level by the hydroxylase enzyme PHD2 and by the tumor suppressor protein VHL, so it is an important domain for the stability and activity of the HIF2α protein [6,72,73].

1.3.6. Congenital Erythrocytosis Type 5 (ECYT5)

ECYT5 is molecularly characterized by a heterozygous pathogenic variants in the *EPO* gene (OMIM: 133170 - Table 1). The *EPO* gene is located on chromosome 7 at position q22.1, has 1 transcript with 1662 base pairs, 193 amino acids, 5 exons, and encodes the Erythropoietin protein (Epo_Human - UniProt P01588). This protein has 1 functional domain: the erythropoietin/thrombopoietin domain (EPO_TPO - PF00758), located at p.31-p.191 (Figure 8).

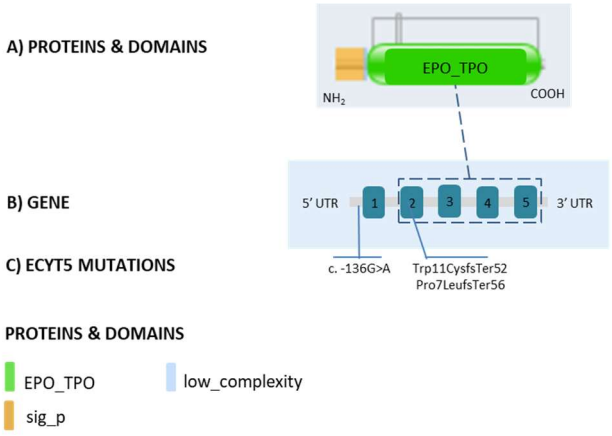

FIGURE 8: Schematic illustration of *EPO* gene and protein structure.
A) Erythropoietin protein (Epo_HUMAN) has different features and 1 domain. erythropoietin/thrombopoietin domain (EPO_TPO); [28,30]. Adapted from Pfam- EMBL-EBI, 2021 [28,30]. B) *EPO* gene has 5 exons [74]. Location of pathogenic variants associated with ECYT5: distributed in 5' UTR and exon 2 [5].

As mentioned before, Epo is a cytokine produced by kidney cells through the hypoxia pathway. Once Epo is released into the bloodstream, it activates the JAK2/STAT5 pathway through EPOR, promoting the differentiation and proliferation of erythrocytes; in other words, Epo controls erythropoiesis [2,4,75,3]. Only three pathogenic variants in the *EPO* have been described to cause CE (Supplementary Table 5). The first pathogenic variant linking *EPO* and CE was described in 2015 by Taylor *et al*. [26] in a study, the multidisease WGS500 Program whole-genome sequencing. The authors described the finding of a variation in the 5'UTR of the *EPO* (c.-136G>A) in to independent families with erythrocytosis, establishing a cosegregation with the disease [56,76,77].

A functional study with *EPO* variants was recently performed by Zmajkovic *et al*. [78], where patients diagnosed with IE revealed a heterozygous single-base deletion in exon 2 of *EPO*, c.32delG and c.19delC variants. The c.32delG creates a frameshift that truncates the erythropoietin and generates a novel peptide, terminating after an additional 51 amino acids, which would predict a loss-of-erythropoietin-function. However, a study for functional characterization of the mutated *EPO* was performed using CRISPR to introduce c.32delG variant into *EPO* in Hep3B cells, a human hepatoma cell line. It showed that assayed erythropoietin in the supernatants of these clonal cell lines was capable of stimulating the growth of an erythropoietin dependent cell line, therefore, concluding that the c.32delG deletion is, in fact, a gain-of-function variant. The same study was performed with the *EPO* c.19delC variant, where the long P2 and short P2 transcripts with the c.19delC also produced high amounts of erythropoietin that was biologically active. In conclusion, both deletions cause a frameshift and convert a noncoding mRNA transcribed from an alternative (P2) promoter inside *EPO* intron 1 into an mRNA that produces functional erythropoietin protein. The investigation of patients with idiopathic erythrocytosis performed by Camps *et al*. [73] found the same *EPO* c.19delC variant in a child with familial history, corroborating the relationship between *EPO* pathogenic variant and CE [76,78,47].

1.3.7. Congenital Erythrocytosis Type 6-8 (ECYT 6-8) and Hb with increased O2 affinity

ECYT 6-8 is characterized by autosomal dominant transmission in the *beta globin* gene (*HBB* - OMIM: 141900 – ECYT6) and *alpha-globin* genes (*HBA1*- OMIM: 141800 or *HBA2*- OMIM: 141850 – ECYT7) or by autosomal recessive transmission in Bisphosphoglycerate Mutase gene (*BPGM*- OMIM: 613896) or autosomal dominant transmission in Pyruvate Kinase, Liver and RBC gene (*PKLR*).

- High oxygen-affinity Hb (ECYT6 - ECYT7)

HB encode alpha and beta chains of hemoglobin and are located on chromosome 16 at position p13.3 and chromosome 11 at position p15.4, respectively (Table 1). The hemoglobin

molecules are globular and tetramer proteins, formed by four polypeptide chains, the globins. Each globin has a protein unit called heme, that is linked to an iron atom, a region of affinity for oxygen molecules and may be divided into two α-subunits (α1 and α2) and two β-subunits (β1 and β2). The two αβ dimers, also named α1β1 and α2β2, are arranged around a 2-fold axis of symmetry resulting in a large central water cavity. This cavity has different conformations when it's in tense (T) state, which exhibits low affinity for O_2, or relax (R) state, which exhibits high affinity for O_2. This system is controlled by endogenous heterotropic ligands, such as 2,3-bisphosphoglycerate (2,3-BGP) [79–81]. Since the first described molecular defect associated with CE, over a hundred of Hb pathogenic variants with high affinity for O_2 has been described. The more frequent pathogenic variants involve the *HBB* gene and are predominantly inherited, with only a few cases of *de novo* variants [5,80,82].

- *BPGM* variants - ECYT8

The bisphosphoglycerate mutase gene (*BPGM*- OMIM: 613896 – ECYT8) is located on chromosome 7 at position 33 (Table 1) and encodes the multifunctional enzyme that catalyzes 1,3-bisphosphoglycerate (1,3-BPG) to 2,3- bisphosphoglycerate (2,3-BPG). 2,3-BPG function is to promote changes in the conformation of hemoglobin, binding to Hb and promoting the binding of Sphingosine-1-Phosphate (S1P). This sphingolipid is known to regulate diverse biological processes as an important signaling molecule. In this process, S1P only binds to the surface of Hb upon binding of 2,3-BPG at the β-cleft and this process orientates the heterodimer R state to T state, promoting the regulation of the affinity of hemoglobin for oxygen and allows the efficient unloading of O_2 into tissues [42,79].

Only 5 pathogenic variants in *BPGM* correlated with CE have been described (Supplementary Table 6): a compound heterozygosity for a c.268C>T missense variant (p.Arg90Cys) and a deletion c.61delC (p.Arg21Valfs*28), and homozygosity for a missense variant c.185G>A (p.Arg62Gln) [6]. Petousi and collaborators [26] identified a novel heterozygous missense variant in the *BPGM*, c.269G>A that predicts the substitution of arginine at residue 90 with histidine (R90H). The functional effects of the variant were confirmed by showing significantly lower levels of BPGM and its product 2,3-BPG in red blood cells [26]. The last *BPMG* pathogenic variant was described in the literature by Camps [73], who developed a Next-Generation Sequencing (NGS) panel directed to idiopathic erythrocytosis and applied it to 125 patients with idiopathic erythrocytosis. Among these patients, one carried a new heterozygous missense variant c.304C>A (p.Gln102Lys) [73]. In cases, when pathogenic variants in *HB* or *BPGM* are not found, variants in *PKLR* should be analysed.

PKLR (OMIM: 609712) encodes the protein pyruvate kinase that is responsible for catalyzes the transphosphorylation of phosphoenolpyruvate to pyruvate and ATP synthesis, which are the end products of the second phase of glycolysis. An autosomal dominant variant in *PKLR* (c.110G>A- p.Gly37Glu) can lead to CE by PK hyperactivity with high ATP levels, recruitment of

a large amount of 1,3-BPG, to the glycolytic pathway, subsequent deficiency 2,3-BPG and a consequent increase in HB affinity for oxygen (Table 1) [83,84].

1.4. New Candidate Genes associate with IE.

- Piezo-type mechanosensitive ion channel component 1 gene (*PIEZO1*)

Filser and collaborators [85] described variants in a new candidate gene that can be associated with CE, the *PIEZO1*. They identified them in patients who presented with elevated Hct and Hb concentrations [85]. This gene encodes a transmembrane non-selective cationic channel, responsible for the modification of the lipidic bilayer in response to stress and to an increase in the membrane tension, allowing the erythrocytes to cross the microcapillaries. Pathogenic variants in this gene can lead to congenital hemolytic anemia (CHA) know as Hereditary xerocytosis (HX) a rare red blood cell membrane disorder. The *PIEZO1* variants associated with Hereditary xerocytosis are the gain-of-function ones, and are responsible for the abnormal activation of the Gardos channel, enabling a massive loss of potassium and water by the cell, and leading it to hemolysis [85–88].

In this research, Filser *et al.* [85] showed that 4% of the population investigated with idiopathic erythrocytosis carry pathogenic alterations in the *PIEZO1*. They also found that the venous P50 measurement was frequently low (<25 mmHg), which indicates an increase in Hb oxygen affinity. The induction of hypoxia leads to an intense glycolytic activity in order to maintain high ATP concentration required for cell homeostasis. Glycolysis then leads to a decrease in 2,3-DPG concentration, associated with low P50 and erythrocytosis. This was previously reported in HX patients with mutated *PIEZO1*, and recently confirmed by Kiger *et al*. [89] in the metabolome study of HX-RBC. They showed how *PIEZO1* HX affects red cell glycolysis and leads to increased Hb oxygen affinity. These two combined studies suggest that some cases of erythrocytosis are related to *PIEZO1* variants and could be considered in the diagnosis of idiopathic erythrocytosis. The recent association of variants in *PIEZO1* with CE consists of four pathogenic variants: p. Arg2110Trp, p. Thr2127Met, and p. Ala2020Thr previously reported in HX and a new *PIEZO1* variant, p. Ile2462Met, in a heterozygous state. They also found one more variant that was classified as uncertain significance, p.Val1223Ile [85,89].

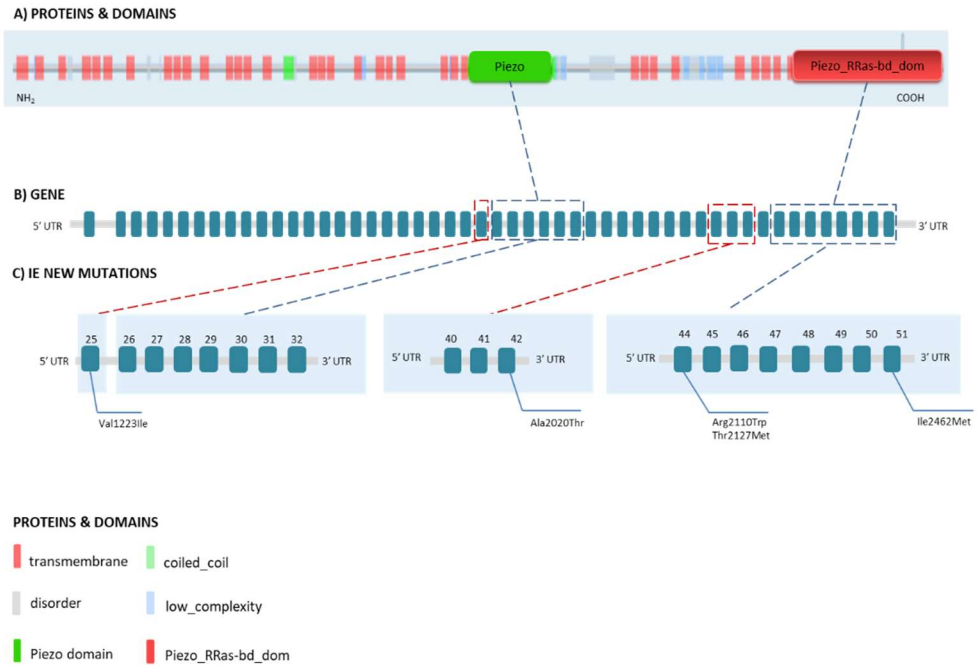

FIGURE 9: Illustration of *PIEZO1* gene and protein structure.
A) Mechanosensitive non-specific cation channel protein (PIEZ1_HUMAN - PF15917) has different features and 2 domains. Piezo domain: located p.1235 - p.1465; Piezo non-specific cation channel 1, R-Ras-binding domain (Piezo_RRas-bd_dom - PF12166) located at p.2111 - p.2519 - UniProt: Q92508 [28,30]. **B)** *PIEZO1* has 51 exons [90]. **C)** Location of new pathogenic variants that can be associated with IE, found by Filser *et al* exons 25, 42, 44, 51 [85].

- Homeostatic Iron Regulator gene (*HFE*)

Pathogenic missense variants in the Homeostatic Iron Regulator gene (HFE) also are frequently found in studies of patients with erythrocytosis, the most frequently are *HFE* c.187C>G (p.His63Asp) and c.845G>A (p.Cys282Tyr) usually in a heterozygous state. Although these variants have a very high frequency in the population (MAF: 0.073) [91], this association between *HFE* variants and erythrocytosis has drawn the attention of researchers. Since 1979, Raphael and his collaborators have described how patients with Hereditary hemochromatosis (HH) could develop erythrocytosis [92]. Recently, this association has been widely discussed among groups studying blood cell diseases, until now, there are no functional studies that prove the relationship between HFE variants with the development of erythrocytosis [93–96]. HH is a syndrome of dysregulated iron homeostasis resulting in the excessive deposition of iron in parenchymal cells. Defects in genes encoding *HFE* are major causes of HH, and the most common variants are C282Y and H63D both with autosomal recessive inheritance [95,94,93,96,91].

1.5. Idiopathic erythrocytosis

With the advent of genomics and with novel genetic panels aiming to investigate erythrocytosis, Next-Generation Sequencing has proven to be largely effective as a tool for diagnoses previously given as idiopathic causes of erythrocytosis. In fact, despite a large number of pathogenic variants already described in association to CE, approximately 60% of cases have no identified cause and are called idiopathic erythrocytosis (IE). This fact relates to patients who have different genetic patterns, such as unexplained heterozygotes, and also patients with family history without a known pathogenic variants. Non-diagnosis is a problem because the therapeutic options for a patient with IE are limited and largely unproven to be effective [34,41,73,97].

1.6. NGS Studies

The Next-Generation Sequencing (NGS) study is one of the main and most used tools for diagnostic because it allows the sequencing of several genes, analysing not only the exons but also non-coding regions, such as the promoter region, the introns, and the UTR regions. With the possibility to study a large number of samples simultaneously and carry out a personalized approach for each type of disease with exclusive design panels, this methodology is faster, more sensitive, with high performance, and cost-effective [98,99].

Nowadays, with novel genetic panels investigating erythrocytosis, NGS has proven to be largely effective as a tool for diagnosis of previously given as idiopathic causes. There are still many questions about the difficulties in providing a diagnosis to patients with erythrocytosis, and the vast majority is diagnosed as idiopathic erythrocytosis, but still, it's the better technique to use. Other candidate genes can be studied dynamically, as genes related with the hypoxia pathway *EGLN2* (PHD1), *EGLN3* (PHD3), *HIF1α*, *HIF3α*, in addition to *PKLR*, *SH2B3*, *HFE* and *PIEZO1* among others correlated with hematological diseases. Recent studies performed with NGS have included findings in genes that are associated with erythrocytosis and also added new genes to this list [72,73,85,100,101]. The big question is perhaps not about the technique used, but about which new genes and new pathways we should explore to provide a diagnosis for IE patients.

TABLE 1: Classification of erythrocytosis and the mutant proteins that cause the phenotype.

Gene OMIM[1]	Cytogenetic location	Protein	Mutant protein effects	Inheritance	Phenotype MIM[2]
JAK2 147796	chr9p24.1	Janus Kinase 2 (JAK2)	Gain of function - Impaired auto-inhibitory domain	Somatic variants	PV 133100
EPOR 133171	chr19p13.2	Erythropoietin Receptor (EPOR)	Gain of function - Increase activation of signaling cascade	Autosomal dominant	ECYT1 133100
SH2B3 605093	chr12q24.12	Lymphocyte Adaptor Protein (LNK)	Loss of function - Impaired inhibitory domain of JAK2/STAT	Autosomal dominant, *de novo,* Somatic variants	133100
VHL 608537	chr3p25.3	Von Hippel-Lindau Tumor Suppressor (VHL)	Loss of function - Impaired degradation of HIF2α	Autosomal recessive	ECYT2 263400
EGLN1 606425	chr1q42.2	Prolyl hydro-xylase domain containing protein 2 (PHD2)	Loss of function - Impaired degradation of HIF2α	Autosomal dominant	ECYT3 609820
EPAS1 603349	chr2p21	Hypoxia-Inducible Factor 2 Alpha (HIF2 α)	Gain of function - increase stability in normoxia	Autosomal dominant	ECYT4 611783
EPO 133170	chr7q22.1	Erythropoietin (Epo)	Gain of function - increase expression in normoxia	Autosomal dominant	ECYT5 617907
HBB 141900	chr11p15.4	Hemoglobin Subunit Beta (HBB)	Loss of function - increases Hb affinity for O_2	Autosomal dominant	ECYT6 617980
HBA1 141800	chr16p13.3	Hemoglobin Subunit Alpha 1 (HBA1)	Loss of function - increases Hb affinity for O_2	Autosomal dominant	ECYT7 617981
HBA2 141850	chr16p13.3	Hemoglobin Subunit Alpha 2 (HBA2)	Loss of function - increases Hb affinity for O_2	Autosomal dominant	ECYT7 617981
BPGM 613896	chr7q33	Bisphosphoglycer ate Mutase (BPGM)	Loss of function - impaired synthesis of 2,3-BPG	Autosomal recessive	ECYT8 222800
PKLR 609712	chr1q22	Pyruvate Kinase, Liver And RBC (PKLR)	Gain of function PK hyperactivity high ATP, reduce 2,3-BPG levels, increases Hb affinity for O2	Autosomal dominant	102900

[1]Online Mendelian Inheritance in Man, [2]Mendelian Inheritance in Man: numerical assignment for inherited diseases.

2. Aims

The present study is focused on molecular pathophysiology of CE. The aim was to contribute to the elucidation of the molecular mechanisms of congenital erythrocytosis through the molecular characterization of a group of samples from patients with IE followed in Laboratório de Hematologia Molecular - Centro Hospitalar e Universitário de Coimbra, during the last two years.

From this main goal several specific objectives were considered:

- To collect and characterize a group of samples from patients with IE.
- To search for new variants and/or new candidate genes involved in the pathophysiology of this disease using the NGS technology.
- To predict the pathogenicity of the identified variants by NGS using *in silico* tools.

3. Design and methods

3.1. Cohort

A group of 77 DNA samples of individuals of European origin was recruited in the Laboratório de Hematologia Molecular - Centro Hospitalar e Universitário de Coimbra (CHUC) for genetic evaluation for erythrocytosis. The study was approved by the CHUC Ethical Committees and all subjects or representatives gave their informed consent. Patients in this study included children and adults (aged 1 to 88 years-old), men (n = 61) and women (n = 16). For all patients the acquired causes of erythrocytosis were excluded. Diagnosis was performed based on clinical and family history, Epo levels, and HPLC suggestive of a hemoglobin variant.

3.2. Techniques

3.2.1. HPLC

The presence of hemoglobin variants was searched by High-Performance Liquid Chromatography (HPLC) through a VARIANT II™, with the β-Thalassemia Short Program (Bio-Rad Laboratories, Hercules, CA, USA). The samples with an abnormal variant in the chromatogram were sequenced by Sanger sequencing (genes *HBB*, *HBA1 and HBA2*).

3.2.2. DNA Extractions

- Whole blood

Genomic DNA was extracted from whole blood collected in EDTA-containing tubes, by automatic isolation with the DNA extraction kit QIAsymphony® DSP DNA Mini Kit using QIAsymphony SP (QIAGEN) extractor, according to manufacturer's instructions.

- Hair follicle

Genomic DNA was extracted from hair follicle by manual procedure with the DNA extraction kit QIAamp® Tissue Kit (QIAGEN) according to manufacturer's instructions.

3.2.3. NGS Study

To identify causal variants in these patients, a custom NGS gene panel was developed (Table 2). This panel includes the exons, promoter regions and intronic flanking regions of the 11 genes previously described as associated with CE - including the *VHL* E1' - as well as of the newly

described genes and candidate genes [53,73]. This panel was designed to include also the genes associated with hemolytic anemia (HA), as some of them are common with CE. This has the advantage of analyzing new variants found in genes that are associated with HA, thus looking at all genes associated with erithropathology.

TABLE 2: Customized NGS gene panel.

Congenital Erythrocytosis		Common for both panels	Hemolytic anemia						
BPGM NM_199186.2	*HIF1A* NM_001530.3	*HBA1* NM_000558.4	*ABCB7* NM_004299.5	*ANK1* NM_000037.3	*FECH* NM_000140.3	*GPX1* NM_000581.3	*HSPA9* NM_004134.6	*PGK1* NM_000291.3	*SLC4A1* NM_000342.3
EGLN1 NM_022051.2	*HIF3A* NM_0152795.4	*HBA2* NM_000517.4	*ABCG5* NM_022436.2	*ATP11C* NM_001010986.2	*G6PD* NM_001042351.2	*GSR* NM_000637.4	*KCNN4* NM_002250.2	*PIGA* NM_002641.3	*SPTA1* NM_003126.3
EGLN2 NM_080732.3	*JAK2* NM_004972.3	*HBB* NM_000518.4	*ABCG8* NM_022437.2	*C15ORF41* NM_001130010.2	*GATA1* NM_002049.3	*GSS* NM_000178.3	*KIF23* NM_138555.3	*PIGT* NM_015937.5	*SPTB* NM_001355436.1
EGLN3 NM_022073.4	*SH2B3* NM_005475.2	*HFE* NM_000410.3	*ALAS2* NM_000 032.4	*CDAN1* NM_138477.3	*GCLC* NM_001498.3	*GYPA* NM_002099.7	*KLF1* NM_006563.4	*RHAG* NM_000324.2	*TPI1* NM_000365.5
EPAS1 (HIF2A) NM_001430.4	*VHL* NM_000551.3	*PKLR* NM_000298.5	*ALDOA* NM_000 034.3	*EPB41* NM_004437.3	*GLRX5* NM_016417.2	*GYPC* NM_002101.4	*LPIN2* NM_014646.2	*SEC23B* NM_006363.5	*UGT1A1* NM_000463.2
EPO NM_000799.3	*VHL* E1' Chr3:10142600 - 10143400	*PIEZO1* NM_001142864.3	*AK1* NM_000476.2	*EPB42* NM_000119.2	*GPI* NM_000175.4	*HK1* NM_000188.2	*NT5C3A* NM_016489.12	*SLC25A38* NM_017875.3	*UROS* NM_000375.2
EPOR NM_000121.3							*PFKM* NM_000289.5	*SLC2A1* NM_006516.2	*XK* NM_021083.3

Genes associate and candidate genes for CE (red), genes that are common in both diseases (purple) and genes associated with HA (blue). NM ref seq by Ensembl

- Library preparation and NGS

The sequencing libraries were prepared using the AmpliSeq Library Kit 2.0, and pooled barcoded libraries were clonally amplified by Ion OneTouch2 both following the manufacturer's protocol (Thermo Fisher Scientific). The enriched template-positive particles were loaded onto an Ion 316 or 318 chip and sequenced using Ion Hi-Q Sequencing 200 Kit chemistry on an Ion Torrent PGM sequencing system (Thermo Fisher Scientific).

The NGS results were analyzed using Ion Torrent Suite™ (v3.6; Thermo Fisher Scientific) and using T-MAP (version 3.6.58977) to align the sequences to the human genome version 19 (hg19). Variants were called using the Torrent Variant Caller, with Germ Line – Low Stringency settings (version 3.6.59049). These were annotated using Ion Reporter ™ (Thermo Fisher Scientific). The minimum depth of coverage required for each nucleotide in the ROI to be identified/considered a candidate variant was ≥ 100x. The mean coverage was highly variable; however, the mean for all the ROIs was at least 500-fold in all patient [102].

3.2.4. Sanger Sequencing process

The Sanger sequencing was performed to confirm significant variants found in NGS, familial inheritance of these variants, and to characterize Hb variants found by HPLC.

❖ PCR Amplification

PCR was performed in order to obtain a high number of copies of fragments of specific regions of genes to increase the amount of DNA for the next step, the Sanger sequencing procedure. The specific primers used in this study, previously described in the literature [18] or designed in the Laboratory of Molecular Hematology (CHUC), are detailed in Supplementary Table 8. For each gene, different PCR reactions were made and for all PCR reactions 100 ng/µL DNA were used. The thermocycler used in the PCR reactions was the BioRad C1000™ Thermal Cycler and the reagents were from QIAGEN®. PCR reactions and programs to specific primers for the genes under study are detailed in Supplementary Table 9.

❖ Agarose Gel Electrophoresis

The amplification of DNA by PCR was confirmed in a 2% agarose gel. This gel was prepared with 1g of agarose mixed in 50 mL 0.5% TBE (Tris 1.78 M; EDTA 0.04 M; Boric Acid 1.77 M). The solution was heated until the agarose was completely dissolved, and then 4 µL of SYBR™ Safe was added. 4 µL of each PCR sample was loaded into the gel, mixed with a drop of loading dye. Finally, the gel was run at 120 V for 30 minutes. At the end of the run, DNA fragments were visualized with UV light and photographed.

❖ Sanger Sequencing

The Sanger sequencing technique was made using the three steps: i) EXOSAP® purification; ii) sequencing reaction, and iii) resin column purification. After the EXOSAP® purification of PCR products were submitted to the sequencing reactions using the BigDye® Terminator v1.1 Cycler Sequencing kit (Applied Biosystems, Carlsbad, CA, USA) and custom primers to sequence the region of interest. After that, the non-incorporated BigDye® was eliminated by applying the sequencing sample to a resin column. Finally, the collected product was loaded into a sequencing plate and analyzed on the sequencer Applied Biosystems™ SeqStudio™ Genetic Analyzer.

The obtained sequences were read with the software Chromas version 2.6.6 (Technelysium Pty Ltd, South Brisbane, Queensland, Australia, 2018), and aligned with Clustal Omega-EMBL-EBI [103–105]. The reference sequence for the alignment was obtained from Ensembl Genome online database - EMBL-EBI GRCh38.p13, all RefSeq are describe in Table 2 [104,106].

3.3. Bioinformatic analysis

To analyze the results obtained in the NGS, the criteria used was the inclusion of variants with MAF ≤ 0.01 and coverage >100. All variants that resulted from these filters were searched in the Varsome database, Human Gene Mutation Database (HGMD) and in the ClinVar database. The results were categorized and grouped into Benign, Likely Benign (L. Benign), Variant of Uncertain Significance (VUS), Likely Pathogenic (L. Pathogenic) and Pathogenic [107] and the ratings obtained in these databases [108–110]. These results were divided into three groups of samples:

Group 1: samples with pathogenic variants in genes that were already described as causing the CE disease.

Group 2: samples with variants in genes associated with CE pathways, but not yet described; this group was divided into two subgroups: 2A) Primary erythrocytosis genes; 2B) Secondary erythrocytosis genes.

Group 3: samples with variants in candidate genes not yet correlated with CE.

The obtained results are summarized in tables 4-6 in the Results section. New variants were analysed by *in silico* tools to classify its pathogenicity.

❖ *In silico* analysis

In silico analysis was performed to missense variants using six tools that predict the possible consequences of the amino acid changes on the structure and function of the proteins using straightforward physical and comparative considerations. The impact of missense changes was analyzed using: PROVEAN, SIFT, PolyPhen-2, MutationAssessor, MutationTaster, and MetaDome, followed by complementary comparative analyzes in the Varsome, Clinvar and HGMD databases. The variant classification was executed follow established pattern:

***In Silico* tools**	**Verdict**
4 Pathogenic	Pathogenic
3 Pathogenic	Likely Pathogenic
2 Pathogenic	VUS
1 Pathogenic	Likely Benign
0 Pathogenic	Benign
NA = Not Applicable	

- PROVEAN, predictions for a list of genome variants. It is based on the assembly of the species and the Ensembl genome annotation [111].

- SIFT, which predicts whether an amino acid substitution affects protein function based on the degree of conservation of amino acid residues in a sequence alignment derived from closely related sequences [111,112].
- PolyPhen-2, a tool that predicts possible consequences of an amino acid substitution on the structure and function of human proteins [113].
- MutationAssessor, the functional impact is assessed based on evolutionary conservation of the affected amino acid in protein homologs [114].
- MutationTaster, employs a Bayes classifier to eventually predict the disease potential of an alteration. The Bayes classifier is fed with the outcome of all tests and the features of the alterations and calculates probabilities for the alteration to be either a disease variants or a harmless polymorphism [115].
- MetaDome, analyses the variant tolerance at each position in a human protein using data from gnomAD and ClinVar. MetaDome enhances the analysis of the gene of interest by parallel analysis of all homologous domains in the whole human genome [116].

3.4. Variants Classification

According to the American College of Medical Genetics (ACMG) standards and guidelines for interpreting variants, are recommended that a description of the process for classifying the variants be carried out in the three categories: i) based on criteria that use typical types of evidence for variants, (*e.g.*: population data, computational data, functional data); ii) interpretation of results by a certified laboratory and iii) certified clinical molecular geneticist or molecular genetic pathologist or equivalent [107].

4. Results

4.1.Cohort

A total of 77 DNA samples from individuals of European origin was selected in the Laboratório de Hematologia Molecular - Centro Hospitalar e Universitário de Coimbra for a genetic evaluation of erythrocytosis. Patients in this study included children and adults with an age median of 45 years, where the youngest is 1 year old and the oldest is 88 years old. There are 16 women and 61 men. They are not related, except samples N41 and N69 that are mother and daughter. The hematological pattern medians of these samples are Hb 17.6 g/dL , Hct 51.35% , RBC mass 5.86 x10^12/L , and Epo 9.85 mUI/mL (Table 3).

TABLE 3: Clinical features and hematological patterns of the patients with erythrocytosis.

Sex	Total of Patients	Median of Age[1]	Hb* (g/dL)	Hct* (%)	RBC* x10^12/L	Epo* mUI/mL
Female	16	41 [6-88]	16.55 [14.2-18.9]	50.10 [40.7-53.4]	5.87 [4.9-6.2]	8.07 [2.5-17]
Male	61	45 [1-86]	17.60 [15.3-20]	52.00 [44.3-60.4]	5.84 [5-6.9]	10.10 [4.3-85.7]
Total	77	45 [1-88]	17.60 [14.2-20]	51.35 [40.7-60.4]	5.86 [4.9-6.9]	9.65 [2.5-85.7]

[1]Current age; *Median of current hematological patterns [minimum and maximum values].

4.2.Molecular study of the patients with erythrocytosis

A genetic study was performed by a dedicated NGS panel and Sanger sequencing in a total of 77 samples. 28 samples presented variants with MAF ≤0.01 and coverage greater than 100x, this pool was used in this work, and represents 37% of the total. Unfortunately, until now it was not possible to perform functional or familiar studies in some of these patients to understand the inheritance of the variants. The other 49 samples, representing 63% of total, were annotated for further study, as some of them had variants with MAF ≤ 0.01 in deep intronic regions or apparently non-pathogenic variants. These patients remain as unknown cause (IE) as shown in Figure 10.

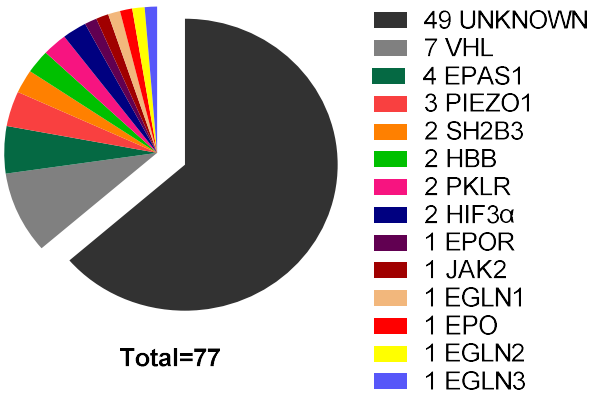

FIGURE 10: Graphic representation of genes and variants found in the 77 patients studied. 49 patients remain as unknown (IE). 4 pathogenic variants to 5 patients and 19 new variants distributed among genes associated with PV and CE and new candidate genes to 23 patients. [Variant criteria: MAF ≤ 0.01 and coverage ≥ 100x.]

The results were grouped into three groups:

Group 1: pathogenic variants already described and associated with CE. A pool with 5 samples and 4 variants. These variants were found in genes: *SH2B3* (c.622G>C), *HBB* (c.392A>G, c.294C>G), and *VHL* (c.340 + 816A>C), as shown in Table 4.

Group 2: variants found in genes already associated with CE that were not yet classified as causing CE disease (Table 5). A pool with 15 samples and 13 new variants. This group was divided into 2 subgroups: 2A) 3 variants found in different genes associated with primary erythrocytosis: *EPOR* (c.296C>T), *JAK2* (c.2681A>G), and *SH2B3* (c.1393G>A); 2B) 10 variants found in different genes associated with secondary erythrocytosis: *EGLN1* (c.1072C>A), *EPAS1* (c.181A>G, c.587C>T, c.1104G>A), *PKLR* (c.1614A>T), *VHL* (c.340+648T>C, c.28G>A, c.74C>T, c.154G>T, c.241C>T).

Group 3: new variants found in candidate genes and that are not yet described as causing CE (Table 6). A pool with 8 samples and 6 variants. These variants were found in genes: *EGLN2* (c.266G>A), *EGLN3* (c.613A>G), *HIF1*α (c.1762G>A), *HIF3*α (c.1247G>A, c.1896C>A), *PIEZO1* (c.6829C>A).

TABLE 4: Group 1. Genetic pathogenic variants already described as associated with congenital erythrocytosis.

Group 1							
ID[1]	Sex	Age[2]	Gene	Nucleotide Change	Amino Acid Change	Zygosity	Reference
N6	F	65	*SH2B3*	c.622G>C	Glu208Gln	Htz	Spolverini *et al.*, 2013
N31	F	61	*HBB*	c.392A>G	Tyr131Cys	Htz	Kister *et al.*, 2005
N16	F	11		c.294C>G	His98Gln	Htz	Lorkin *et al.*, 1970
N60	M	86	*VHL*	c.340+816A>C	-	Hmz	Lenglet *et al.*, 2018
N3	M	46				Hmz	

[1]ID: sample/Patient identification; [2]Age at diagnosis. Htz = Heterozygous; Hmz = Homozygous.

TABLE 5: Group 2. New variants found in genes associated with (A) Primary Erythrocytosis, (B) Secondary Erythrocytosis.

Group 2A							
ID[1]	Sex	Age[2]	Gene	Nucleotide Change	Amino Acid Change	Zygosity	Varsome Verdict[3]
N29	M	39	*EPOR*	c.296C>T	Ala99Val	Htz	Benign
N55	F	8	*JAK2*	c.2681A>G	Asp894Gly	Htz	VUS
N73	M	76	*SH2B3*	c.1393G>A	Val465Ile	Htz	VUS
Group 2B							
ID[1]	Sex	Age[2]	Gene	Nucleotide Change	Amino Acid Change	Zygosity	Varsome Verdict[3]
N20	F	88	*EGLN1*	c.1072C>A	Pro358Thr	Htz	VUS
N56	M	18	*EPAS1*	c.181A>G	Ile61Val	Htz	Benign
N77	F	72		c.587C>T	Thr196Met	Htz	L. Benign
N41	F	59		c.1104G>A	Met368Ile	Htz	Benign
N69	F	31				Htz	
N13	M	58	*PKLR*	c.1614A>T	Glu538Asp	Htz	VUS
N18	M	53				Htz	
N48	M	56	*VHL*	c.340+648T>C	Intronic	Htz	Benign
N23	M	57		c.28G>A	Glu10Lys	Htz	VUS
N27	M	71		c.74C>T	Pro25Leu	Htz	Benign
N66	M	51		c.154G>T	Glu52*	Htz	Pathogenic
N10	M	76		c.241C>T	Pro81Ser	Htz	Pathogenic

[1]ID: sample/Patient identification; [2]Age at diagnosis; [3]Varsome Classification of variants on Human genome (hg38) [108]. Htz = Heterozygous; Hmz = Homozygous; *Stop codon.

TABLE 6: Group 3. New variants found in candidate genes in the hypoxia pathway.

Group 3							
ID[1]	Sex	Age[2]	Gene	Nucleotide Change	Amino Acid Change	Zygosity	Varsome Verdict[3]
N32	M	69	*EGLN2*	c.266G>A	Arg89Gln	Htz	L. Benign
N51	M	1	*EGLN3*	c.613A>G	Arg205Gly	Htz	L. Pathogenic
N45	M	74	*HIF1α*	c.1762G>A	Ala588Thr	Htz	L. Benign
N64	M	82	*HIF3α*	c.1247G>A	Arg416His	Htz	L. Benign
N66	M	51		c.1896C>A	Asn632Lys	Htz	L. Benign
N30	M	29	*PIEZO1*	c.6829C>A	Leu2277Met	Htz	VUS
N35	M	64				Htz	
N74	M	24				Htz	

[1]ID: sample/Patient identification; [2]Age at diagnosis; [3]Varsome Classification of variants on Human genome (hg38) [108]. Htz = Heterozygous; Hmz = Homozygous.

The results obtained with *in silico* tools, predicting the effect of the new missense variants found in the NGS study, are detailed in Table 7.

TABLE 7: *In silico* analysis of the new variants found by NGS study.

Gene	Protein effect	*In Silico* tools							
		MetaDome	SIFT	PROVEAN	PolyPhen-2	Mutation Taster	Mutation Assessor	*In silico* Verdict	Varsome Verdict
EGLN1 (PHD2)	Pro358Thr	Intolerant	Damaging	Deleterious	Probably Damaging	Deleterious	High	Pathogenic	VUS
EGLN2 (PHD1)	Arg89Gln	Slightly intolerant	Tolerated	Neutral	Benign	Benign	Neutral	L. Benign	L. Benign
EGLN3 (PHD3)	Arg205Gly	Intolerant	Damaging	Deleterious	Probably Damaging	Deleterious	High	Pathogenic	L. Pathogenic
EPAS1 (HIF2α)	Ile61Val	Intolerant	Damaging	Neutral	Probably Damaging	Deleterious	Medium	Pathogenic	Benign
EPAS1 (HIF2α)	Thr196Met	Intolerant	Damaging	Deleterious	Probably Damaging	Deleterious	Medium	Pathogenic	L. Benign
EPAS1 (HIF2α)	Met368Ile	Intolerant	Tolerated	Neutral	Benign	Deleterious	Medium	VUS	Benign
EPOR	Ala99Val	Intolerant	Tolerated	Neutral	Benign	Benign	Low	L. Benign	Benign
HIF1α	Ala588Thr	Intolerant	Tolerated	Neutral	Benign	Benign	Neutral	L. Benign	L. Benign
HIF3α	Arg416His	Highly tolerant	Damaging	Neutral	Probably Damaging	Benign	Low	VUS	L. Benign
HIF3α	Asn632Lys	Neutral	Tolerated	Neutral	Benign	Benign	NA	Benign	L. Benign

Table 7- *Cont.*

JAK2	Asp894Gly	Intolerant	Tolerated	Deleterious	Probably Damaging	Deleterious	Low	Pathogenic	VUS
PIEZO1	Leu2277Met	Slightly intolerant	Damaging	Neutral	Probably Damaging	Deleterious	Medium	L. Pathogenic	VUS
PKLR	Glu538Asp	Slightly intolerant	Tolerated	Neutral	Benign	Deleterious	Neutral	VUS	VUS
SH2B3	Val465Ile	Slightly intolerant	Tolerated	Neutral	Benign	Benign	Medium	L. Benign	VUS
VHL	Glu10Lys	Highly tolerant	NA	Neutral	Possibly Damaging	Benign	Neutral	L. Benign	VUS
VHL	Pro25Leu	Tolerant	NA	Neutral	Benign	Benign	Neutral	Benign	Benign
VHL	Glu52*	Intolerant	NA	NA	NA	Deleterious	NA	VUS	Pathogenic
VHL	Pro81Ser	Intolerant	Damaging	Deleterious	Probably Damaging	Benign	Low	Pathogenic	Pathogenic

4.2.1. Group 1

- *SH2B3* (LNK) variants

Patient N6, a 65-year-old female. The molecular study was performed by NGS and we found a missense variant in exon 2, *SH2B3:*c.622G>C, in heterozygous state, causing the amino acid substitution Glu208Gln [44] (Table 4). The presence of this variant was confirmed by Sanger sequencing in DNA isolated from the hair follicle, to eliminate the possibility of a somatic variant. The germline origin was confirmed by the presence in both cell lines, blood and hair follicle, as shown in Figure 5.

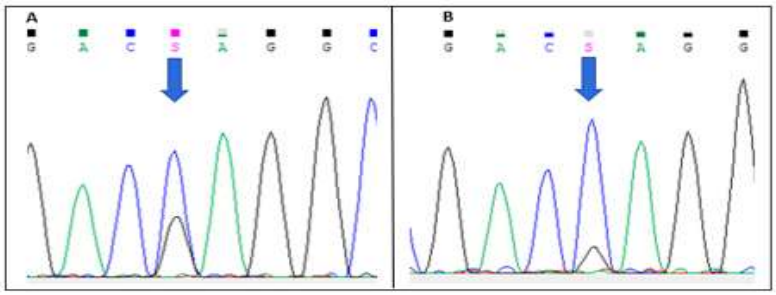

FIGURE 11: *SH2B3* gene pathogenic variant in Patient N6.
Identification of a variant in exon 2 of SH2B3 c.622G>C with amino acid substitution Glu208Gln in heterozygous state. **A**- Sanger sequencing performed with DNA from a blood sample. **B**- Sanger sequencing performed with sample DNA from the hair follicle.

- High oxygen affinity variant - Hb variants

Two different missense heterozygous variants, in two different patients, were found in *HBB* gene (Table 4). Pathogenic variants in the *HBA1* or *HBA2* genes were not found in any of our studied patients.

Patient N31, a 61-year-old female (Table 4). The NGS study was performed and a missense variant (c.392A>G), causing amino acid substitution Tyr131Cys, was found in heterozygous state, subsequently confirmed by Sanger sequencing of the *HBB* exon 3 (Figure 6). This variant was previously characterized as a Hb variant with high oxygen affinity, called Hb Montfermeil [117].

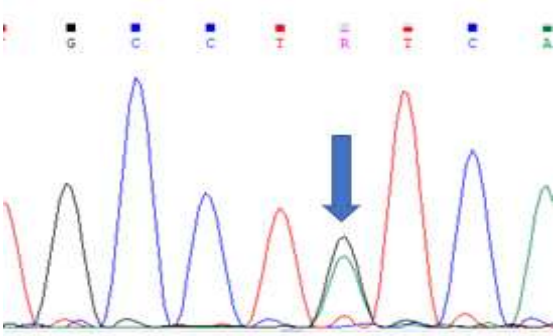

FIGURE 12: *HBB* pathogenic variant in Patient N31.
Identification of Hb Montfermeil in *HBB* c.392A>G, by Sanger sequencing in Patient N31.

Patient N16, an 11-year-old female child (Table 4). The HPLC study was performed, and the presence of an abnormal peak (Hbx) suggests a high oxygen affinity variant (Figure 7A). The Sanger sequencing was made for all 3 exons of the *HBB* gene and the missense variant (c.294C>G) causing amino acid substitution His98Gln, found in a heterozygous state (Figure 7B), confirmed the presence of Hb variant with high oxygen affinity, called Hb Malmö [118].

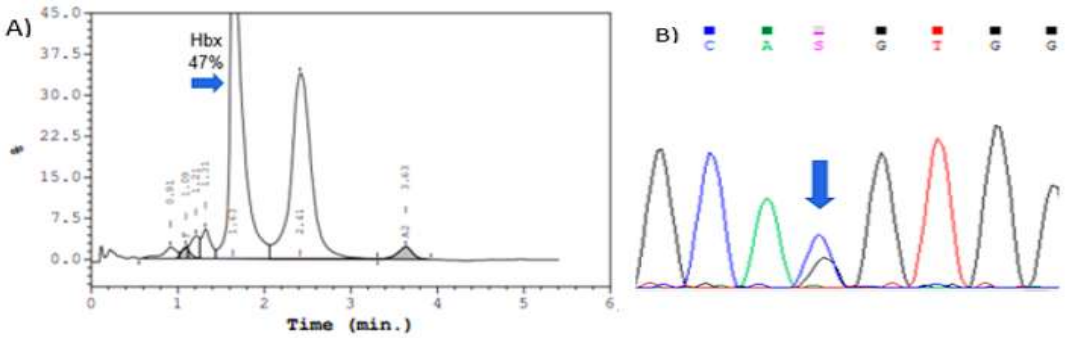

FIGURE 13: HPLC and Sanger sequencing of the Patient N16.
A- HPLC showing the presence of a peak that suggests a high oxygen affinity variant, hemoglobin profile AXA2 - Hbx = 47%. **B-** Identification of Hb Malmö in *HBB* c.294C>G, by Sanger sequencing.

- *VHL* variants

In the *VHL* gene, the pathogenic variant was found in the deep intronic region in the homozygous state, in two different non-consanguineous patients (Table 4).

Patient N60, a 86-year-old male, and Patient N3, a 46-year-old male (Table 4). The deep-intronic change, c.340+816A>C, was found in both patients. This pathogenic variant leads to an alternative splicing site that results in a new cryptic exon (E1'), which combined with exon 1 encodes a non-functional protein [53].

The NGS study was made for Patient N60 and N3, followed by Sanger sequencing to confirm the pathogenic variant and the homozygous state, as shown in Figure 8.

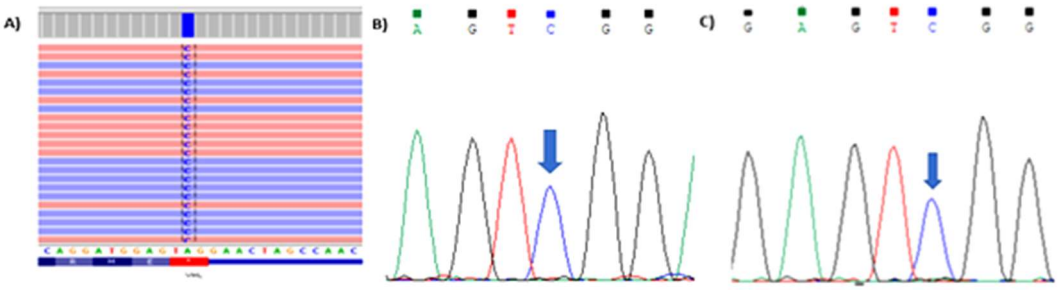

FIGURE 14: *VHL* pathogenic variant in Patient N60 and S1.
Identification of *VHL* deep-intronic c.340+816A>C variant, in homozygous state, by NGS (A) and (B and C) Sanger Sequencing.

4.2.2. Group 2

- *EPOR variant*

Patient N29 a 39-year-old male. A variant was found in exon 2 of the *EPOR* in heterozygous state by NGS study, where the nucleotide change C>T in position 296 causing to the Ala99Val amino acid substitution. The Varsome and Clinvar classify this variant as Benign (Table 5), and the *in silico* analysis predicted to be Likely Benign (Table 7). The analysis performed with the MetaDome and Pfam tools, showed that the variant is located in the non-cytoplasmic domain, at located at p.24 - p.250 in EPOR, specifically the ligand-binding domain p.37 - p.140.

- *JAK2* variant

Patient N55, a 8-year-old female child. A variant was found in heterozygous state in *JAK2* exon 20 that is nucleotide change A>G in position 2681, causing the Asp894Gly amino acid substitution as shown in Figure 9. The Varsome classify this variant as VUS (Table 5), and the *in silico* analysis predicted to be Pathogenic (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is located in the JH1 domain.

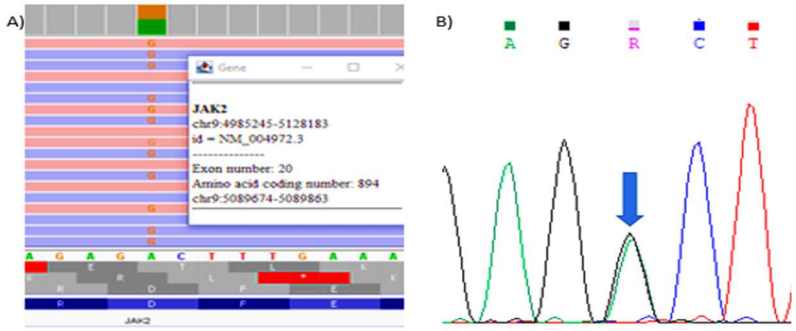

FIGURE 15: *JAK2* variant in Patient N55.
Identification of the *JAK* c.2681A>G variant, in heterozygous state, by NGS (A) and Sanger Sequencing (B).

- *SH2B3* (LNK) variant

Patient N73, a 76-year-old male. A variant was found in exon 7 of the *SH2B3,* in heterozygous state, the nucleotide change G>A in position 1393 that leads to the Val465Ile amino acid substitution (Table 5). The Varsome classify this variant as VUS, and the *in silico* analysis predicted to be Likely Benign (Table 7). The analysis performed with the MetaDome and Pfam tools, showed that variant is located 24 amino acids after Src homology 2 domain (SH2 domain - PF00017), located at p.364 - p.441 in LNK.

- *EGLN1* (PHD2) variant

Patient N20, a 88-year-old female. A variant was found in heterozygous state in *EGLN1* by NGS, the nucleotide change C>A in position 1072 causing the Pro358Thr amino acid substitution (Table 5). The Varsome classify this variant as VUS, and the *in silico* analysis predicted to be Pathogenic (Table 7). An analysis performed with the MetaDome and Pfam tools showed that the variant is located in 2OG-Fe(II) oxygenase superfamily (PF13640) domain, located at p.299 - p.390 in PHD2.

- *EPAS1* (HIF2α) variants

Patient N56, a 18-year-old male. A variant was found in heterozygosity of the *EPAS1* exon 2 (c.181A>G), causing the amino acid substitution Ile61Val (Table 5). The Varsome classify this variant as Benign (Score 11 Pathogenic/12 Benign), and the *in silico* analysis predicted to be Pathogenic (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is not located in any of the domains, and it is 24 amino acids before the PAS fold domain located at p.85 - p.183 in HIF2α.

Patient N41 a 59-year-old female (index case) and N69 a 31-year-old female (N41's daughter). A variant was found in heterozygosity of the *EPAS1* exon 9 (c.1104G>A), causing the amino acid substitution Met368Ile for both samples, mother and daughter. The Varsome and

Clinvar classify this variant as Benign (Table 5), and the *in silico* analysis predicted to be VUS (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is not located in any of the domains, and it is 27 amino acids after the PAS_3 domain, located at p.254 - p.341 in HIF2α.

Patient N77 a 72-year-old female. A variant was found in heterozygosity of the *EPAS1* exon 6 (c.587C>T), causing the amino acid substitution Thr196Met (Table 5). The Varsome classify this variant as Likely Benign (Score 15 Pathogenic/8 Benign), and the *in silico* analysis predicted to be Pathogenic (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is not located in any of the domains, and it is located 13 amino acids after the PAS fold domain, at p.85 - p.183 in HIF2α.

- *PKLR* variant

Patients N13, a 58-year-old male, and N18, a 53-year-old male. The NGS study allowed to identify a variant in both patients in *PKLR* exon 10, that is the nucleotide change A>T in position 1614, in heterozygous state, leading to Glu538Asp amino acid substitution (Table 5). The Varsome and Clinvar classify this variant as VUS, and the *in silico* analysis predicted to be VUS too (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is located in PK alpha/beta domain (PF02887), located at p.453 - p.571.

- *VHL* variants

Patient N23, a 57-year-old male (Table 5). A variant was found in the *VHL* exon 1 by NGS and confirmed by Sanger sequencing, that is the nucleotide change G>A in position 28, in the heterozygous state, as shown in Figure 10, causing the Glu10Lys amino acid substitution. The Varsome and Clinvar classify this variant as VUS, and the *in silico* analysis predicted to be Likely Benign (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is not located in any of the domains of the VHL, being located 53 amino acids before the β-domain, located at p.63 - p.144.

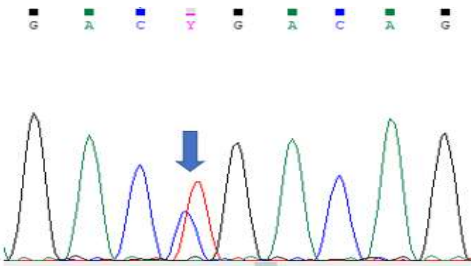

FIGURE 16: *VHL* variant in Patient N23
Variant in *VHL* exon 1 (c.28G>A), in heterozygous state, by Sanger Sequencing.

Patient N48, a 56-year-old male. The NGS study followed by Sanger sequencing (Figure 11) allowed to identify the *VHL* intronic variant c.340+648T>C in heterozygous state (Table 5). The Varsome and Clinvar classify this variant as Benign. The *in silico* analysis was not performed for this intronic variant, because the *in silico* tools used in this study are specific for missense and nonsense variants.

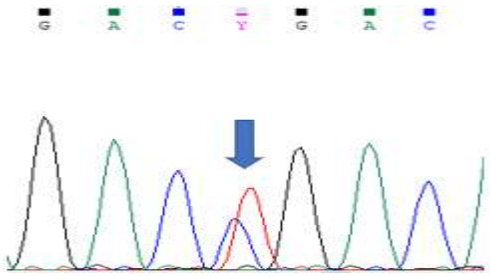

FIGURE 17: *VHL* variant in Patient N48.
Identification of the *VHL* c.340+648T>C intronic variant, in heterozygous state, by Sanger Sequencing.

Patient N27, a 71-year-old male. A variant was found by NGS in exon 1 of *VHL*, which is a C>T nucleotide change at position 74, in the heterozygous state, causing the Pro25Leu amino acid substitution (Table 5). The Sanger sequencing confirmed the presence of this variant (Figure 12). The Varsome and Clinvar classify this variant as Benign, as well as the *in silico* tools (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is not located in any of the domains of the VHL, and it is located 38 amino acids before the Beta domain (PF01847), located at p.63 - p.144.

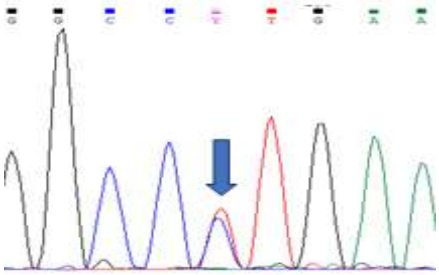

FIGURE 18: *VHL* variant in Patient N27.
Identification of the *VHL* c.74C>T variant exon 1, in heterozygous state, by Sanger Sequencing.

Patient N66, a 51-year-old male. A variant was found in *VHL* exon 1, the nucleotide change G>T in position 154, in heterozygous state, causing the Glu52* stop codon (Table 5), later confirmed by Sanger sequencing (Figure 13). The Varsome and Clinvar classify this variant as Pathogenic and VUS respectively, and the *in silico* analysis predicted to be VUS (Table 7).

In this patient, a second variant was found in *HIF3α* exon 14, which is the nucleotide change C>A in position 1896, in heterozygous state, causing the Asn632Lys amino acid substitution (Table 6). The Varsome classify this variant as Likely Benign, and the *in silico* analysis

predicted to be Benign (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is not located in any of the domains of the HIF3α.

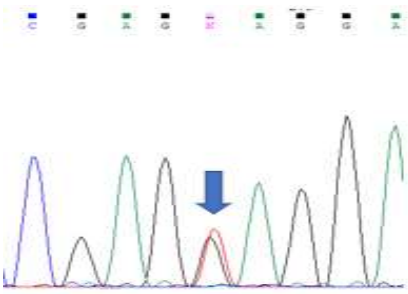

FIGURE 19: *VHL* variant in Patient N66.
Identification of the *VHL* c.154G>T variant exon 1, in heterozygous state, by Sanger Sequencing.

Patient N10, a 76-year-old male (Table 5). A variant was found in the *VHL* exon 1 by NGS and confirmed by Sanger sequencing, which consists in a nucleotide change C>T at position 241, as shown in Figure 14, causing the Pro81Ser amino acid substitution. The Varsome and Clinvar classify this variant as Pathogenic, as well as the *in silico* analysis (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is located in the Beta domain (PF01847), located at p.63 - p.143 of VHL.

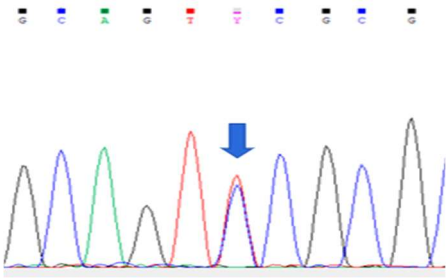

FIGURE 20: *VHL* variant in Patient N10.
Identification of the *VHL* variant c.241C>T, in heterozygous state, by Sanger Sequencing.

4.2.3. Group 3

- *EGLN2* (PHD1) variant

Patient N32, a 69-year-old male. A variant was found in *EGLN2* exon 2, the nucleotide change G>A in position 266, in heterozygous state, causing the Arg89Gln amino acid substitution (Table 6). The Varsome classify this variant as Likely Benign, and the *in silico* analysis predicted to be Benign (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is not located in any of the domains of the PHD1.

- *EGLN3* (PHD3) variant

Patient N51, a 1-year-old male child. A variant was found in *EGLN3* exon 3, the nucleotide change A>G in position 613, at heterozygous state, causing the Arg205Gly amino acid substitution (Table 6). The Varsome classify this variant as Likely Pathogenic, and *in silico* analysis predicted to be Pathogenic (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is located in domain: 2OG-Fe(II) oxygenase superfamily (PF13640), located at p.120 - p.213.

- *HIF1α* variants

Patient N45, a 74-year-old male. A variant was found in *HIF1α* exon 12, the nucleotide change G>A in position 1762, in heterozygous state, causing the Ala588Thr amino acid substitution (Table 6). The Varsome classify this variant as Likely Benign, and *in silico* analysis predicted to be Benign (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is not located in any domain of the HIF1α protein.

- *HIF3α* variant

Patient N64, an 82-year-old male. A variant was found in *HIF3α* exon 10, the nucleotide change G>A in position 1247, in heterozygous state, causing the Arg416His amino acid substitution (Table 6). The Varsome classify this variant as Likely Benign, and *in silico* analysis predicted to be VUS (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is not located in any domain of the HIF1α protein.

- *PIEZO1* variants

In this study many variants were found in *PIEZO*1, as it is a highly polymorphic gene. These variants have been annotated for further detailed study. Only one variant will be described, as it is present in 3 different non-consanguineous patients. Patients N30, a 29-year-old male, N35, a 64-year-old male, and N74, a 24-year-old male, showed the same variant in heterozygous state, in *PIEZO1* exon 47, which is the nucleotide change C>A at position 6829, causing the Leu2277Met amino acid substitution (Table 6). The Varsome classify this variant as VUS, and the *in silico* analysis predicted to be Likely Pathogenic (Table 7). The analysis performed with the MetaDome and Pfam tools showed that the variant is located in Piezo_RRas-bd_domain.

5. Discussion

Congenital Erythrocytosis (CE) is a rare disease that corresponds to an increase in the RBC count, Hb concentration, and Hct above the reference range adapted to age, sex, and living altitude. Congenital causes constitute a very rare subset compared to all acquired causes that are more common, therefore myeloproliferative neoplasms as PV, and secondary acquired causes such as respiratory disease are the most usual causes that course with erythrocytosis. There is no information on the world incidence and prevalence of CE, and the information that exist is mostly reported as individual cases in literature updates [18,119].

In the present study, the medians of hematological patterns of patients, such as Hct and Hb levels, are close to the threshold of normal patterns (Table 3). This fact supports the treatment offered based on their advanced age and the high risk of these patients presenting more critical symptoms, and therefore, regular phlebotomies. The use of low-dose aspirin is usually indicated, as these are effective treatments in myeloproliferative diseases such as PV, and commonly prescribed in the treatment of idiopathic erythrocytosis that controls the rate of Hb, Hct, and RBC mass [12,17]. In this study, it can also be observed that the occurrence of erythrocytosis in men is more elevated (male n = 61; female n = 16). Disease in women could be undiagnosed, notably due to the effect of blood loss associated with menstruation [120,121].

In congenital diseases, the symptoms are usually present in childhood or the stage of a young adult, unlike neoplasms or myeloproliferative diseases that arise in adulthood or old age [122]. Increased Hct and Hb levels are directly related to blood hyperviscosity, and can result in a variety of nonspecific symptoms such as fatigue, dizziness, or headache, which can also lead to a late diagnosis [123].

In this study the average age of patients is 45 years (Table 3), however some of these patients have been diagnosed with idiopathic erythrocytosis a long time ago and now, with the opportunity to perform molecular studies, their doctors advised genetic testing to find the molecular cause of the disease. The knowledge of the molecular cause of a disease can provide personalized and more effective treatment, in addition to the possibility of conducting family studies [5].

With a previously designed NGS panel to CE, we identified the molecular changes that validate the CE phenotype in genes *SH2B3*, *HBB*, and *VHL*. The identified pathogenic variants for a total of 5 patients are already described and associated with erythrocytosis. In a total of 15 patients, we also identified 13 new variants in genes that are already associated with CE; however, these variants were not yet described as causing erythrocytosis; these variants were found in genes *EPOR*, *JAK2*, *SH2B3*, *EGLN1*, *EPAS1*, *EPO*, *PKLR* and *VHL*. Finally, 6 new variants in the candidate genes *EGLN2*, *EGLN3*, *HIF1α*, *HIF3α* and *PIEZO1* were found in a total of 8 patients.

- *JAK2* variant

JAK/STAT5 pathway promotes survival, proliferation, and differentiation of erythroid progenitor cells. It is well established that pathogenic variants in the *JAK2* are responsible for the MPN phenotype, these variants target the myeloid lineage and abnormally activate the JAK2/STAT pathway. Almost all patients with PV have a gain-of-function in JAK2, approximately 96% with variant V617F in exon 14 and 3% with pathogenic variants in exon 12.

In this work, we found in an 8-year-old female child, with hematological patterns suggestive of CE (Hb: 16.80 g/dL; Hct: 48.80%; RBC mass: 5.94 x10^12/L), a new variant in the *JAK2* c.2681A>G (p.Asp894Gly) located at JH1 domain. The Varsome classify this variant as Uncertain Significance (score 21 Pathogenic and 3 Benign), but *in silico* analysis performed in this work predicted it to be Pathogenic. The majority of pathogenic variants in *JAK2* are located in the JH2 domain leading to a loss in the negative control region and consequently confers constitutive activation of JAK2 [119,124], although an autophosphorylation in JH1 mechanism [125] could lead to constant JAK2 activity and consequently, would explain an erythrocytosis phenotype. Functional and familial studies need to be performed to define the pathogenicity of this variant for erythrocytosis.

- *EPOR* variant

EPOR is a member of the type I cytokine receptor superfamily that is predominantly expressed on the surface of immature erythroid cells and has a high affinity for Epo. When Epo selectively associates with EPOR, it is active as a dimer at the cell surface and activates signaling JAK2/STAT5 pathway that promotes the proliferation and differentiation of erythrocyte precursors to mature erythrocytes [14,19,40].

We found a variant in *EPOR* exon 3 (c.296C>T; p.Ala99Val) classified by Varsome as Benign and by *in silico* analysis as Likely Benign. This variant is located in the extracellular region of the protein, more specifically at the erythropoietin receptor ligand-binding site. *EPOR* pathogenic variants can lead to ECYT1, and usually show an autosomal dominant pattern leading to a premature stop codon in the cytoplasmic site with the loss of carboxy-terminal amino acids (exon 8); these truncated EPOR lacks an important negative regulatory region [19,101,126,127,127].

Although this patient carrying this *EPOR* variant has a family history of erythrocytosis and hematological patterns that support the diagnosis of CE, the variant found does not justify the phenotype. The *in silico* analyses shown that this variant is not pathogenic. The variant is not located in the EPOR regulatory region, similar to the other variants already described in the literature. Functional and family studies would be needed to understand better how this variant could cause an ECYT1.

- *SH2B3 (LNK)* variants

The *SH2B3* encodes the LNK adaptor protein, which plays a key role as a negative regulator of normal hematopoiesis. The identification of *SH2B3* genetic variations acquired or inherited in MPNs, lymphoid leukemia and nonmalignant hematological diseases such as IE have demonstrated the importance of LNK in these pathological processes [24,43–45]. In this work we found a pathogenic variant in *SH2B3* c.622G>C (p.Glu208Gln), first described by Oh *et al* [45], and a new variant not yet described, c.1393G>A (p.Val465Ile).

The *SH2B3* c.622G>C (p.Glu208Gln) pathogenic variant was found in previous works with patients diagnosed with IE [44,45,73]; however, no one of these studies described the family history and whether this pathogenic variant was searched for in more than one cell line. The p.Glu208Gln LNK variant affects the PH domain, which results in altering the ability to bind MPL/JAK2 and inhibit downstream STAT activation; in other words, the missense variant affecting the PH domain results in partial loss of LNK function[45]. To exclude the possibility of a somatic variant, we sequenced the *SH2B3* exon 2 using DNA extracted from whole blood and hair; the variant was present in both cell lines, supporting the possibility of being a germline variant or a *de novo* variant. Unfortunately, until now it was not possible to perform familiar studies to understand the inheritance of the variant.

The new variant found in *SH2B3* c.1393G>A (p.Val465Ile) is classified as VUS in the Varsome database (score of 3 pathogenic and 19 benign). It is not reported in Clinvar and was identified in Ensembl as rs374482426 with a MAF of 0. According to the position in the protein structure, the 465 residue is not localized in any important domain. *In silico* analysis predicts that this variant is Likely Benign. Functional and familial studies need to be performed to define the pathogenicity of this variant for the erythrocytosis phenotype.

- *HBB* variants

Hemoglobins are tetramer proteins, responsible for transporting O_2 from the lungs to peripheral tissues through blood flow. Structurally consisting of four polypeptide chains, two α (α1 and α2) and two β (β1 and β2), that form two αβ dimers (named α1β1 and α2β2) are arranged around a 2-fold axis of symmetry resulting in a large central water cavity in the Tense (T) or deoxygenated state and a narrower cavity in the R (Relax) or oxygenated state. Each globin subunit has a heme pocket which is a hydrophobic V- shaped cavity, a region with iron (Fe), this configuration allows Fe to bind O_2 by a covalent bond. The conformational changes between the T state and the R state of globin chains are cooperative and regulate the high-affinity oxygen binding [79,15,81].

Two genes encode the α globin chain, *HBA1* and *HBA2*, and only one gene encodes the β globin chain, *HBB*. The development of secondary erythrocytosis in response to tissue hypoxia is

a physiological response, and more than 100 high oxygen affinity hemoglobin variants are well described, most of them being pathogenic alterations in β chains in an autosomal dominant pattern [80,117,128]. In this study, two patients shown β-chain variants, which were already described as high-affinity O_2 variant Hb.

The Hb Montfermeil (*HBB* c.392A>G, p.Tyr131Cys) was found in a 61-year-old female, in heterozygous state. The patient has an unknown CE familial history, normal Epo, Hb 18.3 g/dL, Hct 52.7%, and red cell mass 6.16x10^12/L. Hb Montfermeil is well described as hemoglobin with increased oxygen affinity and classified as pathogenic by Ithanet (IthaID:1249) and Varsome databases. Functional studies performed by J. Kister *et al.*[117] showed that Tyr (H8) is located inside the Heme pocket where it interacts with other hydrophobic residues. The change of Tyr to Cys leads to alteration in the network of hydrophobic interactions and an increase in oxygen affinity [117,129].

The Hb Malmö (*HBB* c.294C>G, p.His98Gln) was found in an 11-year-old female child, with a strong CE familial history and a Hb 16.3 g/dL. Hb Malmö is also well described as hemoglobin with increased oxygen affinity and classified as pathogenic by Ithanet (IthaID: 1130) and Varsome databases. The histidine is replaced by glutamine and this alteration causes a disorder in the connection between the α1-subunit and β2-subunit of the Hb structure, which implicates that there is an increased binding for O_2 and decreased release of O_2 in tissues, associated with physiological erythrocytosis [130,118].

- *VHL* variants

The main function of the VHL is to act as a tumor suppressor, where pVHL30 and pVHL19 act as signaling proteins in the ubiquitin-proteasome degradation process. Their main target is the hydroxylated HIF2α molecule, thus performing upregulation in the transcription of *EPO* and other genes which are transcribed into the oxygen-sensing pathway [54]. A new exon from a deep-intronic region was found between exon 1 and exon 2 that results from an alternative splice and was named E1', which together with Exon 1 encodes a 193aa protein [53].

Pathogenic variants in *VHL* are related to different phenotypes, such as different tumor types, von Hippel-Lindau disease, Chuvash Polycythemia and ECYT2, the last two are characterized by variants in *VHL* with an autosomal recessive pattern in homozygous or compound heterozygous state. Moreover, there are some clinical cases where the family genetic study demonstrates inexplicable heterozygous patterns [6,17,54]. In this study we found one *VHL* pathogenic variant in two different patients, c.340+816A>C, that was already described to promote erythrocytosis. Others five new variants were found in five different patients, in the heterozygous state: c.340+648T>C, c.28G>A (p.Glu10Lys), c.74C>T (p. Pro25Leu), c.154G>T (p. Glu52*), c.241C>T (p.Pro81Ser).

The intronic *VHL* c.340+816A>C was found in two different and non-consanguineous patients, in homozygous state. The Patient N60 is a 86 year-old male, with high Epo levels (45.4 mUI/mL), and the N3 Patient is a 46 year-old man, also high Epo levels (85,7 mUI/mL); both patients have high levels of Hct and Hb. The study performed by Lenglet *et al*. [53] propose that pathogenic variants in E1' induce severe retention of this cryptic E1 exon, which correlated with a defect in the overall expression of the VHL protein. This splicing dysregulation, with consequent down-regulation of VHL expression, is the underlying cause of ECTY2, through insufficient levels of VHL and not by the reduction of HIF binding by the mutant VHL.

The *VHL* c.340+648T>C, another intronic variant, was found in heterozygous state in Patient N48, a 56 year-old male with elevated hematological patterns (Hb 20 g/dL). This variant is classified as Benign in the Varsome and ClinVar databases. The *in silico* analysis was not performed for this variant, because all tools which we used in this study are specific for missense and nonsense variants. In the study performed by Lenglet *et al*. [53], above cited, shown the same variant in compound heterozygosity (c.340+617C>G / c.340+648T>C) in a patient with VHL disease. Therefore, we can infer that this variant is probably not responsible for the erythrocytosis presented by the N48 patient, mainly because it was previously correlated with VHL disease and, in this specific case, it presents a heterozygous pattern [53].

The *VHL* c.28G>A (p.Glu10Lys) was found in heterozygous state in Patient N23, a 57 year-old male. The variant is classified as VUS in the Varsome and ClinVar databases (score of 4 pathogenic and 19 benign) and *in silico* analysis predict it as Likely Benign. This sequence change replaces glutamic acid with lysine at codon 10 of the VHL protein. The glutamic acid residue is moderately conserved and there is a small physicochemical difference between glutamic acid and lysine [131]. The Mutation Update performed by Bento *et al.* [5] describe that a patient from France presented a p.Glu10* also in a heterozygous state. This particular variant is located between the two translation initiation codons and has the ability to produce a pVHL19 isoform still capable of regulating HIF. Therefore, this fact corroborates the analyzes carried out with the *in silico* tools. Perhaps this variant is not responsible for the presented phenotype, although functional studies could better elucidate the mechanism of this variant.

The *VHL* c.74C>T (p. Pro25Leu) was found in heterozygous state in Patient N27, a 71 year-old male. This variant is classified as Benign in the Varsome and ClinVar databases. The *in silico* analysis also predicted to be Benign. According to molecular location, this variant is not found in any domain of the protein, and ClinVar report this variant as a conservative change: it occurs at a poorly conserved position in the protein, it is predicted to be benign by multiple *in silico* algorithms, and/or has population frequency not consistent with diseases [132]. In accordance with the facts presented, we can infer that this variant is probably not responsible for the erythrocytosis presented by the N27 patient.

The *VHL* c.154G>T (p.Glu52*) was found in Patient N66, a 51-year-old male, with a family history of polyglobulia and hematological pattern suggestive of erythrocytosis. The Varsome classifies this variant as Pathogenic and ClinVar as VUS. The *in silico* analysis also predicted it to be a VUS. This nonsense variant predicts translation termination of the long VHL isoform (p30) leading to translation of only the short and alternative form of VHL (p19). Curiously, two previous studies performed by Camps *et al.*[73] and Oliveira *et al.* [133] with IE patients also related the same p.Glu52* variant in a heterozygous state [73,133]. Nonsense variats in the *VHL* are most commonly associated with VHL disease. However, the other two nonsense variants reported in the literature, p.E10*, and p.E46* are associated with apparently isolated erythrocytosis and with Hippel-Lindau syndrome, respectively [5,134]. The role of the heterozygote VHL p.E52* in the production of erythrocytosis is not clear and, in this case, the patient should be recommended to perform further tests to exclude the possibility of VHL disease.

The *VHL* c.241C>T (p. Pro81Ser) was found in Patient N10, a 76-year-old male, with suggestive familiar history of the erythrocytosis. The Varsome classified this variant as Pathogenic and ClinVar as VUS. The *in silico* analysis also predicted it to be Pathogenic. This variant is located in VHL beta domain and other pathogenic variants in this site are described with increased Risk of tumor development [135]. Intriguingly, this variant was also found in a patient with IE in a recent study by Oliveira *et al.* [133] from Minnesota USA. Functional and familial studies need to be performed to define the pathogenicity of this variant for the erythrocytosis phenotype and the patient should be recommended to perform further tests to exclude the possibility of development renal cell carcinoma or hemangioblastomas.

- *EGLN1- EGLN2, EGLN3* (PHD) variants

EGLN1 encodes the PHD2 protein, which is described to be a key enzyme in catalyzing the prolyl hydroxylation of HIFα. Changes in PHD2, that reduce hydroxylase activity, are correlated with erythrocytosis [5,63]. Although, how other mutated PHD isoforms can promote erythrocytosis remain unknown. Among all the variants identified in this study, we identified 3 new variants in PHD, each variant in a different isoform [136,137].

The *EGLN1* variant c.1072C>A (p.Pro358Thr) was found in the heterozygous state in Patient N20, an 88-year-old female. This variant is classified as a VUS by Varsome, and *in silico* analysis predicts it as Pathogenic. Percy *et al.* [63,64] reported two variants p.Pro317Arg and p.Arg371His associated with CE, they are able to reduce hydroxylase activity and promote the erythrocytosis phenotype, after this find other variants were also found in nearby amino acids. These variants happen within the 2OG-Fe(II) oxygenase domain [5,63,64,138]. The present results suggest that this variant may be responsible for the CE phenotype, even though family and functional studies are needed to validate the pathogenicity of the variant.

The *EGLN2* variant c.266G>A (p.Arg89Gln) was found in the heterozygous state in Patient N32, a 69-year-old male. This variant is classified as L. Benign by Varsome, and the *in silico* analysis also predicts it as Likely Benign. This variant is not within any domain of PHD1, as mentioned above, thus the mechanisms of how PHD isoforms can lead to erythrocytosis are uncertain. Further functional studies are needed to elucidate the effects of this variant.

The *EGLN3* variant c.613A>G (p.Arg205Gly) was found in the heterozygous state in Patient N51, a 1-year-old child male. This variant is classified as Likely Pathogenic by Varsome, and *in silico* analysis also predicts it as Pathogenic. These variant occurs within the 2OG-Fe(II) oxygenase domain of PHD3. In fact, it is important to emphasize that this variant is within an important domain in the protein, as well as the pathogenic variants described in *EGLN1* that lead to erythrocytosis. Based on the age of the patient and in their high hematological patterns (Hb 16.5 g/dL; adapted by age), we can infer that this new variant is a strong candidate for a functional and familial study.

- *EPAS1* (HIF2α) variants

The HIF2α is the primary transcription factor to stimulates *EPO* expression in response to hypoxia. Their degradation occurs by hydroxylation of the residues Pro405 and Pro531 and subsequently VHL E3 ligase signaling to the ubiquitin-proteasome system. The first pathogenic variants associated with CE were described by Percy *et al.* [64,71] these include four variants within the NAD domain in exon 12 lead to gain of function in HIF2α and impaired interaction of HIF-2α with PHD2 and VHL [64,139].

The *EPAS1* variant c.181A>G p.Ile61Val was found in the heterozygous state in Patient N56, an 18-year-old male. This variant is classified as Benign by Varsome (score 11 pathogenic and 12 benign), and *in silico* analysis also predicts it as Pathogenic. According to InterPro database, this variant is within the bHLH domain, a conserved protein domain [140]. The bHLH domain function as a diverse set of regulatory factors because they recognize different DNA sequences and dimerize with different proteins. They are involved in cell fate determination and act in neurogenesis, cardiogenesis, myogenesis, and hematopoiesis [62,141]. To date, no variant associated with erythrocytosis has been described within this highly conserved domain: Therefore, functional and familial studies need to be carried out to confirm how this change could lead to the pathogenic phenotype.

The *EPAS1* variant c.587C>T (p.Thr196Met) was found in the heterozygous state in Patient N77, a 72-year-old female. This variant is classified as L. Benign by Varsome (score 15 Pathogenic and 8 Benign), and *in silico* analysis also predicts it as Pathogenic. No pathogenic variants in exon 6 associated with CE have been described to date. With a MAF <0.01 and because it is a highly conserved exon, there is a high possibility that this variant is associated with the erythrocytosis

phenotype. According to shared information among CE study groups in Europe, this variant was found in a man from France with suggestive symptoms of the CE (personal communication, not yet published). This data confirms the need to carry out functional and familial studies to elucidate the mechanisms of this missense variant.

The *EPAS1* variant c.1104G>A (p.Met368Ile) was found in the heterozygous state in Patient N41, a 59-year-old female, and N69, a 31-year-old female (Mother and Daughter). This variant is classified as Benign by Varsome (score 4 pathogenic and 18 benign), and the *in silico* analysis also predicts it as VUS. The p.Met368Ile variant is located in an exclusive conserved site at exon 9. Lorenzo *et al*. [142] found a pathogenic variant in the same exon 9 at a nearby amino acid. As described, p.Phe374Tyr variant causes conformational changes at pVHL b-domain sites and exhibited significant changes in impairment of ElonginC interaction, leading to gain-of-function of HIF-2α. This pathogenic variant is associated with CE that is also associated with multiple paragangliomas tumors.

- *HIF1α* and *HIF3α* variants

In this study, we also found variants in the other isoforms of HIF, in heterozygous state and with MAF >0.01. Two of them are in the HIF3α isoform (Arg416His, Asn632Lys) and one in the HIF1α isoform (Ala588Thr). According to evaluations performed with the *in silico* tools, only HIF3a Arg416His is classified as a VUS. However, the mechanisms that can lead to CE as a result of the changes found in the HIF1a and HIF3a isoforms are still not clearly understood. Based on the high sequence homology of the 3 isoforms, the role of its isoforms in disease development should not be neglected [72]. Anyway, it would be interesting to proceed with functional and family studies of the HIF3α Arg416His in order to better understand this variant.

- *PKLR* variant

Pathogenic variants in *PKLR* can lead to CE by PK hyperactivity with high ATP levels, recruitment of 1,3-BPG to the glycolytic pathway, subsequent deficiency 2,3-BPG and a consequent increase in Hb affinity for oxygen. Few pathogenic variants in *PKLR* are described associated with CE [83,84].

In this work we found one variation in *PKLR* c.1614A>T (p.Glu538Asp) in two non-consanguineous patients, N13, a 58-year-old male, and N18, a 53-year-old male, both in heterozygous state. The PKLR Glu538Asp is classified by Varsome and Clinvar as VUS, and the *in silico* analysis performed in this work predict this variant as VUS. This variant was found by Manco and collaborators [143] at heterozygous state in a woman from São Tomé e Príncipe (West Africa) and classified as a variant potentially causing Pyruvate kinase deficiency. According to the data presented herein, a study of PK enzymatic properties and immunological specific activity, may

help to better understand whether in these two isolated cases the erythrocytosis is caused by the same variant.

- *PIEZO* variant

The recent study performed by Filser et al. [85] associates PIEZO1 pathogenic variants with erythrocytosis by an increase in Hb oxygen affinity with intense glycolytic activity, leading to a decreased 2,3-DPG concentration to maintain the high ATP concentration required for cell homeostasis.

In your study we found many variants in *PIEZO1* with MAF>0.01 and normally classified by Varsome and Clinvar as benign polymorphisms; therefore, these variants were not submitted to posterior studies. A single *PIEZO1* variant was chosen to be analyzed in this work because it was present in three non-consanguineous male patients, aged 24-29 to 64 years. The *PIEZO1* c.6829C>A (p.Leu2277Met), variant is located in an extracellular domain at the C-terminus of the Piezo1 protein, it is classified as VUS by Varsome and as Likely Pathogenic by *in silico* analysis. Based on the results found and the clinical patterns presented by the patients, this variant is a strong candidate for family and functional studies.

6. Conclusions

In conclusion, this study allowed us to obtain a significant contribution to the molecular basis underlying CE in a cohort of European patients. These are the strong acknowledgments:

✓ With a previously established NGS panel of 13 genes associated with erythrocytosis, 45 genes associated with hemolytic anemia, and 6 genes for both diseases, we identified high-reliability variants among patients with IE and discovered other potential disease-causing variants in patients who previously did not have one specific diagnosis.

✓ We were able to provide a real genetic cause for erythrocytosis congenital for 5 patients. Other candidate variants were evaluated by *in silico* tools and 11 possible variants can be disease-causing: 6 were classified as Pathogenic, 1 as Likely Pathogenic, and 4 as VUS. The other 7 variants were classified as Benign or Likely Benign.

✓ Performing family and functional studies of the new variants found are extremely important for a genetic study, aiming to elucidate the mechanisms of transmission, and whether these variants are in fact congenital or acquired variants. The first step of a molecular family study by NGS was carried out, and, as a result, mother and daughter have an *EPAS1* variant (p.Met368Ile) classified by *in silico* analysis as VUS.

✓ We were able to collect important information to proceed with familial and functional studies of potentially pathogenic variants.

7.Future Perspective

The prospect of this work for the future is the possibility of carrying out new studies, such as family and functional studies, which can elucidate the definitive role of these new variants in the role of erythrocytosis.

Regarding the 63% that remain patients with an unknown genetic cause, therefore being within idiopathic erythrocytosis, the first step is to perform a Whole-genome analysis.

The epigenetic study will also be a way to better understand how methylations could alter the production of erythropoietin or even alterations in the intrinsic control of the differentiation and proliferation of erythroblasts.

We should not forget that erythrocytosis can be congenital or acquired, secondary to the exposure to exogenous sources. Maybe the answer is in not look for them separately but think that in same individuals both of the situations can be present. Probably some of the benign variants not valuable can have an influence in the Hb levels if associated with environmental characteristics. Research on the human Exposome is a growing area to understand the causal pathways leading to common diseases [144–146], congenital or not, and it could also bring an answer to the unsolved cases of CE.

8. References

1. Guyton AC, Hall JE. GUYTON AND HALL - TEXTBOOK OF MEDICAL PHYSIOLOGY. 13th ed. Rio de Janeiro: Elsevier Inc.; 2017. 3333 p.

2. Palis J. Primitive and definitive erythropoiesis in mammals. Front Physiol. 2014;5.

3. Palis J, Robertson S, Kennedy M, Wall C, Keller G. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. Dev Camb Engl. 1999 Nov;126(22):5073–84.

4. Hattangadi SM, Wong P, Zhang L, Flygare J, Lodish HF. From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. Blood. 2011 Dec 8;118(24):6258–68.

5. Bento C, Percy MJ, Gardie B, Maia TM, van Wijk R, Perrotta S, Della Ragione F, Almeida H, Rossi C, Girodon F, Åström M, Neumann D, Schnittger S, Landin B, Minkov M, Randi ML, Richard S, Casadevall N, Vainchenker W, Rives S, Hermouet S, Ribeiro ML, McMullin MF, Cario H, on behalf of ECE-Consortium, Chauveau A, Gimenez-Roqueplo A-P, Bressac-de-Paillerets B, Altindirek D, Lorenzo F, Lambert F, Dan H, Gad-Lapiteau S, Catarina Oliveira A, Rossi C, Fraga C, Taradin G, Martin-Nuñez G, Vitória H, Diaz Aguado H, Palmblad J, Vidán J, Relvas L, Ribeiro ML, Luigi Larocca M, Luigia Randi M, Pedro Silveira M, Percy M, Gross M, Marques da Costa R, Beshara S, Ben-Ami T, Ugo V. Genetic Basis of Congenital Erythrocytosis: Mutation Update and Online Databases. Hum Mutat. 2014 Jan;35(1):15–26.

6. Bento C. Genetic basis of congenital erythrocytosis. Int J Lab Hematol. 2018 May;40:62–7.

7. Yang M, Su H, Soga T, Kranc KR, Pollard PJ. Prolyl hydroxylase domain enzymes: important regulators of cancer metabolism. Hypoxia. 2014 Aug 30;2:127–42.

8. Patnaik MM, Tefferi A. The complete evaluation of erythrocytosis: congenital and acquired. Leukemia. 2009 May;23(5):834–44.

9. Aljabry M. Primary familial and congenital polycythemia; The forgotten entity. J Appl Hematol. 2018;9(2):39.

10. McMullin MF. Congenital erythrocytosis. Int J Lab Hematol. 2016 May;38:59–65.

11. West JB, Richalet J-P. Denis Jourdanet (1815–1892) and the early recognition of the role of hypoxia at high altitude. Am J Physiol-Lung Cell Mol Physiol. 2013 Sep 1;305(5):L333–40.

12. McMullin MFF, Mead AJ, Ali S, Cargo C, Chen F, Ewing J, Garg M, Godfrey A, Knapper S, McLornan DP, Nangalia J, Sekhar M, Wadelin F, Harrison CN, the British Society for Haematology Guideline. A guideline for the management of specific situations in polycythaemia vera and secondary erythrocytosis: A British Society for Haematology Guideline. Br J Haematol. 2019 Jan;184(2):161–75.

13. Lewandowski K. Medical dilemmas – erythrocytosis. Acta Haematol Pol. 2020 Dec 1;51(4):203–11.

14. Sarper nazan. EPOR mutasyonuna bağlı konjenital eritrositozu olan Türk ailenin klinik özellikleri. Çocuk ve ergenlerde düzenli flebotomi gereklimidir? Türk Pediatri Arş [Internet]. 2019; Available from: http://turkarchpediatr.org//en/clinical-characteristics-of-a-turkish-family-with-congenital-erythrocytosis-due-to-an-epor-mutation-is-routine-phlebotomy-indicated-in-children-and-adolescents-13188

15. Mangin O. High oxygen affinity hemoglobins. Rev Médecine Interne. 2017 Feb;38(2):106–12.

16. McMullin MF. The classification and diagnosis of erythrocytosis. Int J Lab Hematol. 2008;30(6):447–59.

17. McMullin MF. Genetic Background of Congenital Erythrocytosis. Genes. 2021 Jul 28;12(8):1151.

18. Members of the European Congenital Erythrocytosis Consortium (ECEC), Members of the COST Action BM0902 "MPN&MPNr-EuroNet," Bento C, Cario H, Gardie B, Hermouet S, McMullin MF. Congenital Erythrocytosis and Hereditary Thrombocytosis. Europe; 2015. 195 p.

19. Bento C, McMullin MF, Percy M, Cario H. Primary Familial and Congenital Polycythemia. 2016 Nov;15.

20. Sokol L, Luhovy M, Guan Y, Prchal J, Semenza G, Prchal J. Primary familial polycythemia: a frameshift mutation in the erythropoietin receptor gene and increased sensitivity of erythroid progenitors to erythropoietin. Blood. 1995 Jul 1;86(1):15–22.

21. Tondeur S, Paul F, Riou J, Mansier O, Ranta D, Le Clech L, Lippert E, Tavitian S, Chaoui D, Mercier M, De Renzis B, Cottin L, Cassinat B, Chrétien J-M, Ianotto J-C, Allangba O, Marzac C, Voillat L, Boyer F, Orvain C, Hunault-Berger M, Girodon F, Kiladjian J-J, Ugo V, Luque Paz D. Long-term follow-up of JAK2 exon 12 polycythemia vera: a French Intergroup of Myeloproliferative Neoplasms (FIM) study. Leukemia [Internet]. 2020 Jul 21 [cited 2020 Jul 24]; Available from: http://www.nature.com/articles/s41375-020-0991-x

22. Cross NCP. Genetic and Epigenetic Complexity in Myeloproliferative Neoplasms. Hematology. 2011 Dec 10;2011(1):208–14.

23. Tondeur S, Paul F, Riou J, Mansier O, Ranta D, Le Clech L, Lippert E, Tavitian S, Chaoui D, Mercier M, De Renzis B, Cottin L, Cassinat B, Chrétien J-M, Ianotto J-C, Allangba O, Marzac C, Voillat L, Boyer F, Orvain C, Hunault-Berger M, Girodon F, Kiladjian J-J, Ugo V, Luque Paz D. Long-term follow-up of JAK2 exon 12 polycythemia vera: a French Intergroup of Myeloproliferative Neoplasms (FIM) study. Leukemia [Internet]. 2020 Jul 21 [cited 2020 Jul 25]; Available from: http://www.nature.com/articles/s41375-020-0991-x

24. McMullin MF, Cario H. LNK mutations and myeloproliferative disorders: LNK and MPN. Am J Hematol. 2016 Feb;91(2):248–51.

25. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M, Vardiman JW. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016 May 19;127(20):2391–405.

26. Petousi N, Copley RR, Lappin TRJ, Haggan SE, Bento CM, Cario H, Percy MJ, The WGS Consortium, Ratcliffe PJ, Robbins PA, McMullin MF, Donnelly P, Bell J, Bentley D, McVean G, Ratcliffe P, Taylor J, Wilkie A, Donelly P, Broxholme J, Buck D, Cazier J-B, Cornall R, Gregory L, Knight J, Lunter G, McVean G, Tomlinson I, Wilkie A, Buck D, Allan C, Attar M, Green A, Gregory L, Humphray S, Kingsbury Z, Lamble S, Lonie L, Pagnamenta A, Piazza P, Polanco G, Trebes A, McVean G, Donnelly P, Cazier J-B, Broxholme J, Copley R, Fiddy S, Grocock R, Hatton E, Holmes C, Hughes L, Humburg P, Kanapin A, Lise S, Lunter G, Martin H, Murray L, McCarthy D, Rimmer A, Sahgal N, Wright B, Yau C. Erythrocytosis associated with a novel missense mutation in the BPGM gene. Haematologica. 2014 Oct 1;99(10):e201–4.

27. Transcript: JAK2-201 (ENST00000381652.4) - Summary - Homo_sapiens - Ensembl genome browser 104 [Internet]. [cited 2021 Oct 27]. Available from: http://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000096968;r=9:4984390-5129948;t=ENST00000381652

28. The UniProt Consortium, Bateman A, Martin M-J, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, Bye-A-Jee H, Coetzee R, Cukura A, Da Silva A, Denny P, Dogan T, Ebenezer T, Fan J, Castro LG, Garmiri P, Georghiou G, Gonzales L, Hatton-Ellis E, Hussein A, Ignatchenko A, Insana G, Ishtiaq R, Jokinen P, Joshi V, Jyothi D, Lock A, Lopez R, Luciani A, Luo J, Lussi Y, MacDougall A, Madeira F, Mahmoudy M, Menchi M, Mishra A, Moulang K, Nightingale A, Oliveira CS, Pundir S, Qi G, Raj S, Rice D, Lopez MR, Saidi R, Sampson J, Sawford T, Speretta E, Turner E, Tyagi N, Vasudev P, Volynkin V, Warner K, Watkins X, Zaru R, Zellner H, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter M-C, Bolleman J, Boutet E, Breuza L, Casals-Casas C, de Castro E, Echioukh KC, Coudert E, Cuche B, Doche M, Dornevil D, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Hyka-Nouspikel N, Jungo F, Keller G, Kerhornou A, Lara V, Le Mercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto TB, Paesano S, Pedruzzi I, Pilbout S, Pourcel L, Pozzato M, Pruess M, Rivoire C, Sigrist C, Sonesson K, Stutz A, Sundaram S, Tognolli M, Verbregue L, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Yeh L-S, Zhang J, Ruch P, Teodoro D. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D480–9.

29. CDD Conserved Protein Domain Family: FERM_C_JAK2 [Internet]. Available from: https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cd13333

30. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D412–9.

31. CDD Conserved Protein Domain Family: SH2_Jak2 [Internet]. Available from: https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cd10379

32. CDD Conserved Protein Domain Family: PTK_Jak2_rpt1 [Internet]. [cited 2021 Oct 10]. Available from: https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cd05078

33. CDD Conserved Protein Domain Family: PTKc_Jak2_rpt2 [Internet]. Available from: https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=cd14205

34. Randi ML, Bertozzi I, Cosi E, Santarossa C, Peroni E, Fabris F. Idiopathic erythrocytosis: a study of a large cohort with a long follow-up. Ann Hematol. 2016 Jan;95(2):233–7.

35. Pietra D, Li S, Brisci A, Passamonti F, Rumi E, Theocharides A, Ferrari M, Gisslinger H, Kralovics R, Cremonesi L, Skoda R, Cazzola M. Somatic mutations of JAK2 exon 12 in patients with JAK2 (V617F)-negative myeloproliferative disorders. Blood. 2008 Feb 1;111(3):1686–9.

36. Scott LM, Tong W, Levine RL, Scott MA, Beer PA, Stratton MR, Futreal PA, Erber WN, McMullin MF, Harrison CN, Warren AJ, Gilliland DG, Lodish HF, Green AR. *JAK2* Exon 12 Mutations in Polycythemia Vera and Idiopathic Erythrocytosis. N Engl J Med. 2007 Feb;356(5):459–68.

37. Pietra D, Li S, Brisci A, Passamonti F, Rumi E, Theocharides A, Ferrari M, Gisslinger H, Kralovics R, Cremonesi L, Skoda R, Cazzola M. Somatic mutations of JAK2 exon 12 in patients with JAK2 (V617F)-negative myeloproliferative disorders. Blood. 2008 Feb 1;111(3):1686–9.

38. Bento C, Almeida H, Maia TM, Relvas L, Oliveira AC, Rossi C, Girodon F, Fernandez-Lago C, Aguado-Diaz A, Fraga C, Costa RM, Araújo AL, Silva J, Vitória H, Miguel N, Silveira MP, Martin-Nuñez G, Ribeiro ML. Molecular study of congenital erythrocytosis in 70 unrelated patients revealed a potential causal mutation in less than half of the cases (Where is/are the missing gene(s)?). Eur J Haematol. 2013 Aug;n/a-n/a.

39. Transcript: EPOR-201 (ENST00000222139.11) - Exons - Homo_sapiens - Ensembl genome browser 104 [Internet]. [cited 2021 Oct 25]. Available from: http://www.ensembl.org/Homo_sapiens/Transcript/Exons?db=core;g=ENSG00000187266;r=19:11377207-11384342;t=ENST00000222139

40. Watowich SS. The Erythropoietin Receptor: Molecular Structure and Hematopoietic Signaling Pathways. J Investig Med Off Publ Am Fed Clin Res. 2011 Oct;59(7):1067–72.

41. McMullin MF. Investigation and Management of Erythrocytosis. Curr Hematol Malig Rep. 2016 Oct;11(5):342–7.

42. Bittorf T, Seiler J, Zhang Z, Jaster R, Brock J. SHP1 Protein Tyrosine Phosphatase Negatively Modulates Erythroid Differentiation and Suppression of Apoptosis in J2E Erythroleukemic Cells. Biol Chem [Internet]. 1999 Jan 12 [cited 2020 Jul 23];380(10). Available from: https://www.degruyter.com/view/j/bchm.1999.380.issue-10/bc.1999.152/bc.1999.152.xml

43. Maslah N, Cassinat B, Verger E, Kiladjian J-J, Velazquez L. The role of LNK/SH2B3 genetic alterations in myeloproliferative neoplasms and other hematological disorders. Leukemia. 2017 Aug;31(8):1661–70.

44. Spolverini A, Pieri L, Guglielmelli P, Pancrazzi A, Fanelli T, Paoli C, Bosi A, Nichele I, Ruggeri M, Vannucchi AM. Infrequent occurrence of mutations in the PH domain of LNK in patients with JAK2 mutation-negative "idiopathic" erythrocytosis. Haematologica. 2013 Sep 1;98(9):e101–2.

45. Oh ST, Simonds EF, Jones C, Hale MB, Goltsev Y, Gibbs KD, Merker JD, Zehnder JL, Nolan GP, Gotlib J. Novel mutations in the inhibitory adaptor protein LNK drive JAK-STAT signaling in patients with myeloproliferative neoplasms. Blood. 2010 Aug 12;116(6):988–92.

46. Langabeer SE, Conneally E, Flynn CM. Myelodysplastic Syndrome/Acute Myeloid Leukemia Arising in Idiopathic Erythrocytosis. Case Rep Hematol. 2018;2018:1–4.

47. Camps C, Petousi N, Bento C, Cario H, Copley RR, McMullin MF, van Wijk R, Ratcliffe PJ, Robbins PA, Taylor JC. Gene panel sequencing improves the diagnostic work-up of patients with idiopathic erythrocytosis and identifies new mutations. Haematologica. 2016 Nov;101(11):1306–18.

48. Transcript: VHL-201 (ENST00000256474.3) - Summary - Homo_sapiens - Ensembl genome browser 104 [Internet]. [cited 2021 Oct 21]. Available from: http://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000134086;r=3:10141778-10153667;t=ENST00000256474

49. Schoenfeld A, Davidowitz EJ, Burk RD. A second major native von Hippel-Lindau gene product, initiated from an internal translation start site, functions as a tumor suppressor. Proc Natl Acad Sci. 1998 Jul 21;95(15):8817–22.

50. Chesnel F, Hascoet P, Gagné JP, Couturier A, Jouan F, Poirier GG, Le Goff C, Vigneau C, Danger Y, Verite F, Le Goff X, Arlot-Bonnemains Y. The von Hippel–Lindau tumour suppressor gene: uncovering the expression of the pVHL172 isoform. Br J Cancer. 2015 Jul;113(2):336–44.

51. Hascoet P, Chesnel F, Jouan F, Goff CL, Couturier A, Darrigrand E, Mahe F, Rioux-Leclercq N, Goff XL, Arlot-Bonnemains Y. The pVHL172 isoform is not a tumor suppressor and up-regulates a subset of pro-tumorigenic genes including *TGFB1* and *MMP13*. Oncotarget. 2017 Sep 29;8(44):75989–6002.

52. Stebbins CE, Kaelin WG, Pavletich NP. Structure of the VHL-ElonginC-ElonginB Complex: Implications for VHL Tumor Suppressor Function. Science. 1999 Apr 16;284(5413):455–61.

53. Lenglet M, Robriquet F, Schwarz K, Camps C, Couturier A, Hoogewijs D, Buffet A, Knight SJL, Gad S, Couvé S, Chesnel F, Pacault M, Lindenbaum P, Job S, Dumont S, Besnard T, Cornec M, Dreau H, Pentony M, Kvikstad E, Deveaux S, Burnichon N, Ferlicot S, Vilaine M, Mazzella J-M, Airaud F, Garrec C, Heidet L, Irtan S, Mantadakis E, Bouchireb K, Debatin K-M, Redon R, Bezieau S, Bressac-de Paillerets B, Teh BT, Girodon F, Randi M-L, Putti MC, Bours V, Van Wijk R, Göthert JR, Kattamis A, Janin N, Bento C, Taylor JC, Arlot-Bonnemains Y, Richard S, Gimenez-Roqueplo A-P, Cario H, Gardie B. Identification of a new VHL exon and complex splicing alterations in familial erythrocytosis or von Hippel-Lindau disease. Blood. 2018 Aug 2;132(5):469–83.

54. Hwang S-H, Bang S, Kim W, Chung J. Von Hippel–Lindau tumor suppressor (VHL) stimulates TOR signaling by interacting with phosphoinositide 3-kinase (PI3K). J Biol Chem. 2020 Jan 20;jbc.RA119.011596.

55. Bader HL, Hsu T. Systemic *VHL* gene functions and the VHL disease. FEBS Lett. 2012 Jun 4;586(11):1562–9.

56. Lappin TR, Lee FS. Update on mutations in the HIF: EPO pathway and their role in erythrocytosis. Blood Rev. 2019 Sep;37:100590.

57. Ang SO, Chen H, Hirota K, Gordeuk VR, Jelinek J, Guan Y, Liu E, Sergueeva AI, Miasnikova GY, Mole D, Maxwell PH, Stockton DW, Semenza GL, Prchal JT. Disruption of oxygen homeostasis underlies congenital Chuvash polycythemia. Nat Genet. 2002 Dec;32(4):614–21.

58. Ang SO, Chen H, Gordeuk VR, Sergueeva AI, Polyakova LA, Miasnikova GY, Kralovics R, Stockton DW, Prchal JT. Endemic Polycythemia in Russia: Mutation in the VHL Gene. Blood Cells Mol Dis. 2002 Jan;28(1):57–62.

59. Buffet A, Calsina B, Flores S, Giraud S, Lenglet M, Romanet P, Deflorenne E, Aller J, Bourdeau I, Bressac-de Paillerets B, Calatayud M, Dehais C, De Mones Del Pujol E, Elenkova A, Herman P, Kamenický P, Lejeune S, Sadoul JL, Barlier A, Richard S, Favier J, Burnichon N, Gardie B, Dahia PL, Robledo M, Gimenez-Roqueplo A-P. Germline mutations in the new E1' cryptic exon of the *VHL* gene in patients with tumours of von Hippel-Lindau disease spectrum or with paraganglioma. J Med Genet. 2020 Jan 29;jmedgenet-2019-106519.

60. Transcript: EGLN1-201 (ENST00000366641.4) - Summary - Homo_sapiens - Ensembl genome browser 104 [Internet]. [cited 2021 Oct 27]. Available from: http://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000135766;r=1:231363751-231422287;t=ENST00000366641

61. Astuti D, Ricketts CJ, Chowdhury R, McDonough MA, Gentle D, Kirby G, Schlisio S, Kenchappa RS, Carter BD, Kaelin WG, Ratcliffe PJ, Schofield CJ, Latif F, Maher ER.

Mutation analysis of HIF prolyl hydroxylases (PHD/EGLN) in individuals with features of phaeochromocytoma and renal cell carcinoma susceptibility. Endocr Relat Cancer. 2010 Dec 21;18(1):73–83.

62. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 2020 Jan 8;48(D1):D265–8.

63. Percy MJ, Zhao Q, Flores A, Harrison C, Lappin TRJ, Maxwell PH, McMullin MF, Lee FS. A family with erythrocytosis establishes a role for prolyl hydroxylase domain protein 2 in oxygen homeostasis. Proc Natl Acad Sci. 2006 Jan 17;103(3):654–9.

64. Percy MJ, Furlow PW, Beer PA, Lappin TRJ, McMullin MF, Lee FS. A novel erythrocytosis-associated PHD2 mutation suggests the location of a HIF binding groove. Blood. 2007 Sep 15;110(6):2193–6.

65. Song D, Li L, Arsenault PR, Tan Q, Bigham AW, Heaton-Johnson KJ, Master SR, Lee FS. Defective Tibetan PHD2 Binding to p23 Links High Altitude Adaption to Altered Oxygen Sensing. J Biol Chem. 2014 May 23;289(21):14656–65.

66. Bonnin A, Gardie B, Girodon F, Airaud F, Garrec C, Bézieau S, Vignon G, Mottaz P, Labrousse J, Lellouche F. Polyglobulie rare par mutation du gène EGLN1 : à propos d'un cas et revue de la littérature. Rev Médecine Interne. 2020 Mar;41(3):196–9.

67. Transcript: EPAS1-201 (ENST00000263734.5) - Summary - Homo_sapiens - Ensembl genome browser 104 [Internet]. [cited 2021 Oct 27]. Available from: http://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000116016;r=2:46293667-46386697;t=ENST00000263734

68. Hall JE, Lawrence ES, Simonson TS, Fox K. Seq-ing Higher Ground: Functional Investigation of Adaptive Variation Associated With High-Altitude Adaptation. Front Genet. 2020 May 14;11:471.

69. Li C, Li X, Xiao J, Liu J, Fan X, Fan F, Lei H. Genetic changes in the *EPAS1* gene between Tibetan and Han ethnic groups and adaptation to the plateau hypoxic environment. PeerJ. 2019 Oct 28;7:e7943.

70. Percy MJ, Lee FS. Familial erythrocytosis: molecular links to red blood cell control. Haematologica. 2008 Jul 1;93(7):963–7.

71. Percy MJ, Beer PA, Campbell G, Dekker AW, Green AR, Oscier D, Rainey MG, van Wijk R, Wood M, Lappin TRJ, McMullin MF, Lee FS. Novel exon 12 mutations in the HIF2A gene associated with erythrocytosis. Blood. 2008 Jun 1;111(11):5400–2.

72. Kristan A, Debeljak N, Kunej T. Genetic variability of hypoxia-inducible factor alpha (HIFA) genes in familial erythrocytosis: Analysis of the literature and genome databases. Eur J Haematol. 2019;103(4):287–99.

73. Camps C, Petousi N, Bento C, Cario H, Copley RR, McMullin MF, van Wijk R, Ratcliffe PJ, Robbins PA, Taylor JC. Gene panel sequencing improves the diagnostic work-up of patients with idiopathic erythrocytosis and identifies new mutations. Haematologica. 2016 Nov;101(11):1306–18.

74. Gene: EPO (ENSG00000130427) - Summary - Homo_sapiens - Ensembl genome browser 104 [Internet]. Available from: http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000130427;r=7:100720468-100723700;t=ENST00000252723

75. Formenti F, Constantin-Teodosiu D, Emmanuel Y, Cheeseman J, Dorrington KL, Edwards LM, Humphreys SM, Lappin TRJ, McMullin MF, McNamara CJ, Mills W, Murphy JA, O'Connor DF, Percy MJ, Ratcliffe PJ, Smith TG, Treacy M, Frayn KN, Greenhaff PL, Karpe F, Clarke K, Robbins PA, ... WGK. Regulation of human metabolism by hypoxia-inducible factor. Proc Natl Acad Sci U S A. 2010;107(28):12722–7.

76. Vočanec D, Prijatelj T, Debeljak N, Kunej T. Genetic variants of erythropoietin (*EPO*) and EPO receptor genes in familial erythrocytosis. Int J Lab Hematol. 2019 Apr;41(2):162–7.

77. Taylor JC, Martin HC, Lise S, Broxholme J, Cazier J-B, Rimmer A, Kanapin A, Lunter G, Fiddy S, Allan C, Aricescu AR, Attar M, Babbs C, Becq J, Beeson D, Bento C, Bignell P, Blair E, Buckle VJ, Bull K, Cais O, Cario H, Chapel H, Copley RR, Cornall R, Craft J, Dahan K, Davenport EE, Dendrou C, Devuyst O, Fenwick AL, Flint J, Fugger L, Gilbert RD, Goriely A, Green A, Greger IH, Grocock R, Gruszczyk AV, Hastings R, Hatton E, Higgs D, Hill A, Holmes C, Howard M, Hughes L, Humburg P, Johnson D, Karpe F, Kingsbury Z, Kini U, Knight JC, Krohn J, Lamble S, Langman C, Lonie L, Luck J, McCarthy D, McGowan SJ, McMullin MF, Miller KA, Murray L, Németh AH, Nesbit MA, Nutt D, Ormondroyd E, Oturai AB, Pagnamenta A, Patel SY, Percy M, Petousi N, Piazza P, Piret SE, Polanco-Echeverry G, Popitsch N, Powrie F, Pugh C, Quek L, Robbins PA, Robson K, Russo A, Sahgal N, van Schouwenburg PA, Schuh A, Silverman E, Simmons A, Sørensen PS, Sweeney E, Taylor J, Thakker RV, Tomlinson I, Trebes A, Twigg SRF, Uhlig HH, Vyas P, Vyse T, Wall SA, Watkins H, Whyte MP, Witty L, Wright B, Yau C, Buck D, Humphray S, Ratcliffe PJ, Bell JI, Wilkie AOM, Bentley D, Donnelly P, McVean G. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet. 2015 Jul;47(7):717–26.

78. Zmajkovic J, Lundberg P, Nienhold R, Torgersen ML, Sundan A, Waage A, Skoda RC. A Gain-of-Function Mutation in *EPO* in Familial Erythrocytosis. N Engl J Med. 2018 Mar 8;378(10):924–30.

79. Ahmed MH, Ghatge MS, Safo MK. Hemoglobin: Structure, Function and Allostery. In: Hoeger U, Harris JR, editors. Vertebrate and Invertebrate Respiratory Proteins, Lipoproteins and other Body Fluid Proteins [Internet]. Cham: Springer International Publishing; 2020 [cited 2021 Jul 20]. p. 345–82. (Subcellular Biochemistry; vol. 94). Available from: http://link.springer.com/10.1007/978-3-030-41769-7_14

80. Jorge SE, Bringas M, Petruk AA, Arrar M, Marti MA, Skaf MS, Costa FF, Capece L, Sonati MF, Estrin D. Understanding the molecular basis of the high oxygen affinity variant human hemoglobin Coimbra. Arch Biochem Biophys. 2018 Jan;637:73–8.

81. Esquerra RM, Bibi BM, Tipgunlakant P, Birukou I, Soman J, Olson JS, Kliger DS, Goldbeck RA. Role of Heme Pocket Water in Allosteric Regulation of Ligand Reactivity in Human Hemoglobin. Biochemistry. 2016 Jul 26;55(29):4005–17.

82. Nabhani IA, Aneke JC, Verhovsek M, Eng B, Kuo KHM, Rudinskas LC, Waye JS. Novel High Oxygen Affinity Hemoglobin Variant in a Patient with Polycythemia: Hb Kennisis [β85(F1)Phe→Leu (TTT>TTG); HBB: c.258T>G]. Hemoglobin. 2020 Jan 2;44(1):10–2.

83. Beutler E, Westwood B, van Zwieten R, Roos D. GÃT Transition at cDNA nt 110 (K37Q) in the PKLR (Pyruvate Kinase) Gene Is the Molecular Basis of a Case of Hereditary Increase of Red Blood Cell ATP. :4.

84. van Oirschot BA, Francois JJJM, van Solinge WW, van Wesel ACW, Rijksen G, van Amstel HKP, van Wijk R. Novel type of red blood cell pyruvate kinase hyperactivity predicts a remote regulatory locus involved in *PKLR* gene expression: Pyruvate kinase hyperactivity unlinked to *PKLR*. Am J Hematol. 2014 Apr;89(4):380–4.

85. Filser M, Giansily-Blaizot M, Grenier M, Monedero Alonso D, Bouyer G, Pérès L, Egée S, Aral B, Airaud F, Da Costa L, Picard V, Cougoul P, Palach M, Béziau S, Garrec C, Aguilar-Martinez P, Gardie B, Girodon F. Increased incidence of germline *PIEZO1* mutations in individuals with idiopathic erythrocytosis. Blood. 2021 Apr 1;137(13):1828–32.

86. Cahalan SM, Lukacs V, Ranade SS, Chien S, Bandell M, Patapoutian A. Piezo1 links mechanical forces to red blood cell volume. eLife. 2015 May 22;4:e07370.

87. Botello-Smith WM, Jiang W, Zhang H, Ozkan AD, Lin Y-C, Pham CN, Lacroix JJ, Luo Y. A mechanism for the activation of the mechanosensitive Piezo1 channel by the small molecule Yoda1. Nat Commun. 2019 Dec;10(1):4503.

88. Maciak K, Adamowicz-Salach A, Siwicka A, Poznanski J, Urasinski T, Plochocka D, Gora M, Burzynska B. Hereditary xerocytosis - spectrum and clinical manifestations of variants in the PIEZO1 gene, including co-occurrence with a novel β-globin mutation. Blood Cells Mol Dis. 2020 Feb;80:102378.

89. Kiger L, Oliveira L, Guitton C, Bendélac L, Ghazal K, Proulle V, Galacteros F, Junot C, Fenaille F, Roméo P-H, Garçon L, Picard V. Piezo1-xerocytosis red cell metabolome shows impaired glycolysis and increased hemoglobin oxygen affinity. Blood Adv. 2021 Jan 12;5(1):84–8.

90. Transcript: PIEZO1-201 (ENST00000301015.14) - Summary - Homo_sapiens - Ensembl genome browser 104 [Internet]. [cited 2021 Oct 24]. Available from: http://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000103335;r=16:88715338-88785220;t=ENST00000301015

91. Toste S, Relvas L, Pinto C, Bento C, Abade A, Ribeiro ML, Manco L. Intragenic haplotype analysis of common HFE mutations in the Portuguese population. J Genet. 2015 Jun;94(2):329–33.

92. Raphael B, Cooperberg AA, Nillof P. The triad of hemochromatosis, hepatoma and erythrocytosis. AMERICAN CANCER SOCIETY. 1979;43(2):5.

93. Gurnari C, Lombardi AM, Cosi E, Biagetti G, Buccisano F, Franceschini L, Di Veroli A, Falconi G, Fabiani E, Cantonetti M, Nasso D, Lo-Coco F, Randi ML, Voso MT. Genetic analysis of erythrocytosis reveals possible causative and modifier gene mutations. Br J Haematol [Internet]. 2019 Aug [cited 2021 Jul 20];186(4). Available from: https://onlinelibrary.wiley.com/doi/10.1111/bjh.15931

94. Asif S, Begemann M, Raza S. Polycythemia in Patients With Hereditary Hemochromatosis: Real or Myth? J Clin Med Res. 2019;11(6):422–7.

95. Khan AA, Hadi Y, Hassan A, Kupec J. Polycythemia and Anemia in Hereditary Hemochromatosis. Cureus [Internet]. 2020 Apr 9 [cited 2021 Jul 20]; Available from: https://www.cureus.com/articles/29793-polycythemia-and-anemia-in-hereditary-hemochromatosis

96. Langabeer SE. Hemochromatosis, erythrocytosis and the JAK2 p.V617F mutation. 2017 Feb;2.

97. Ruggeri M, Tosetto A, Frezzato M, Rodeghiero F. The Rate of Progression to Polycythemia Vera or Essential Thrombocythemia in Patients with Erythrocytosis or Thrombocytosis. Ann Intern Med. 2003 Sep 16;139(6):470.

98. Kohlmann A, Grossmann V, Nadarajah N, Haferlach T. Next-generation sequencing - feasibility and practicality in haematology. Br J Haematol. 2013 Mar;160(6):736–53.

99. Metzker ML. Sequencing technologies — the next generation. Nat Rev Genet. 2010 Jan;11(1):31–46.

100. McMullin MF. Diagnostic workflow for hereditary erythrocytosis and thrombocytosis. Hematology. 2019 Dec 6;2019(1):391–6.

101. Filser M, Aral B, Airaud F, Chauveau A, Bruce A, Polfrit Y, Thiebaut A, Gauthier M, Le Maréchal C, Lippert E, Béziau S, Garrec C, Gardie B, Girodon F. Low incidence of *EPOR* mutations in idiopathic erythrocytosis. Haematologica. 2020 Mar 12;106(1):299–301.

102. Fidalgo T, Martinho P, Pinto CS, Oliveira AC, Salvado R, Borràs N, Coucelo M, Manco L, Maia T, Mendes MJ, Del Orbe Barreto R, Corrales I, Vidal F, Ribeiro ML. Combined study of ADAMTS13 and complement genes in the diagnosis of thrombotic microangiopathies using next-generation sequencing. Res Pract Thromb Haemost. 2017 Jul;1(1):69–80.

103. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. Fast, scalable generation of high-

quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011 Jan;7(1):539.

104. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. 2013 Jul 1;41(W1):W597–600.

105. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. 2010 Jul 1;38(Web Server):W695–9.

106. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Charkhchi M, Cummins C, Da Rin Fioretto L, Davidson C, Dodiya K, El Houdaigui B, Fatima R, Gall A, Garcia Giron C, Grego T, Guijarro-Clarke C, Haggerty L, Hemrom A, Hourlier T, Izuogu OG, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Gonzalez Martinez J, Marugán JC, Maurel T, McMahon AC, Mohanan S, Moore B, Muffato M, Oheh DN, Paraschas D, Parker A, Parton A, Prosovetskaia I, Sakthivel MP, Salam AIA, Schmitt BM, Schuilenburg H, Sheppard D, Steed E, Szpak M, Szuba M, Taylor K, Thormann A, Threadgold G, Walts B, Winterbottom A, Chakiachvili M, Chaubal A, De Silva N, Flint B, Frankish A, Hunt SE, IIsley GR, Langridge N, Loveland JE, Martin FJ, Mudge JM, Morales J, Perry E, Ruffier M, Tate J, Thybert D, Trevanion SJ, Cunningham F, Yates AD, Zerbino DR, Flicek P. Ensembl 2021. Nucleic Acids Res. 2021 Jan 8;49(D1):D884–91.

107. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015 May;17(5):405–23.

108. Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, Massouras A. VarSome: the human genomic variant search engine. Wren J, editor. Bioinformatics. 2019 Jun 1;35(11):1978–80.

109. Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, Hayden M, Heywood S, Millar DS, Phillips AD, Cooper DN. The Human Gene Mutation Database (HGMD®): optimizing its use in a clinical diagnostic or research setting. Hum Genet. 2020 Oct;139(10):1197–207.

110. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018 Jan 4;46(D1):D1062–7.

111. PROVEAN Human Genome Variants [Internet]. Available from: http://provean.jcvi.org/genome_submit_2.php?species=human

112. SIFT - Predict effects of nonsynonmous / missense variants [Internet]. [cited 2021 Oct 22]. Available from: https://sift.bii.a-star.edu.sg/

113. PolyPhen-2: prediction of functional effects of human nsSNPs [Internet]. Available from: http://genetics.bwh.harvard.edu/pph2/

114. MutationAssessor.org /// functional impact of protein mutations [Internet]. Available from: http://mutationassessor.org/r3/

115. MutationTaster [Internet]. Available from: https://www.mutationtaster.org/

116. MetaDome web server [Internet]. Available from: https://stuart.radboudumc.nl/metadome/dashboard

117. Kister J, Baudin-Creuza V, Kiger L, Préhu C, Papassotiriou I, Riou J, Galactéros F, Wajcman H. Hb Montfermeil [β 130(H8) Tyr→Cys]: suggests a key role for the interaction between helix A and H in oxygen affinity of the hemoglobin molecule. Blood Cells Mol Dis. 2005 Mar;34(2):166–73.

118. Landin B, Berglund S, Wallman K. Two Different Mutations in Codon 97 of the p-Globin Gene Cause Hb Malmo in Sweden. 1996;5.

119. Kristan A, Pajič T, Maver A, Režen T, Kunej T, Količ R, Vuga A, Fink M, Žula Š, Podgornik H, Anžej Doma S, Preložnik Zupan I, Rozman D, Debeljak N. Identification of Variants Associated With Rare Hematological Disorder Erythrocytosis Using Targeted Next-Generation Sequencing Analysis. Front Genet. 2021 Jul 19;12:689868.

120. Toriumi N, Kaneda M, Hatakeyama N, Manabe H, Okajima K, Sakurai Y, Yamamoto M, Sarashina T, Ikuta K, Azuma H. A case of primary familial congenital polycythemia with a novel EPOR mutation: possible spontaneous remission/alleviation by menstrual bleeding. Int J Hematol. 2018 Sep;108(3):339–43.

121. Mallik N, Sharma P, Kaur Hira J, Chhabra S, Sreedharanunni S, Kumar N, Naseem S, Sachdeva MUS, Ahluwalia J, Malhotra P, Varma N, Varma S, Das R. Genetic basis of unexplained erythrocytosis in Indian patients. Eur J Haematol. 2019 Aug;103(2):124–30.

122. Congenital anomalies [Internet]. Available from: https://www.who.int/news-room/fact-sheets/detail/congenital-anomalies

123. Lee G, Arcasoy MO. The clinical and laboratory evaluation of the patient with erythrocytosis. Eur J Intern Med. 2015 Jun;26(5):297–302.

124. Ma W, Kantarjian H, Zhang X, Yeh C-H, Zhang ZJ, Verstovsek S, Albitar M. Mutation Profile of JAK2 Transcripts in Patients with Chronic Myeloproliferative Neoplasias. J Mol Diagn. 2009 Jan;11(1):49–53.

125. Mol CD, Lim KB, Sridhar V, Zou H, Chien EYT, Sang B-C, Nowakowski J, Kassel DB, Cronin CN, McRee DE. Structure of a c-Kit Product Complex Reveals the Basis for Kinase Transactivation. J Biol Chem. 2003 Aug;278(34):31461–4.

126. Kralovics R, Indrak K, Stopka T, Berman BW, Prchal JF, Prchal JT. Two New EPO Receptor Mutations: Truncated EPO Receptors Are Most Frequently Associated With Primary Familial and Congenital Polycythemias. Blood. 1997 Sep 1;90(5):2057–61.

127. Sokol L, Prchal JF, D'Andrea A, Rado TA, Prchal JT. Mutation in the negative regulatory element of the erythropoietin receptor gene in a case of sporadic primary polycythemia. Exp Hematol. 1994 May;22(5):447–53.

128. Thom CS, Dickson CF, Gell DA, Weiss MJ. Hemoglobin Variants: Biochemical Properties and Clinical Correlates. Cold Spring Harb Perspect Med. 2013 Mar 1;3(3):a011858–a011858.

129. Stephanou C, Kountouris P, Tamana S, Minaidou A, Lederer CW, Kleanthous M. ITHANET: An Information and Database Community Portal for Hemoglobinopathies. Hemoglobin. 2019 Nov 2;43(6):363–363.

130. Santbergen B, van der Heul C. At high altitude in the Netherlands: secondary erythrocytosis due to HB-Malmo. Case Rep. 2014 Mar 5;2014(mar05 1):bcr2014203701–bcr2014203701.

131. VCV000374982.9 - ClinVar - NCBI [Internet]. Available from: https://www.ncbi.nlm.nih.gov/clinvar/variation/374982/

132. VCV000093330.22 - ClinVar - NCBI [Internet]. [cited 2021 Oct 12]. Available from: https://www.ncbi.nlm.nih.gov/clinvar/variation/93330/

133. Oliveira JL, Coon LM, Frederick LA, Hein M, Swanson KC, Savedra ME, Porter TR, Patnaik MM, Tefferi A, Pardanani A, Grebe SK, Viswanatha DS, Hoyer JD. Genotype-phenotype correlation of hereditary erythrocytosis mutations, a single center experience. Am J Hematol. 2018 Aug;93(8):1029–41.

134. Olschwang S, Richard S, Boisson C, Giraud S, Laurent-Puig P, Resche F, Thomas G. Germline mutation profile of the VHL gene in von Hippel-Lindau disease and in sporadic hemangioblastoma. :7.

135. Hes F, Zewald R, Peeters T, Sijmons R, Links T, Verheij J, Matthijs G, Legius E, Mortier G, van der Torren K, Rosman M, Lips C, Pearson P, van der Luijt R. Genotype-phenotype correlations in families with deletions in the von Hippel-Lindau (VHL) gene. Hum Genet. 2000 Apr 25;106(4):425–31.

136. Takeda K, Aguila HL, Parikh NS, Li X, Lamothe K, Duan L-J, Takeda H, Lee FS, Fong G-H. Regulation of adult erythropoiesis by prolyl hydroxylase domain proteins. Blood. 2008 Mar 15;111(6):3229–35.

137. Fong G-H, Takeda K. Role and regulation of prolyl hydroxylase domain proteins. Cell Death Differ. 2008 Apr;15(4):635–41.

138. Al-Sheikh M, Mazurier E, Gardie B, Casadevall N, Galacteros F, Goossens M, Wajcman H, Prehu C, Ugo V. A study of 36 unrelated cases with pure erythrocytosis

revealed three new mutations in the erythropoietin receptor gene. Haematologica. 2008 Jan 1;93(7):1072–5.

139. Percy MJ, Furlow PW, Beer PA, Lappin TRJ, McMullin MF, Lee FS. A novel erythrocytosis-associated PHD2 mutation suggests the location of a HIF binding groove. Blood. 2007 Sep 15;110(6):2193–6.

140. Q99814 (Q99814) - protein - InterPro [Internet]. [cited 2021 Oct 28]. Available from: https://www.ebi.ac.uk/interpro/protein/UniProt/Q99814/

141. CDD Conserved Protein Domain Family: bHLH_SF [Internet]. Available from: https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi?uid=412148

142. Lorenzo FR, Yang C, Ng Tang Fui M, Vankayalapati H, Zhuang Z, Huynh T, Grossmann M, Pacak K, Prchal JT. A novel EPAS1/HIF2A germline mutation in a congenital polycythemia with paraganglioma. J Mol Med. 2013 Apr;91(4):507–12.

143. Novel human pathological mutations. Hum Genet. 2009 Apr;125(3):333–52.

144. Wild CP. Complementing the Genome with an "Exposome": The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. Cancer Epidemiol Biomarkers Prev. 2005 Aug 1;14(8):1847–50.

145. Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. The Blood Exposome and Its Role in Discovering Causes of Disease. Environ Health Perspect. 2014 Aug;122(8):769–74.

146. Vermeulen R, Schymanski EL, Barabási A-L, Miller GW. The exposome and health: where chemistry meets biology. Science. 2020 Jan 24;367(6476):392–6.

147. Jang J-H, Seo JY, Jang J, Jung CW, Lee K-O, Kim S-H, Kim H-J. Hereditary gene mutations in Korean patients with isolated erythrocytosis. Ann Hematol. 2014 Jun;93(6):931–5.

148. Peroni E, Bertozzi I, Gherlinzoni F, Stefani PM, Lombardi A, Biagetti G, Fabris F, Randi ML. Two novel missense mutations in EPOR gene causes erythrocytosis in two unrelated patients. Br J Haematol. 2018;180(6):908–11.

149. Arcasoy MO, Karayal AF, Segal HM, Sinning JG, Forget BG. A novel mutation in the erythropoietin receptor gene is associated with familial erythrocytosis. Blood. 2002 Apr 15;99(8):3066–9.

150. O'Rourke K, Fairbairn DJ, Jackson KA, Morris KL, Tey S-K, Kennedy GA. A novel mutation of the erythropoietin receptor gene associated with primary familial and congenital polycythaemia. Int J Hematol. 2011 Apr;93(4):542–4.

151. Perrotta S, Cucciolla V, Ferraro M, Ronzoni L, Tramontano A, Rossi F, Scudieri AC, Borriello A, Roberti D, Nobili B, Cappellini MD, Oliva A, Amendola G, Migliaccio AR, Mancuso P, Martin-Padura I, Bertolini F, Yoon D, Prchal JT, Della Ragione F. EPO Receptor Gain-of-Function Causes Hereditary Polycythemia, Alters CD34+ Cell

Differentiation and Increases Circulating Endothelial Precursors. Ng IOL, editor. PLoS ONE. 2010 Aug 5;5(8):e12015.

152. Brorson Petersen K, Hokland P, Bruun Petersen G, Guldborg Nyvold C. Erythropoietin receptor defect: a cause of primary polycythaemia. Br J Haematol. 2004 May;125(4):537–8.

153. Kralovics R, Prchal JT. Genetic heterogeneity of primary familial and congenital polycythemia. Am J Hematol. 2001 Oct;68(2):115–21.

154. Kralovics R, Sokol L, Prchal JT. Absence of polycythemia in a child with a unique erythropoietin receptor mutation in a family with autosomal dominant primary polycythemia. J Clin Invest. 1998 Jul 1;102(1):124–9.

155. Watowich SS, Xie X, Klingmuller U, Kere J, Lindlof M, Berglund S, de la Chapelle A. Erythropoietin Receptor Mutations Associated With Familial Erythrocytosis Cause Hypersensitivity to Erythropoietin in the Heterozygous State. Blood. 1999 Oct 1;94(7):2530–2.

156. Arcasoy MO, Degar BA, Harris KW, Forget BG. Familial Erythrocytosis Associated With a Short Deletion in the Erythropoietin Receptor Gene. Blood. 1997 Jun 15;89(12):4628–35.

157. Furukawa T, Narita M, Sakaue M, Otsuka T, Kuroha T, Masuko M, Azegami T, Kishi K, Takahashi M, Utsumi J, Koike T, Aizawa Y. Primary familial polycythaemia associated with a novel point mutation in the erythropoietin receptor. Br J Haematol. 1997 Oct;99(1):222–7.

158. de la Chapelle A, Traskelin AL, Juvonen E. Truncated erythropoietin receptor causes dominantly inherited benign human erythrocytosis. Proc Natl Acad Sci. 1993 May 15;90(10):4495–9.

159. Rives S, Pahl HL, Florensa L, Bellosillo B, Neususess A, Estella J, Debatin K-M, Kohne E, Schwarz K, Cario H. Molecular genetic analyses in familial and sporadic congenital primary erythrocytosis. Haematologica. 2007 May 1;92(5):674–7.

160. Chauveau A, Luque Paz D, Lecucq L, Le Gac G, Le Maréchal C, Gueguen P, Berthou C, Ugo V. A new point mutation in *EPOR* inducing a short deletion in congenital erythrocytosis. Br J Haematol. 2016 Feb;172(3):475–7.

161. Le Couedic J, Mitjavila M, Villeval J, Feger F, Gobert S, Mayeux P, Casadevall N, Vainchenker W. Missense mutation of the erythropoietin receptor is a rare event in human erythroid malignancies. Blood. 1996 Feb 15;87(4):1502–11.

162. Bento MC, Chang KT, Guan Y, Liu E, Caldas G, Gatti RA, Prchal JT. Congenital polycythemia with homozygous and heterozygous mutations of von Hippel-Lindau gene: five new Caucasian patients. Haematologica. 2005 Jan 1;90(1):128–9.

163. Cario H, Schwarz K, Jorch N, Kyank U, Petrides PE, Schneider DT, Uhle R, Debatin K-M, Kohne E. Mutations in the von Hippel-Lindau (VHL) tumor suppressor gene and

VHL-haplotype analysis in patients with presumable congenital erythrocytosis. Haematologica/the hematology journal. 2005;6.

164. Lorenzo FR, Yang C, Lanikova L, Butros L, Zhuang Z, Prchal JT. Novel compound VHL heterozygosity (*VHL* T124A/L188V) associated with congenital polycythaemia. Br J Haematol. 2013 Sep;162(6):851–3.

165. Bond J, Gale DP, Connor T, Adams S, de Boer J, Gascoyne DM, Williams O, Maxwell PH, Ancliff PJ. Dysregulation of the HIF pathway due to VHL mutation causing severe erythrocytosis and pulmonary arterial hypertension. Blood. 2011 Mar 31;117(13):3699–701.

166. Pastore YD, Jelinek J, Ang S, Guan Y, Liu E, Jedlickova K, Krishnamurti L, Prchal JT. Mutations in the VHL gene in sporadic apparently congenital polycythemia. Blood. 2003 Feb 15;101(4):1591–5.

167. Pastore Y, Jedlickova K, Guan Y, Liu E, Fahner J, Hasle H, Prchal JF, Prchal JT. Mutations of von Hippel-Lindau Tumor-Suppressor Gene and Congenital Polycythemia. Am J Hum Genet. 2003 Aug;73(2):412–9.

168. Percy MJ, Furlow PW, Jones FGC, Lappin TRJ, Lee FS, McMullin MF. Erythrocytosis caused by mutations in the PHD2 and VHL genes. In: ASH Annual Meeting. 2007. p. Abstracts 110:3663.

169. Lorenzo FR, Simonson TS, Yang Y, Ge R, Prchal JT. Novel PHD2 mutation associated with Tibetan genetic adaptation to high altitude hypoxia. In: ASH 52nd Annual Meeting (ASH,. Orlando, FL.; 2010.

170. Ladroue C, Hoogewijs D, Gad S, Carcenac R, Storti F, Barrois M, Gimenez-Roqueplo A-P, Leporrier M, Casadevall N, Hermine O, Kiladjian J-J, Baruchel A, Fakhoury F, Bressac-de Paillerets B, Feunteun J, Mazure N, Pouyssegur J, Wenger RH, Richard S, Gardie B. Distinct deregulation of the hypoxia inducible factor by PHD2 mutants identified in germline DNA of patients with polycythemia. Haematologica. 2012 Jan 1;97(1):9–14.

171. Albiero E, Ruggeri M, Fortuna S, Finotto S, Bernardi M, Madeo D, Rodeghiero F. Isolated erythrocytosis: study of 67 patients and identification of three novel germ-line mutations in the prolyl hydroxylase domain protein 2 (PHD2) gene. Haematologica. 2012 Jan 1;97(1):123–7.

172. Schelker RC, Herr W, Grassinger J. A new exon 12 mutation in the *EPAS1* gene possibly associated with erythrocytosis. Eur J Haematol. 2019 Jul;103(1):64–6.

173. Perrotta S, Stiehl DP, Punzo F, Scianguetta S, Borriello A, Bencivenga D, Casale M, Nobili B, Fasoli S, Balduzzi A, Cro L, Nytko KJ, Wenger RH, Della Ragione F. Congenital erythrocytosis associated with gain-of-function HIF2A gene mutations and erythropoietin levels in the normal range. Haematologica. 2013 Oct 1;98(10):1624–32.

174. Furlow PW, Percy MJ, Sutherland S, Bierl C, McMullin MF, Master SR, Lappin TRJ, Lee FS. Erythrocytosis-associated HIF-2α Mutations Demonstrate a Critical Role for Residues C-terminal to the Hydroxylacceptor Proline. J Biol Chem. 2009 Apr;284(14):9050–8.

175. Percy MJ, Chung YJ, Harrison C, Mercieca J, Hoffbrand AV, Dinardo CL, Santos PCJL, Fonseca GHH, Gualandro SFM, Pereira AC, Lappin TRJ, McMullin MF, Lee FS. Two new mutations in the HIF2A gene associated with erythrocytosis. Am J Hematol. 2012 Apr;87(4):439–42.

176. Percy MJ, Furlow PW, Lucas GS, Li X, Lappin TRJ, McMullin MF, Lee FS. A Gain-of-Function Mutation in the HIF2A Gene in Familial Erythrocytosis. N Engl J Med. 2008 Jan 10;358(2):162–8.

177. Martini M, Teofili L, Cenci T, Giona F, Torti L, Rea M, Foa R, Leone G, Larocca LM. A novel heterozygous HIF2AM535I mutation reinforces the role of oxygen sensing pathway disturbances in the pathogenesis of familial erythrocytosis. Haematologica. 2008 Jan 1;93(7):1068–71.

178. van Wijk R, Sutherland S, Van Wesel ACW, Huizinga EG, Percy MJ, Bierings M, Lee FS. Erythrocytosis associated with a novel missense mutation in the HIF2A gene. Haematologica. 2010 May;95(5):829–32.

179. Lemarchandel V, Joulin V, Valentin C, Rosa R, Galacteros F, Rosa J, Cohen- Solal M. Compound heterozygosity in a complete erythrocyte bisphosphoglycerate mutase deficiency. Blood. 1992 Nov 15;80(10):2643–9.

180. Hoyer JD, Allen SL, Beutler E, Kubik K, West C, Fairbanks VF. Erythrocytosis due to bisphosphoglycerate mutase deficiency with concurrent glucose-6-phosphate dehydrogenase (G-6-PD) deficiency. Am J Hematol. 2004 Apr;75(4):205–8.

181. McMullin MF, Wu C, Percy MJ, Tong W. A nonsynonymous LNK polymorphism associated with idiopathic erythrocytosis. Am J Hematol. 2011 Nov;86(11):962–4.

9. Supplementary Material

Supplementary Table 1: Previously described and novel *EPOR* gene pathogenic variants.

***EPOR* pathogenic variants (ECYT1)**			
Exon	**Nucleotide exchanged**	**Protein effect**	**References**
5	c.634G>T	p.Gly212Cys	Jang *et al.*[147]
8	c.1013G>A	p.Cys338Tyr	Peroni *et al.* [148]
8	c.1023C>T	p.Thr341Met	Peroni *et al.* [148]
8	c.1138C>G	p.Pro380Ala	Bento *et al.*[38]
8	c.1141_1142delCC	p.Pro381Glnfs*2	Al-Sheikh *et al.* [138]
8	c.1195G>T	p.Glu399*	Arcasoy *et al.* [149]
8	c.1234delT	p.Ser412Argfs*41	O'Rourke *et al.* [150]
8	c.1235C>A	p.Ser412*	Bento *et al.*[38]
8	c.1242_1276del35	p.Ser415Hisfs*18	Bento *et al.*[38]
8	c.1249G>T	p.Glu417*	Perrotta *et al.* [151]
8	c.1252_1255delGGAG	p.G418Profs*34	Brorson *et al.* [152]
8	c.1271_1272delTT	p.Phe424*	Al-Sheikh *et al.* [138]
8	c.1273G>T	p.Glu425*	Kralovics *et al.* [153]
8	c.1278C>G	p.Tyr426*	Kralovics *et al.*[154]
8	c.1281dupT	p.Ile428Tyrfs*17	Kralovics *et al.* [126]
8	c.1282_1289dupATCCTGGA	p.Asp430Glufs*26	Watowich *et al.* [155]
8	c.1285delC	p.Leu429Trpfs*24	Al-Sheikh *et al.* [138]
8	c.1288dupG	p.Asp430Glyfs*15	Sokol *et al*. [127]
8	c.1299_1305delCCAGCTC	p.Gln434Cysfs*17	Arcasoy *et al.*[156]
8	c.1300C>T	p.Gln434*	*Furukawa et al.* [157]
8	c.1311_1312delTC	p.Pro438Metfs*6	Bento *et al.*[38]
8	c.1310G>A	p.Arg437His	Bento *et al.*[38]
8	c.1316G>A	p.Trp439*	de la Chapelle *et al*. [158]
8	c.1317G>A	p.Trp439*	Rives *et al.* [159]
8	c.1362C>G	p.Tyr454*	Chauveau *et al.* [160]
8	c.1460A>G	p.Ans487Ser	Le Couedic *et al.* [161]
8	c.1462C>T	p.Pro488Ser	Sokol *et al*. [127]

*Stop Codon

Supplementary Table 2: Previously described and novel *VHL* pathogenic variants.

***VHL* pathogenic variants (ECYT2)**			
Exon	**Nucleotide exchanged**	**Protein effect**	**References**
1	c.28G>T	p.Glu10*	Vainchenker reported by Bento *et al.* [5]
1	c.235C>T	p.Arg79Cys	Bento *et al.* [162]
1	c.241C>G	p.Pro81Ala	Casadevall reported by Bento *et al.* [5]
1	c.311G>T	p.Gly104Val	Cario *et al.* [163]
1'	c .340 + 574A>T		Lenglet *et al.* [53]
1'	c.340 + 694_711dup	---------	Lenglet *et al.* [53]
1'	c.340+770T>C	---------	Lenglet *et al.* [53]
1'	c.340 + 816A> C	----------	Lenglet *et al.* [53]
2	c.370A>G	p.Thr124Lys	Lorenzo *et al*. [164]
2	c.376G>A	p.Asp126Asn	Bond *et al.* [165]
2	c.376G>T	p.Asp126Tyr	Pastore *et al.* [166]
2	c.388G>C	p.Val130Leu	Pastore *et al.* [166]
2	c.413C>T	Pro138Leu	Lanikova *et al.* 2013
2	c.430G>A	p.Gly144Arg	Randi *et al*. 2005
3	c.524A>G	p.Tyr175Cys	Bento *et al.*[162]
3	c.548C>T	p.Ser183Leu	Bond *et al.* [165]
3	c.562C>G	p.Leu188Val	Pastore *et al.* [167]
3	c.571C>T	p.His191Asp	Pastore *et al.* [167]
3	c.574C>T	p.Pro192Ser	Pastore *et al.* [167]
3	c.574C>A	p.Pro192Thr	Percy *et al.* [168]
3	c.586A>G	p.Lys196Glu	Bento *et al.* [162]
3	c.598C>T	p.Arg200Trp	Ang *et al.* [57]

*Stop Codon

Supplementary Table 3: Previously described and novel *EGLN1* pathogenic variants.

***EGLN1* (PHD2) pathogenic variants (ECYT3)**			
Exon	**Nucleotide exchanged**	**Protein effect**	**References**
1	c.12C>A	p.Asp4Glu	Lorenzo *et al*. [169]
1	c.380G>	p.Cys127Ser	Lorenzo *et al*. [169]
1	c.400C>T	p. Gln134*	Bonnin *et al*. [66]
1	c.599C>A	p.Pro200Gln	Ladroue *et al*. [170]
1	c.606delG	p.Met202Ilefs∗72	Al-Sheikh *et al.* [138]
1	c.609C>G	p.Asn203Lys	Albiero *et al*. [171]
1	c.610G>A	p.Lys204Glu	McMullin reported by Bento *et al*. [5]
1	c.760G>C	p.Asp254His	Ladroue *et al*. [170]
1	c.840dupA	p.Arg281Thrfs∗4	Al-Sheikh *et al.* [138]
1	c.853G>C	p.Gly285Arg	McMullin reported by Bento *et al.* [5]
1	c.872A>T	p.Lys291Ile	Albiero *et al*. [171]
2	c.950C>G	p.Pro317Arg	Percy *et al.* [63]
2	c.1000T>C	p.Trp334Arg	Bento *et al*. [38]
3	c.1010dup	p.Val338Glyfs*18	McMullin reported by Bento *et al.* [5]
3	c.1112G>A	p.Arg371His	Percy *et al.* [63]
3	c.1121A>G	p.His374Arg	Ladroue *et al*. [170]
3	c.1129C>T	p.Gln377*	Al-Sheikh *et al.* [138]
4	c.1192C>T 4	p.Arg398*	Ladroue *et al*. [170]
5	c.1267A>G	p.Lys423Glu	Albiero *et al*. [171]

*Stop Codon

Supplementary Table 4: Previously described and novel *EPAS1* pathogenic variants.

***EPAS1* (HIF2a) pathogenic variants (ECYT4)**			
Exon	**Nucleotide exchanged**	**Protein effect**	**References**
2	c.47_49delAGG	p.Lys16_Glu17del	Camps *et al*. [47]
2	c.49_51delGAG	p.Glu17del	Camps *et al*. [47]
9	c.1121T>A	p.Phe374Tyr	Lorenzo *et al*. [142]
12	c.1573G>C	p.Asp525His	Schelker *et al*. [172]
12	c.1594T>C	p.Tyr532His	Camps *et al*. [47]
12	c.1597A>G	p.Ile533Val	Perrotta *et al* [173]
12	c.1601C>T	p.Pro534Leu	Furlow *et al*. [174]
12	c.1601C>G	p.Pro534Arg	Oliveira *et al*.[133]
12	c.1603A>G	p.Met535Leu	Oliveira *et al*. [133]
12	c.1603A>G	p.Met535Val	Percy *et al.*[175]
12	c.1604T>C	p.Met535Thr	Percy *et al.* [176]
12	c.1605G>A	p.Met535Ile	Martini *et al*. [177]
12	c.1609G<A	p.Gly537Arg	Percy *et al.* [176]
12	c.1609G>T	p.Gly537Trp	Percy *et al.* [176]
12	c.1609G>C	p.Gly537Arg	Oliveira *et al*. [133]
12	c.1615G>A	p.Asp539Asn	Oliveira *et al*. [133]
12	c.1617C>G	p.Asp539Glu	vanWijk *et al*. [178]
12	c.1620C>G	p.Phe540Leu	Percy *et al.* [175]
12	c.1620C>G	p.Phe540Leu	Percy *et al.* [175]
12	c.1631C>G	p.Pro544Arg	Oliveira *et al*. [133]
13	c.2120A>C	p.Lys707Thr	Kristan *et al.* [119]
16	c.2465T>C	p.Met822Thr	Camps *et al*. [73]

Supplementary Table 5: Previously described and novel *EPO* pathogenic variants.

***EPO* pathogenic variants (ECYT5)**			
Exon	**Nucleotide exchanged**	**Protein effect**	**References**
5'UTR	c. -136G>A	----	Zmajkovic *et al* [78]
2	c.32delG	p.Trp11CysfsTer52	Zmajkovic *et al* [78]
2	c.19delC	p.Pro7LeufsTer56	Taylor *et al.* [77]

Supplementary Table 6: Previously described and novel *BPGM* pathogenic variants.

***BPGM* pathogenic variants (ECYT8)**			
Exon	**Nucleotide exchanged**	**Protein effect**	**References**
3	c.61delC	p.Arg21Valfs*28	Lemarchandel *et al*.[179]
3	c.185G>A	p.Arg62Gln	Hoyer *et al.*[180]
3	c.268C>T	p.Arg90Cys	Lemarchandel *et al*.[179]
3	c.268G>A	p.Arg90His	Petousi *et al.*[26]
3	c.304C>A	p.Gln102Lys	Camps *et al.* [73]

*Stop Codon

Supplementary Table 7: Previously described and novel *SH2B3* gene pathogenic variants.

***SH2B3* (LNK) pathogenic variants**			
Exon	**Nucleotide exchanged**	**Protein effect**	**References**
2	c.232G>A	p.Glu78Lys	Camps *et al.* [73]
2	c.622G>C	p.Glu208Gln	Camps *et al.* [73]
4	c.901G>A	p.Glu301Lys	Camps *et al.* [73]
6	c.1198G>A	p.Glu400Lys	McMullin *et al.*[181]
7	c.1244G>C	p.Arg415Pro	Camps *et al.* [73]

Supplementary Table 8: Primers *HBB, JAK2, SH2B3, VHL* genes, used for PCR and Sanger sequencing.

Gene	Exon	Primer	Sense	T°C Annealing
HBB	1	5'- GAGCCAAGGACAGGTACGG-3' 5'- CAAAGGACTCAAAGAACCTC-3'	Forward*/** Reverse *	54°C
HBB	2	5'-AGACTCTTGGGTTTCTGA-3' 5'CATTCGTCTGTTTCCCATTCTA-3'	Forward*/** Reverse *	54°C
HBB	3	5'-CAATGTATCATGCCTCTTTGCACC-3' 5'-GCAGCCTCACCTTCTTTCATGG-3'	Forward* Reverse */**	54°C
JAK2	14	5'- TCCTCAGAACGTTGATGGCAG-3' 5'- ATTGCTTTCCTTTTTCACAAGAT-3'	Forward* Reverse */**	58°C
JAK2	20	5'-GACAGTCTGCTAATTCCAGC-3' 5'-CTGGGCATTGGCATAAGTC-3'	Forward* Reverse */**	60°C
SH2B3 (LNK)	2	5' - CCAGCACTGGGTGTTATG - 3' 5' - CAGCTGGAAAGCCATCAC - 3'	Forward*/** Reverse */**	56°C
VHL	INTRON E'1	5' - GGCCGAATAGTTTGCATTTGAA – 3' 5' - TAAGCATCCTCTCCCAAGACT – 3'	Forward* Reverse */**	58°C
VHL	1	5'– AGCGCGTTCCATCCTCTAC – 3' 5'– GCTTCAGACCGTGCTATCGT – 3'	Forward*/** Reverse *	62°C

*used in PCR; **used in Sanger sequencing.

Supplementary Table 9: PCR reactions for *HBB, JAK2, SH2B3, VHL* genes

HBB exons 1-2-3 *JAK2* exons 14-20		*SH2B3* exon 2		*VHL* exon 1		*VHL* intron E1'	
Reagent [Concentration]	Volume µL	Reagent [Concentration]	Volume µL	Reagent [Concentration]	Volume µL	Reagent [Concentration]	Volume µL
dH_2O	21	dH_2O	14	dH_2O	4.5	dH_2O	21
Qiagen buffer [10x]	2.5	Qiagen buffer [10x]	2.5	Master Mix [2x]	6.25	*Glucose buffer [10x]	2.5
dNTPs [10 mM]	0.50	Q solution [5x]	5	Q solution [5x]	1.25	dNTPs [10 mM]	0.50
Forward primer [10 mM]	0.50	dNTPs [10 mM]	0.50	Forward primer [10 mM]	0.50	Forward primer [10 mM]	0.50
Reverse primer [10 mM]	0.50	Forward primer [10 mM]	0.75	Reverse primer [10 mM]	0.50	Reverse primer [10 mM]	0.50
Taq polymerase 5 U/µL	0.2	Reverse primer [10 mM]	0.75			Taq polymerase 5 U/µL	0.2
		Taq polymerase 5 U/µL	0.2				

*Glucose buffer [10x]: Tris pH8.8 6.7 mM, $(NH_4)_2SO_4$ 1.66 mM, Na_2 EDTA 6.7 µM, $MgCl_2$ 2.5 mM, BSA 16 µg/mL, B-Me 10 mM.

Projecto de investigação: Caracterização clínica e molecular de doentes com Eritrocitose Congénita – criação de um registo europeu e pesquisa de variantes patogénicas nos genes associados.
Investigadores responsáveis: Doutora Leticia Ribeiro e Dra. Celeste Bento, do Serviço de Hematologia do Centro Hospitalar e Universitário de Coimbra.