UNIVERSITY OF NAPLES FEDERICO II

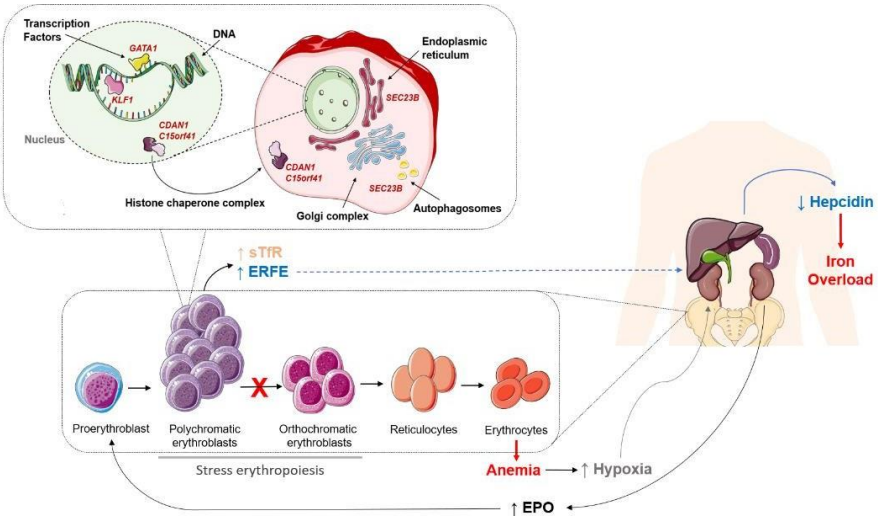
DOCTORATE IN

MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY

XXXIII CYCLE

Roberta Marra

Dissecting the molecular genetics and pathogenesis of Hereditary Dyserythropoietic Anemias



Year 2021

UNIVERSITY OF NAPLES FEDERICO II

DOCTORATE IN

MOLECULAR MEDICINE AND MEDICAL BIOTECHNOLOGY

XXXIII CYCLE

Dissecting the molecular genetics and pathogenesis of Hereditary Dyserythropoietic Anemias

Tutor

Prof. Achille Iolascon

Candidate

Roberta Marra

Co-Tutor

Dr. Roberta Russo

Year 2021

Table of contents

List of Abbreviations ***6***

Abstract ***9***

1. **Background** *11*

1.1 Normal and ineffective erythropoiesis *11*

Biomarkers of dyserythropoiesis *13*

1.2 Classification, diagnostic criteria, and clinical management of CDAs *18*

1.3 From classical CDAs to non-classical forms *24*

CDA type I *24*

CDAN1 *27*

C15orf41 (CDIN1) *28*

CDAII *30*

CDA III and CDA variants *31*

Syndromic CDAs *32*

1.4 Hereditary hemolytic anemia due to Pyruvate Kinase Deficiency *33*

1.5 Differential diagnosis of CDAs and other hereditary hemolytic anemias *34*

1.6 NGS as a diagnostic tool for hereditary anemias *37*

2. Aims of the study **40**

3. Materials and Methods **41**

3.1 Patients and genomic DNA preparation *41*

3.2 Targeted-NGS of HA patients *41*

Libraries establishment *41*

Sequencing and data analysis *42*

Gene Ranking *43*

3.3 Whole Exome Sequencing *43*

Library preparation and sequencing *43*

Bioinformatic analysis of sequencing data 44
3.4 Plasma collection and ELISA assay 44
3.5 Cloning of C15orf41 coding sequence product into expression vector and site-directed mutagenesis 45
3.6 Isolation of mononuclear cells, reticulocytes, and red blood cells membrane from peripheral blood 45
3.7 Gene and protein expression analysis 46
RNA isolation, reverse transcription, and quantitative real-time PCR analysis 46
Protein isolation, subcellular fractionation, and western blotting analysis 46
3.8 Cell cultures 47
Stable clones production 47
Drug treatment 48
3.9 Flow cytometry 49
3.10 Immunofluorescence analysis 49
3.11 Statistical and bioinformatics analyses 50
4. Results **51**
4.1 NGS-based genetic testing improves diagnostic yield of HAs 51
4.1.1 NGS-based genetic testing for hereditary anemias 51
4.1.2 Misdiagnosis of PKD patients as CDAs is due to PKD dyserythropoietic component 55
4.2 Description of novel cases of CDAIb and syndromic CDA 57
4.2.1 Uridine treatment restores the CDA II-like hematological phenotype in a patient with homozygous mutation in the CAD gene 57
4.2.2 Identification of novel causative mutations in the uncharacterized C15orf41 gene 61
4.3 Characterization of C15orf41 in physiological and pathological erythropoiesis 65
4.3.1 C15orf41 and CDAN1 gene expression are directly correlated 65

4.3.3 C15orf41 localization into nuclear and cytosolic compartments *67*
4.3.4 Establishment of K562 C15orf41-WT and mutant stable clones *70*
4.3.5 C15orf41 is regulated by proteasome degradation *72*
4.3.6 C15orf41 is expressed during G1/S phase of cell-cycle *74*
4.3.7 Characterization of C15orf41 expression in Y94S, H230P, and P20T mutants *76*
4.3.8. C15orf41-Y94S and -H230P mutations affect erythroid differentiation and cell cycle dynamics *78*
5. Discussion **80**
6. Conclusions **88**
7. Acknowledgments **89**
8. List of publications **90**
9. References **93**

List of abbreviations

ACMG: American college of medical genetics and genomics

ARC: absolute reticulocyte count

BasoE: basophilic erythroblast

BFU-E: erythroid burst-forming unit

BLAST: basic local alignment search tool

BM: bone marrow

CAF-1: chromatin assembly factor 1

CDA: congenital dyserythropoietic anemia

CDIN1: Codanin-1 Interacting Nuclease 1

CFU-E: erythroid colony-forming unit

DBA: Diamond-Blackfan anemia

DHS: dehydrated hereditary stomatocytosis

EM: electron microscopy

EPO: erythropoietin

EPOr: erythropoietin receptor

ERFE: erythroferrone

GDF11: growth differentiation factor 11

GDF15: growth differentiation factor 15

gDNA: genomic DNA

GM-CSF: granulocyte-macrophage colony-stimulating factor

HA: hereditary anemia

HAMP: hepcidin

Hb: hemoglobin

HC: healthy control

HHA: hereditary hemolytic anemia

HIF2α: Hypoxia Inducible factor alpha 2

HS: hereditary spherocytosis

HSPC: hematopoietic stem and progenitor cells

HSt: hereditary stomatocytosis

HUDEP-2: human umbilical cord blood-derived erythroid progenitors

IE: ineffective erythropoiesis

IF: immunofluorescence

IFNα: interferon alpha

IL-3: interleukin 3

IL-9: interleukin 9

IREB2: iron responsive element binding protein 2

LDH: high lactate dehydrogenase

MAF: minor allele frequency

MCV: mean cell volume

MKLP1: mitotic kinesin-like protein 1

NGS: next generation sequencing

NPM: nucleophosmin

OD: optical density

OMIM: Online Mendelian Inheritance in Man

OrthoE: orthochromatic erythroblast

PBMC: peripheral blood mononuclear cells

PFA: paraformaldehyde

PK: pyruvate kinase

PKD: pyruvate kinase deficiency

PolyE: polychromatophilic erythroblast

ProE: proerythroblast

PTM: post translational modification

RBC: red blood cells

ROS: reactive oxygen species

RP: ribosomal protein

RVIS: residual variation intolerance score

SCF: stem cell factor

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel

SNV: single-nucleotide variant

sTfR: soluble transferrin receptor

TBP: TATA binding protein

TfR: transferrin receptor

T-NGS: targeted-next generation sequencing

TPD: Treponema Pallidum domain

UTR: untranslated region

WB: western blotting

WES: whole exome sequencing

WGS: whole genome sequencing

WT: wild type

XLTDA: X-linked thrombocytopenia with or without dyserythropoietic anemia

Abstract

Hereditary anemias (HAs) embrace a heterogeneous group of chronic disorders with a highly variable clinical picture. Within HAs, congenital dyserythropoietic anemias (CDAs) are a large group of hypo-productive anemias that result from various kinds of abnormalities during late stages of erythropoiesis. Among them, CDAI is characterized by relative reticulocytopenia, and congenital anomalies. It is caused by biallelic mutations in *CDAN1* and *C15orf41*. Differential diagnosis, classification, and patient stratification of CDAs and related HAs are often difficult, particularly between CDAI-II and enzymatic defects, such as pyruvate kinase deficiency (PKD). The classical diagnostic workflow for these conditions includes different lines of investigation, in which genetic testing by next generation sequencing (NGS) approaches has become the frontline system. Indeed, the primary aim of this study was to analyze a large cohort of HAs patients (n=244), by our (t)-NGS RedPanel, to identify the proper molecular diagnosis despite their clinical suspicion. Indeed, only 16.3% of patients originally suspected to suffer from CDA (14/86) showed a matched genotype. Conversely, 64% of patients (72/86) initially suspected for CDA were diagnosed as other HAs, mainly PKD. In agreement with this observation, the analysis of the main erythroid markers demonstrated that PKD patients showed a dyserythropoietic component that may underlie the frequent misdiagnosis with CDAI-II.

Beyond achieving a definitive diagnosis, knowing the genetic basis of these patients is valuable also for guiding treatment. Indeed, in our cohort of patients, we identified a novel case of syndromic CDA due to a novel variant in *CAD* gene, leading to a specific treatment with uridine supplementation. Finally, we described three cases of CDAI, identifying two novel variants in the DNA binding domain of C15orf41, Y94S and P20T, and another one in the nuclease domain of the protein, H230P. Functional characterization of these variants

showed that the H230P leads to reduced gene expression and protein levels, while Y94S and P20T do not affect C15orf41 expression. Moreover, Y94S and H230P variants accounted for impaired erythroid differentiation in K562 cells, and H230P mutant also exhibits an increased S-phase of the cell cycle. Nowadays, *C15orf41* is still an uncharacterized gene, encoding a protein with an unknown function. Thus, we aimed to unravel novel insights on its physiological role. Indeed, we demonstrated that C15orf41 endogenous protein exhibits nuclear and cytosolic localization, being mostly in the nucleus. Our data showed that C15orf41 is a cell-cycle regulated protein, mostly expressed during G1/S phase, and that both the predicted isoforms of the protein are degraded by the ubiquitin-proteasome pathway. Finally, we demonstrated that gene expression of *C15orf41* and *CDAN1*, the other causative gene of CDAI, is tightly correlated, suggesting a shared mechanism of regulation between the two genes.

Overall, these studies pointed out the relevance of genetic testing for the achievement of a correct and definitive diagnosis of CDAs and the related HAs, for the treatment of these conditions, and for elucidating the underlying pathogenic mechanisms of such rare disorders.

1. Background

1.1 Normal and ineffective erythropoiesis

Circulating erythrocytes (or red blood cells, RBCs) are enucleated cells generated from hematopoietic stem and progenitor cells (HSPCs) through the stepwise process of differentiation known as erythropoiesis. Erythropoiesis occurs in the liver during fetal life, while in the post-natal era it is confined to the bone marrow (BM), within erythroblastic islands, which are composed of erythroblasts physically attached to central macrophage cells (Chasis & Mohandas, 2008). The most immature erythroid compartment is the erythroid burst-forming unit (BFU-E), named for its ability to form colonies in semisolid media. BFU-E cells give rise to erythroid colony-forming units (CFU-E). CFU-E progenitors require only 2 and 7 days in mouse and human systems, respectively, to form smaller mature colonies of hemoglobin (Hb) containing cells. Proliferation, maturation, and differentiation of erythroid progenitors are due to the stimulation by stem cell factor (SCF), interleukin 3 (IL-3), granulocyte- macrophage colony-stimulating factor (GM-CSF), interleukin 9 (IL-9), anderythropoietin (EPO) (M. J. Koury & Bondurant, 1990, 1992). EPO is mainly produced in the kidney in hypoxia conditions, and acts through its specific receptor, EPOr, characterized by a single transmembrane domain that dimerizes upon ligand activation. Downstream signals occur through JAK2/STAT5 complex that translocates to the nucleus and induces expression of target genes including iron-responsive element binding protein 2 (IREB2) and anti-apoptotic factors (Camaschella & Nai, 2016). The second erythroid compartment consists of morphologically identifiable, nucleated precursors that progress from proerythroblast (ProE) to basophilic (BasoE), polychromatophilic (PolyE), and orthochromatic (OrthoE) forms (Figure 1.1). Erythroid precursor maturation is characterized by (1) progressive erythroblast expansion, (2) accumulation of Hb, (3) decrease in cell size, (4) nuclear pyknosis, and (5) decrease in RNA content. The morphologic criteria distinguishing these cells rely primarily on the progressive nuclear

condensation combined with changes in cytoplasmic staining, which reflect the degree of hemoglobinization and RNA content. The result of precursor maturation is enucleation, which results in the formation of two cell types: reticulocytes and pyrenocytes. The first contain most of the cytoplasm and Hb, as well as the proteins needed to form a unique cytoskeletal network (Geiduschek & Singer, 1979; S. T. Koury, Koury, & Bondurant, 1989; Lee et al., 2004). The latter, namely the "extruded nuclei", contain the condensed nucleus surrounded by a lipid bilayer and thin rim of cytoplasm (McGrath, Bushnell, & Palis, 2008). Reticulocyte maturation is a complex process that results in approximately 20% loss of plasma membrane surface area, reduced cell volume, increased association of the cytoskeleton to the outer plasma membrane, and the loss of all residual cytoplasmic organelles, including mitochondria and ribosomes (Johnstone, 1992; Merryweather-Clarke et al., 2011). Steady-state levels of RBCs are maintained by the continuous production and release of reticulocytes into the bloodstream to balance the removal of senescent RBCs by macrophages of the spleen (Bennett & Kay, 1981). In adult humans, this mechanism results in the release of more than 2 million reticulocytes every second into the bloodstream. During the dynamic multistep process of erythropoiesis, cell amplification and differentiation are inversely coordinated. Under conditions of hypoxia, to meet the demands of increased RBC generation and oxygenation, EPO production is stimulated by Hypoxia Inducible factor alpha 2 (HIF2α). This phenomenon, known as stress erythropoiesis, is characterized by an imbalance of erythroid proliferation and differentiation axis, resulting in an expansion of the erythroid progenitor pool. Under these conditions, erythropoiesis occurs in extramedullary sites such as the liver and the spleen; iron absorption is increased by the erythroid factor erythroferrone (ERFE) produced by erythroid progenitors, which acts on the liver to suppress hepcidin (*HAMP*) expression, thereby increasing iron availability (Camaschella & Nai, 2016; Kautz et al., 2015). There is a wide spectrum of intrinsic disorders in which stress erythropoiesis is defined as

ineffective erythropoiesis (IE), due to its chronic state and the expansion of abnormal erythroblasts which are unable to generate RBCs. If these conditions are severe, they could result in anemia (Iolascon, Esposito, & Russo, 2012; Iolascon et al., 2011; Nandakumar, Ulirsch, & Sankaran, 2016). (1) Thalassemias are a hereditary group of disorders characterized by insufficient or absent production of α-or β-globin chains, which are the two major components of Hb (Higgs, Engel, & Stamatoyannopoulos, 2012). Different mechanisms contribute to the pathogenesis of IE in α-or β-thalassemia (Oikonomidou & Rivella, 2018): increased apoptosis of thalassemic erythroid precursors (Mathias et al., 2000; Yuan et al., 1993), decreased differentiation of erythroid progenitors (Forster, Cornwall, Finlayson, & Ghassemifar, 2016; Libani et al., 2008), and increased production of reactive oxygen species (ROS) in RBCs due to the aggregation of free α-globins within the cells (Amer, Goldfarb, & Fibach, 2003; Ribeil et al., 2013; Voskou, Aslan, Fanis, Phylactides, & Kleanthous, 2015). (2) Diamond–Blackfan anemia (DBA) is a rare congenital hypoplastic anemia, due to mutations in the genes encoding ribosomal proteins (RP), characterized by a block in erythropoiesis at the progenitor stage, although the exact stage at which this occurs remains to be fully defined (Da Costa, Narla, & Mohandas, 2018; Nathan, Clarke, Hillman, Alter, & Housman, 1978). (3) Finally, CDAs are hereditary diseases that embrace a highly heterogeneous set of rare or very rare anemias that result from various kinds of abnormalities during late stages of erythropoiesis.

Biomarkers of dyserythropoiesis

Since EPO is mainly produced in hypoxia conditions, high serum EPO levels have been observed in patients with a variety of hematological disorders such as acute leukemia, multiple myeloma, myelodysplasia syndrome, aplastic anemia, and pure red cell aplasia. Thus, different cytokines and growth factors, that interconnect the process of erythropoiesis to the maintenance of iron

balance should be evaluated to determine the varying degrees of dyserythropoiesis (Figure 1.2). The iron gateway to cells is the transferrin receptor (TfR), which is truncated from the cell surface and released into the serum as soluble transferrin receptor (sTfR) (Weiss & Goodnough, 2005). sTfR concentrations are directly proportional to the number of transferrin receptors on erythroblasts. Indeed, different studies demonstrated the increase of sTfR levels in patients with non-transfusion-dependent thalassemia, and in splenectomized and non-splenectomized patients with thalassemia intermedia (Danise et al., 2009; Ricchi et al., 2016). Evaluation of hepcidin, the master regulator of iron balance, has a key role in determining iron status due to ineffective erythropoiesis. Hepcidin is a 25-amino acid peptide mainly produced by hepatocytes and secreted into the plasma. This peptide lowers the amount of iron in the serum by inhibiting iron export by ferroportin, a membrane-bound cellular iron exporter present on macrophages and at the basolateral site of enterocytes, which release iron into the circulation. Hepcidin is suppressed by increased erythropoietic iron demand and is upregulated in the presence of increased iron levels or elevated body iron stores (Delaby, Pilard, Goncalves, Beaumont, & Canonne-Hergaux, 2005; Nemeth et al., 2004). Growth Differentiation Factor 15 (GDF15), a member of the TGFβ super family, has been proposed as responsible for the pathological suppression of hepcidin in thalassemia, CDAI, and II, even if alone GDF15 seems unnecessary for physiological hepcidin suppression (Casanovas et al., 2013; Tamary et al., 2008; Tanno et al., 2007). More recently, it was demonstrated that the pathogenesis of iron overload in iron-loading anemias, such as CDAII and beta-thalassemia, is related to the over-expression of the erythroblast-derived hormone ERFE, leading to hepcidin suppression (Kautz et al., 2015; Russo et al., 2016). Moreover, it was demonstrated that a low-frequency missense variant in *ERFE*, *ERFE*-A260S, has a modifier role in the iron overload of CDAII patients by impairing the iron regulation pathways at the hepatic level (Andolfo et al., 2019). Furthermore, Growth Differentiation

Factor 11 (GDF11) has been found over-expressed in beta-thalassemia, myelodysplastic syndrome, and CDAII (De Rosa et al., 2020; Dussiot et al., 2014). Indeed, RAP-011, a ligand trap for GDF11, was recently proposed as a possible therapeutic agent for the management of iron overload in patients with CDAII (De Rosa et al., 2020) (Figure 1.3).

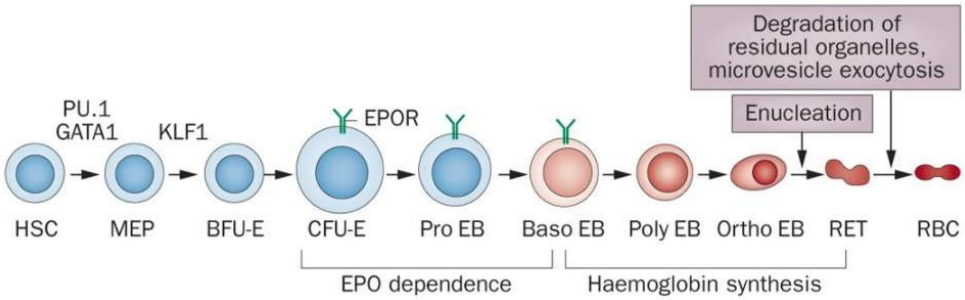


Figure 1.1. Stages of erythroid maturation. Progressive stages of erythroid differentiation showing the relative sizes and presumed or known morphologic appearances of hematopoietic cells at various stages. The transcription factors PU.1 and GATA1 are important in determining whether HSCs will progress towards an erythroid or a non-erythroid fate, whereas KLF1 is important in determining whether MEPs will progress towards an erythroid or a megakaryocytic fate. PU.1 expression continues until the EPO-dependent stages, whereas GATA1 and KLF1 have important roles in differentiation throughout hemoglobin synthesis. Stages of hemoglobin synthesis show relative accumulations of hemoglobin as increasing intensity of red in the cytoplasm. The periods of EPO dependence ending at the early Baso EB stage and hemoglobin synthesis beginning in the late Baso EB stage do not overlap [Adapted from Koury & Haase, Nature Reviews Nephrology, 2015].

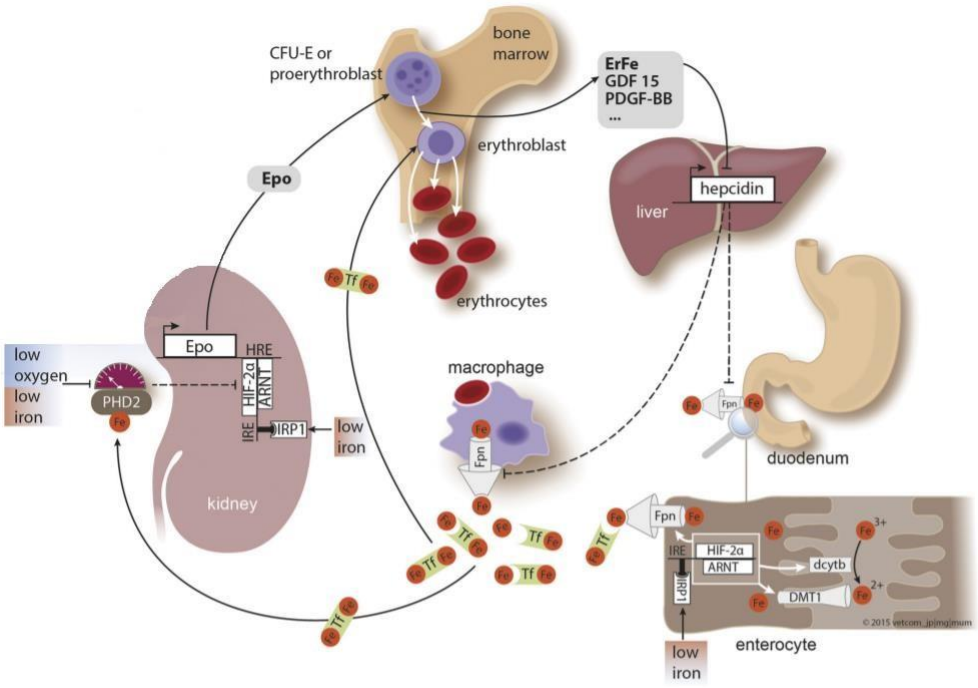


Figure 1.2. Schematic representation of the interconnection between erythropoiesis and maintenance of iron balance. Systemic oxygen levels are monitored by prolyl hydroxylases (PHDs, exemplified by PHD2) in peritubular fibroblasts present in the renal cortex. Iron is a critical cofactor for these dioxygenases. Thus, reduced oxygen or iron levels suppress the activity of PHD2 and cause stabilization of the -subunit of the hypoxia-inducible factor-2 (HIF-2) that heterodimerizes with its partner aryl hydrocarbon receptor nuclear translocator, also termed HIF-2α) to enhance erythropoietin transcription. A second iron-dependent process adjusts EPO levels to iron availability: HIF-2 contains an iron-responsive element (IRE) in its 5= untranslated region. Under iron-deficient conditions, when hemoglobin synthesis is reduced, this RNA structure binds to the iron regulatory protein-1 to inhibit HIF-2 translation. In other words, these mechanisms ensure that EPO synthesis is adjusted to iron availability. Once EPO reaches the bone marrow it promotes red blood cell maturation and proliferation, a process that consumes high amounts of iron. To make sure that sufficient iron is provided systemically, hypoxia-induced soluble factors, such as the EPO-controlled ERFE or GDF15 that are both expressed in erythroid precursor cells, as well as the platelet-derived growth factor BB reach the liver where they reduce expression of hepcidin, the iron hormone that binds the cellular iron exporter ferroportin leading to its internalization and degradation. Thus, suppression of hepcidin allows both elevated iron release from storage organs including macrophages and enhanced absorption of dietary iron by enterocytes. In addition, tissue hypoxia or iron deficiency further augments dietary iron absorption in the intestine. Similar to the situation in the kidney, these conditions stabilize the -subunit of HIF-2 that stimulates transcription of proteins that control iron absorption: the ferrireductase (dcytb), the apical divalent metal transporter-1 (DMT-1), and the iron exporter Fpn. The iron released from macrophages and duodenal enterocytes is transported bound to transferrin (Tf) to ultimately satisfy the iron requirements of erythropoiesis in the bone marrow. CFU-E, colony-forming unit-erythroid; HRE, hypoxia response element [Adapted from Gassmann and Muckenthaler, J Appl Physiol., 2015]

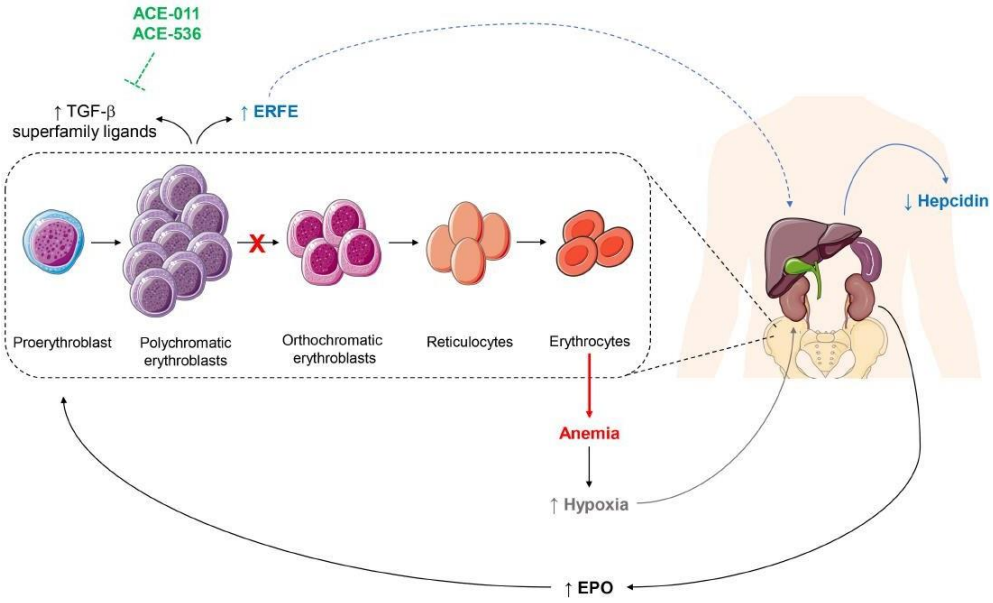


Figure 1.3. Schematic representation of the pathogenic mechanisms of CDAII at the systemic level. The pathogenic mechanism of CDAII at a systemic level, highlighting the role of ERFE in the interplay between the bone-marrow and hepatic compartments. At the bone-marrow level, block of erythroid maturation results in accumulation of erythroblasts that secrete ERFE. The increased levels of ERFE are responsible for the suppression of Hamp expression, which codes for hepcidin, and this can result in liver iron overload. EPO, erythropoietin; TGF-β, transforming growth factor-β [Adapted from Iolascon et Al., Blood, 2020].

1.2 Classification, diagnostic criteria, and clinical management of CDAs

Congenital dyserythropoietic anemias are a large group of hypo-productive anemias that result from various kinds of abnormalities during late stages of erythropoiesis (Russo, Marra, Rosato, Iolascon, & Andolfo, 2020). They are characterized by morphological abnormalities of erythroblasts in the BM and ineffective erythropoiesis as predominant mechanism of anemia, accompanied by a hemolytic component (Iolascon, Heimpel, Wahlin, & Tamary, 2013). These conditions can be suspected in the presence of anemia and hemolytic signs, accompanied with reticulocytosis inadequate to the degree of anemia. In particular, the following criteria are required: (1) evidence of anemia, jaundice, splenomegaly; (2) evidence of ineffective erythropoiesis; (3) occurrence of typical morphological features of erythroblasts at BM examination; and (4) exclusion of other congenital anemias that fulfill criteria 1 and 2, such as thalassemia syndromes and other inherited bone marrow failure syndromes (Figure 1.4) (Gambale, Iolascon, Andolfo, & Russo, 2016). The BM of CDA patients is always hypercellular, due to an exclusive and pronounced increase of erythroblasts, with altered, generally increased, erythropoietic/granulopoietic ratio (E:G) (normal reference values approximately 0.3-0.5). The expansion of the erythropoietic tissue leads to high serum concentration of the soluble transferrin receptor (sTfR), if iron deficiency is excluded (Beguin, 2003). Dyserythropoiesis appears to be a morphological feature common to several conditions, and this could account for the difficulties in the diagnosis of CDAs. However, the specific morphological alterations of the erythroid precursors justify the heterogeneity of these disorders (Figure 1.5). Indeed, the three classical types of CDAs (types I, II, and III) are defined based on BM morphology, even though inclusions of additional CDAs, the so called CDA variants, despite remarkable morphological studies, gradually led to overlapping entities and imposed a limitation on classification. To this end, the identification of the causative genes of the most common forms among CDAs in the last two decades

represented an evident advantage for reclassifying these disorders, as well as in understanding their pathogenesis. Moreover, uncovering the molecular basis of CDAs helped to unravel novel aspects of the molecular biology of erythropoiesis (Gambale et al., 2016). From the genetic point of view, six different types of CDAs are listed in the Online Mendelian Inheritance in Man (OMIM) compendium of human genes and genetic phenotypes (Table 1.1). However, thanks to the development of high-throughput technologies such as next generation sequencing (NGS), the discovery of new causative genes could be a matter of fact. The identification of genetic variations in hereditary disorders, beyond obtaining definitive diagnosis and planning patient management, is crucial in estimating their prevalence and geographical distribution. The majority of CDAs are due to the small number of offspring in most European families, where single cases in one family are the rule rather than the exception. Together with the rarity of the disorder and the need to perform a bone marrow biopsy for diagnosis, this explains why correct diagnosis is often delayed, particularly in mild cases, even when anemia and/or hyperbilirubinemia have been evident for many years (Heimpel and Iolascon, 2009). The main complications of CDAs are associated with chronic hemolytic anemia: iron overload, hydrops fetalis, aplastic crisis, hyperbilirubinemia, gallstones, and splenomegaly. Iron overload is mainly due to ineffective erythropoiesis, and it is also linked to both the transfusion regimen and the hemolytic component. Among CDAII patients, approximately 30% of those who are not transfusion dependent show increased ferritinemia (ferritin >300 ng/mL), while 17% of them show marked hemosiderosis (ferritin >600 ng/mL). CDA patients with severe iron overload also show liver damage (i.e., cirrhosis), heart failure, diabetes mellitus, and hypergonadotropic hypogonadism (with related subfertility and osteopenia). Moreover, the severe phenotypes can be related to co-inheritance of modifier mutations, such as polymorphic variants in the *HFE* gene, which causes hemochromatosis type 2 (Gambale et al., 2016; Iolascon, Andolfo, & Russo, 2020; Liu et al., 2012).

Clinical management of CDAs depends on severity and characteristics of the patients. The standard treatment for cases with severe anemia (hemoglobin, <7 g/dL) is transfusion. In cases of suspected CDA *in utero*, it is important to search for hydrops using prenatal ultrasound scanning, and in cases of fetal anemia, *in-utero* transfusions are needed (Iolascon et al., 2020). Iron chelators are the best treatment for iron overload and its deposition in target organs. The guidelines for chelation in CDAs are generally the same assessed for thalassemia (Angelucci et al., 2008). Periodical evaluation of the iron balance parameters in the serum, such as ferritin, transferrin saturation, and hepcidin, as well as T2*-weighted magnetic resonance imaging for liver and heart iron accumulation are crucial to monitor patients. Splenectomy is one possible therapeutic approach to the management of severely anemic patients, even though it is restricted to those with symptomatic splenomegaly (Iolascon et al., 2017). Recent studies are evaluating the efficacy of drugs that target ineffective erythropoiesis. As previously discussed, activin receptor ligand traps (e.g., luspatercept, sotatercept) have shown promising results in Phase I and II clinical trials for patients with ß-thalassemia (Cappellini et al., 2019). Data on the use of a murine analog of sotatercept, RAP-011, in a CDAII cellular model, propose its use in the management and restoration of erythroid maturation in CDAII patients (De Rosa et al., 2020). Advanced therapies involve hematopoietic stem cell transplantation, which has been successfully performed in CDAI and CDAII patients (Ayas et al., 2002; Miano et al., 2019; Nair, Das, & Sharma, 2009), even though iron overload is an important issue and has to be treated before transplantation (Unal et al., 2014). Finally, great efforts are spent on gene therapy approaches, even though it is still far from clinical practice. For instance, it was demonstrated that *in-vitro* gene therapy with p60-BBF2H7 lentivirus led to up-regulation of *SEC23A* gene, with a subsequent normalization of its paralogue, *SEC23B*, causative gene of CDAII (Pellegrin et al., 2019).

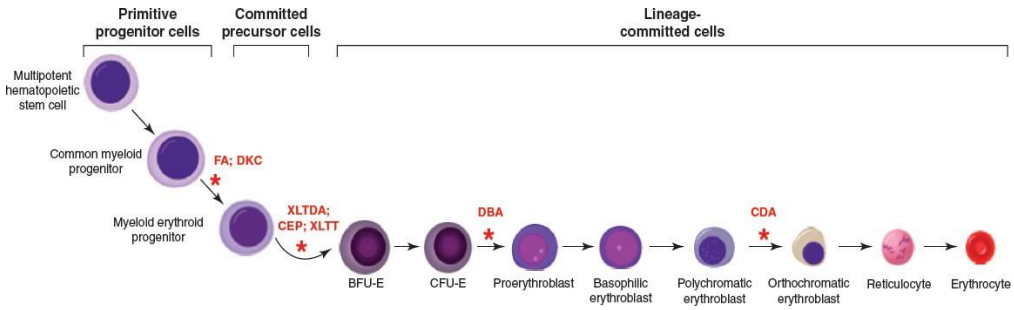


Figure 1.4. Differential diagnosis of congenital dyserythropoietic anemias (CDAs) among marrow failure syndromes. Erythroid maturation arrest can occur at several stages and can result in different marrow failure syndromes. CDA marrow is characterized by erythroid hyperplasia because the maturation arrest occurs later in these patients. [Adapted from Williams Manual of Hematology, Ninth Edition.]

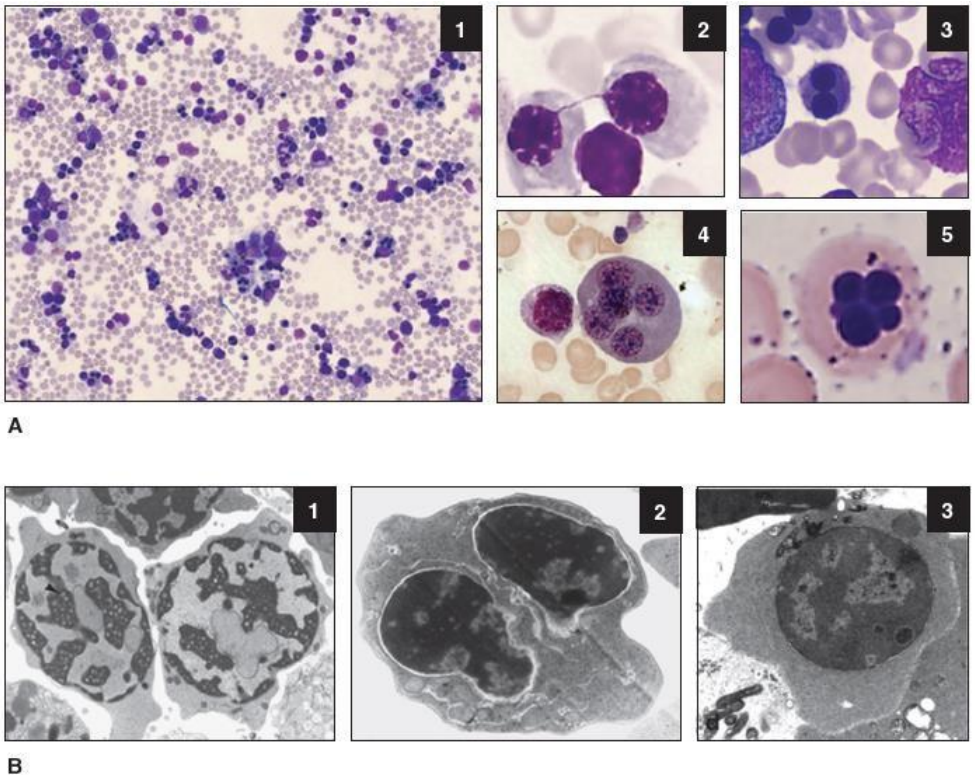


Figure 1.5. Marrow morphologic features of congenital dyserythropoietic anemia (CDA) erythroblasts. (a) Marrow light microscopy of CDA patients almost always shows erythroid hyperplasia (*inset 1*). However, specific morphologic abnormalities in erythroid precursor cells are used as hallmarks for discerning different types of CDAs. The presence of internuclear chromatin bridging is a mark of CDA type I (*inset 2*), whereas the presence of bi – or multinucleated late erythroid precursors is the discriminating feature of CDA type II (*inset 3*). Giant multinucleated erythroblasts (*inset 4*) and multinucleate erythroblasts (*inset 5*) are typical features of CDA type III and IV, respectively. **(b)** Marrow electron microscopy of CDAI erythroblasts shows a typical Swiss cheese- heterochromatin pattern in the nucleus (*inset 1*), whereas CDA type II erythroid precursors display the characteristic double plasma membrane (*inset 2*). CDA type IV has nonspecific features, shared among different CDA types, such as marked heterochromatin, invagination of nuclear membrane, intranuclear precipitated material, and nuclear blebbing (*inset 3*). [Adapted from Williams Manual of Hematology, Ninth Edition]

Table 1.1. Classification of CDAs by OMIM database

Disease symbol	Phenotype	Phenotype MIM number	*Gene* Location	Inheritance	N cases[§]
CDA Ia	Congenital dyserythropoietic anemia type Ia	224120	*CDAN1* 15q15.2	AR	< 100
CDA Ib	Congenital dyserythropoietic anemia type Ib	615631	*C15orf41* 15q14	AR	< 10
CDA II	Congenital dyserythropoietic anemia type II	224100	*SEC23B* 20p11.23	AR	> 400
CDA III	Congenital dyserythropoietic anemia type III	105600	*KIF23* 15q21	AD	< 20
CDA IV	Congenital dyserythropoietic anemia type IV	613673	*KLF1* 19p13.2	AD	≈ 10
XLTDA	Thrombocytopenia X-linked with or without dyserythropoietic anemia	300367	*GATA1* Xp11.23	XLR	≈ 10

[§]Number of cases with positive molecular analysis
AD, Autosomal dominant; AR, Autosomal recessive; XLR, X-linked recessive

1.3 From classical CDAs to non-classical forms

CDA type I

Most of the patients affected by CDAI exhibit lifelong macrocytic anemia with variable values of Hb and mean cell volume (MCV) ranging between 100 and 120 fL; however, it can be normocytic in childhood. While some cases have been identified *in utero* (Kato et al., 2001; Lin et al., 2014; Parez et al., 2000), leading to fetal demise if untreated, most cases are detected in childhood or young adulthood (Shalev, Al-Athamen, Levi, Levitas, & Tamary, 2017). Relative reticulocytopenia is typically observed, accompanied to increased Hb turnover as attested by indirect hyperbilirubinemia, high lactate dehydrogenase (LDH) value and low or absent plasma haptoglobin. All patients develop splenomegaly in adolescence or adulthood, while about 20% of cases show congenital anomalies, particularly syndactyly in hands or feet, absence of nails or supernumerary toes, pigeon chest deformity, and short stature. BM smear shows hypercellularity and erythroid hyperplasia. Approximately 30-60% of polychromatic erythroblasts show abnormalities of nuclear and chromatin structure. Indeed, the morphological pathognomonic feature of CDAI is the presence of thin chromatin bridges between the nuclei pairs of erythroblasts (Figure 1.5 A, panel 2). A minority of erythroblasts shows bi- or multinuclearity, but in contrast to CDAII, the nuclei of binucleated cells are of different size and shape. At electron microscopy (EM), heterochromatin is denser than normal, and forms demarcated clumps with small translucent vacuoles, giving rise to the metaphor of "Swiss cheese appearance" (Heimpel et al., 2010) (Figure 1.5 B, panel 1). Besides the therapeutic approaches shared with the other types of CDAs, treatment with interferon alpha (IFNα) is effective and specific for CDAI patients. The efficacy of IFNα was discovered by chance in a patient who had contracted Hepatitis C through contaminated blood, which became transfusion independent with reduction in ineffective erythropoiesis and amelioration of scanning electron microscopy features on bone marrow aspirate (Lavabre-

Bertrand et al., 1995). Other studies demonstrated that prolonged treatment with IFNα leads to a stable Hb, with ongoing normalization in electron microscopy features, reduction of intramedullary hemolysis, and improvement of ineffective erythropoiesis (Heimpel et al., 2006). Wickramasinghe and colleagues observed that IFN production by Epstein-Barr virus-transformed lymphocytes from CDAI patients is reduced (Wickramasinghe, 1997). Nevertheless, further studies are needed to understand if the response to IFNα in CDAI patients is due to subnormal production *in vivo*, or the molecular defect leading to CDAI can be overcome by the over-expression of IFN-responsive genes (N. B. A. Roy & Babbs, 2019). From a genetic standpoint, CDAI is inherited as autosomal recessive disorder, due to biallelic mutations in either *CDAN1* (CDAIa) or *CDIN1* (or *C15orf41*) (CDAIb) genes (Babbs et al., 2013; Dgany et al., 2002) (Figure 1.6). Of note, approximately 10% of CDAI cases remain unexplained by mutations in either of the known genes, suggesting that there may be cis-acting regulatory mutations affecting these genes or further loci causative of CDAI. CDAIa and CDAIb patients show similar clinical features. Anyhow, a difference in Hb level and MCV value can discriminate between them. Particularly, CDAIb patients exhibit a more severe anemia but normal MCV when compared to CDAIa patients. Nevertheless, more patients should be identified to validate this observation (Gambale et al., 2016).

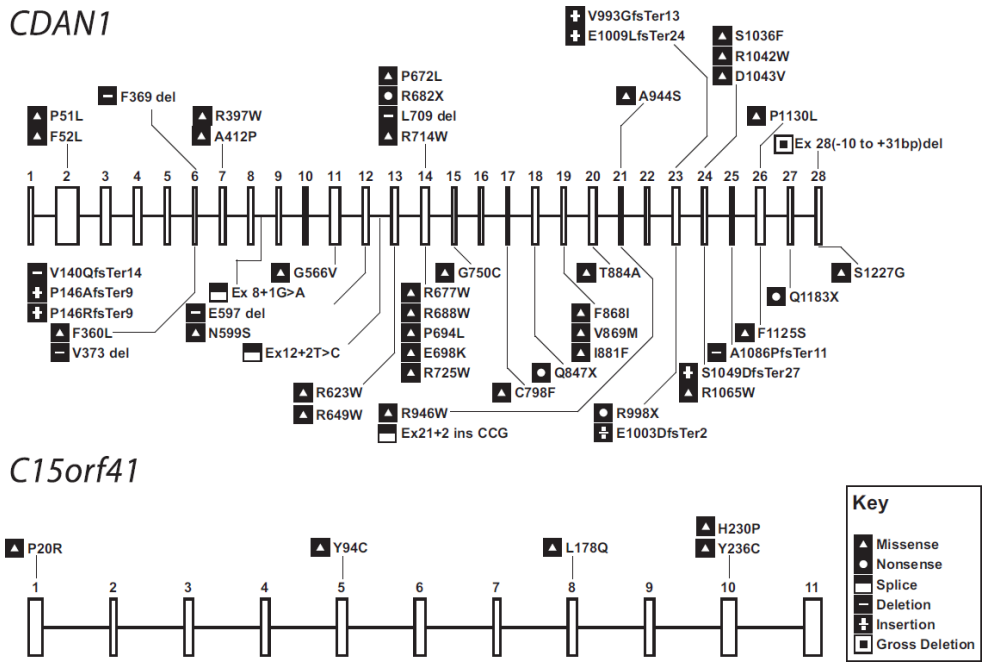


Figure 1.6. Representation of all known pathogenic variants associated with *CDAN1* and *C15orf41*. The exons in each gene are numbered to show the location of each mutation. Coding mutations are shown according to the amino acid change while splicing changes are shown according to their location relative to exons. [Adapted from Roy and Babbs, British Journal of Haematology, 2019]

CDAN1

The first causative gene (CDAIa) in which pathogenic variants were identified has been *CDAN1* (chr15q15.2) (Dgany et al., 2002). More than 100 patients and 51 causative mutations have been described so far (Figure 1.6). *CDAN1* encodes a ubiquitously expressed protein, codanin-1, which is localized to heterochromatin in interphase cells. It is a cell cycle-regulated protein active in the S phase; indeed, at mitosis, codanin-1 underwent phosphorylation, which coincided with its exclusion from condensed chromosomes. The proximal *CDAN1* gene promoter region, containing five putative E2F binding sites, was found to be a direct target of E2F1 (Noy-Lotan et al., 2009). Moreover, it has been demonstrated that codanin-1 is part of the cytosolic Asf1-H3-H4-importin-4 complex, which is implicated in nucleosome assembly and disassembly. It binds directly to Asf1 via a conserved B-domain, implying a mutually exclusive interaction with the chromatin assembly factor 1 (CAF-1) and HIRA. Codanin-1 depletion accelerates the rate of DNA replication and increases the level of chromatin-bound Asf1, suggesting that Codanin-1 guards a limiting step in chromatin replication. Indeed, this function is compromised by two CDAI mutations that impair complex formation with Asf1, providing insight into the molecular basis for CDAI disease (Ask et al., 2012). Very recently, studies on human umbilical cord blood-derived erythroid progenitors (HUDEP-2), mutated for two common heterozygous mutations in CDAN1 patients, demonstrated alterations in histone availability and acetylation, disruption of the cell cycle, and decreased cell viability during terminal erythroid maturation (Murphy et al., 2020). Furthermore, *CDAN1* knockout mice die *in utero* before the onset of erythropoiesis, suggesting a critical role of codanin-1 in developmental processes beyond erythropoiesis (Renella et al., 2011). Of note, no homozygous patients for null mutations have been described so far. Homozygous or compound heterozygous in *CDAN1* gene cover approximately 50% of CDAI patients, while in 30% of cases only a single mutant allele can be identified.

C15orf41 (CDIN1)

The second gene, identified as causative of CDAI, is *C15orf41* (chr15q14) (Babbs et al., 2013), recently named Codanin-1 Interacting Nuclease 1 (*CDIN1*). In cultured erythroblasts, *C15orf41* produces a spliced transcript encoding a protein with homology to the Holliday junction resolvases.

C15orf41 is an uncharacterized gene comprising 11 exons (Figure 1.7). Data from gene expression arrays show that *C15orf41* is widely transcribed, although its expression appears to be elevated in B lymphoblasts, CD34+ cells, cardiomyocytes and fetal liver suggesting a specific requirement in hematopoiesis (Su et al., 2004). Global gene expression analysis throughout erythropoiesis reveals that *C15orf41* is uniformly expressed during erythroid differentiation (Merryweather-Clarke et al., 2011), suggesting a constant requirement for this protein. Corroboration by profile-to-profile comparison methods, provide strong evidence that the C15orf41 protein contains 2 N-terminal AraC/XylS-like wHtH domains followed by a PD-(D/E) XK nuclease domain, suggesting *C15orf41* encodes a divalent metal-ion dependent restriction endonuclease. Biological functions performed by this family include DNA damage repair, Holliday junction resolution, and RNA processing. In some members of the PD-(D/E) XK nuclease superfamily this combination of domains underlies protein-protein interactions (usually dimerization) and may establish additional DNA interactions, thereby improving DNA specificity. Recently, structural studies evidenced that C15orf41 protein is thermostable, unstable, and acidic. Moreover, the protein is predicted to contain the TPD (Treponema Pallidum) domain (135 to 265 residue position), with an unknown function, and three post-translational modification (PTM) sites: K50 (Acetylation), T114 (Phosphorylation) and K176 (Ubiquitination) (Ahmed, Shahjaman, Kabir, & Kamruzzaman, 2018). Phylogenetic analysis demonstrated that the same animal taxa loose both *CDAN1* and *C15orf41* genes, strengthening the hypothesis that the two proteins participate in the

same pathway. Indeed, the demonstration of the interaction between the C-terminal region of Codanin-1 and C15orf41 enlightened their relationship. Moreover, it was observed that Codanin-1 overexpression elevates C15orf41 protein levels and shifts it from the nucleus to the cytoplasm (Shroff, Knebel, Toth, & Rouse, 2020; Swickley et al., 2020).

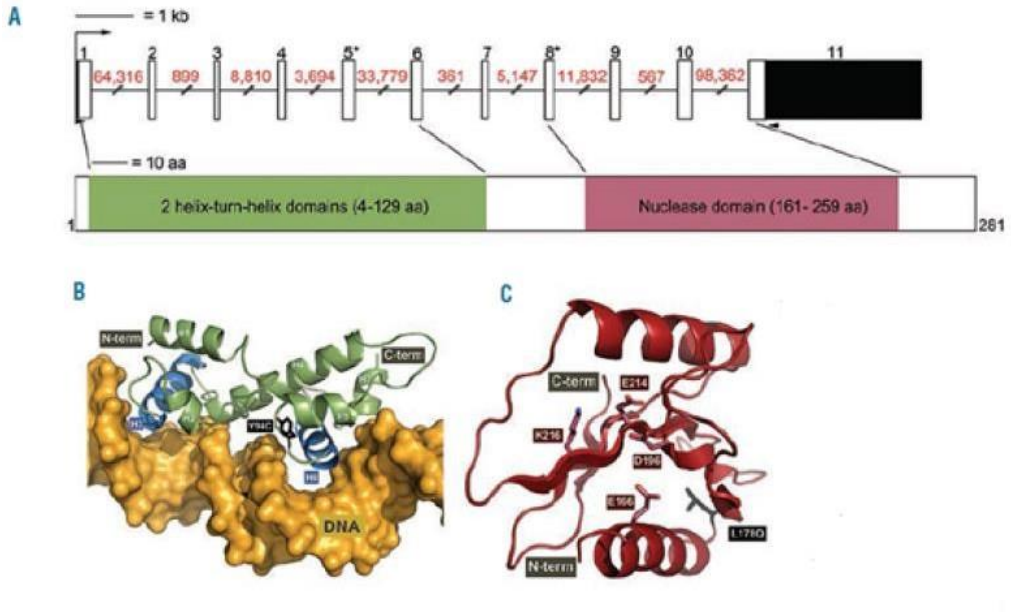


Figure 1.7. *CDIN1* gene and protein structure. (**A**) Schematic representation of the *CDIN1* gene, exons are shown to scale with coding sequence shown in white and untranslated regions (UTRs) in black. Red numbers above lines indicate intron sizes (not to scale); numbers above exons indicate exon number; asterisks indicate the two exons in which the first mutations have been identified. The lower section shows the CDIN1 protein with annotated domains shown to scale. (**B**) Two helix-turn-helix domains predicted for the N-terminal of CDIN1 (amino acids 4-129). Helices are numbered and putative DNA interaction helices are shown in blue (H3 and H6). (**C**) The PD-(D/E) XK nuclease domain predicted for the C-terminal region of CDIN1 (amino acids 161-259) [Adapted from Babbs et al., 2013].

CDA type II

CDAII is the most common form of the CDAs. It is characterized by the presence of bi- and multinucleated erythroblasts in BM (more than 10%), with nuclei of equal size and DNA content, suggesting a cytokinesis disturbance (Figure 1.5 A, panel 3). Development of other hematopoietic lineages is normal. The identification of glycosylation abnormalities with fast-moving band 3 (the anion exchanger 1) and band 4.5 (glucose transporter 1), by the analysis of RBC membrane proteins with sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE) electrophoresis, is a highly sensitive and specific diagnostic tool. CDAII patients show moderate to severe normocytic or microcytic anemia, with normal or insufficiently increased reticulocyte count. Individuals with CDAII show progressive splenomegaly, gallstones, and iron overload potentially with liver cirrhosis or cardiac failure (Iolascon et al., 2012). CDAII is an autosomal recessive disorder, due to bi-allelic mutations in the *SEC23B* gene (20p11.23) (Table 1.1). It encodes the cytoplasmic COPII (coat protein) component SEC23B, which is involved in the secretory pathway of eukaryotic cells. The entire complex mediates accumulation of secretory cargo, deformation of the membrane, and anterograde transport of correctly folded cargo for budding from the endoplasmic reticulum toward the Golgi apparatus (Bianchi et al., 2009; Schwarz et al., 2009). Homozygosity or compound heterozygosity for null mutations was never found as they are presumably lethal. However, the presence of two hypomorphic alleles accounting for mild CDAII clinical forms was described (Russo et al., 2013).

CDA type III and CDA-variants

CDAIII (OMIM #105600) is the rarest among the three types, with autosomal dominant inheritance pattern (Table 1.1). The causative gene of this condition, *KIF23* (15q21), was identified in two unrelated families, carrying the same variant, c.2747C>G (p.P916R). *KIF23* encodes a kinesin-superfamily molecule, mitotic kinesin-like protein 1 (MKLP1), essential for cytokinesis (Figure 1.5 A, panel 4) (Lind et al., 1995; Makyio et al., 2012).

The subgroup of CDA variants includes CDA type IV (CDAIV) (OMIM #613673) and X-linked thrombocytopenia with or without dyserythropoietic anemia (XLTDA) (OMIM #300367). Hypercellular bone marrow and binucleate or multinucleate erythroblasts are observed (Figure 1.5 A, panel 4). Electron microscopy shows immature erythroid progenitors with atypical cytoplasmic inclusions, invagination of the nuclear membrane, and marked heterochromatin. CDAIV is an autosomal dominant condition caused by a unique heterozygous variant c.973G>A (p.E325K) in the *KLF1* gene (19p13.2) (Table 1.1). This gene encodes the homonymous erythroid-specific transcription factor KLF1, which is required in terminal erythroid differentiation, with a critical role in the regulation of the switch between fetal and adult hemoglobin expression (Siatecka & Bieker, 2011). The molecular mechanism of the *KLF1*-E325K variant has been investigated. Its main effect is either impaired recognition by this mutated *KLF1* of its normal cognate site, or incorrect protein complex formation that interferes with the wildtype activity at target sites (Xue, Ankala, Wilcox, & Hegde, 2015). XLTDA is characterized by macro-thrombocytopenia with hypo granulated platelets, bleeding tendency, and mild-to-severe anemia. Bone-marrow features are dyserythropoiesis and impaired megakaryopoiesis. XLTDA is caused by mutations in *GATA1* (Xp11.23) (Table 1.1), an X-linked gene that encodes a DNA-binding protein with two zinc fingers and a transactivation domain. GATA1 has an essential role in development and maintenance of both erythroid and megakaryocytic lineages (Crispino & Horwitz, 2017). Due to the X-linked inheritance, males

are predominantly affected, and the severity and specificity of the phenotype depends on the imbalance in GATA1 function, which is determined by the mutation type and its expression level (Shimizu & Yamamoto, 2016). Due to their role as transcriptional regulators, several mutations identified in KLF1 or GATA1 result in heterogeneous phenotypes. Moreover, genetic variants in KLF1 and GATA1 might also have roles as modifiers of the phenotype. Indeed, co-inheritance of pathogenic variants in KLF1/GATA1 and in other relevant erythrocyte genes has been reported in some cases that show a severe clinical course (Pereira et al., 2016).

Syndromic CDAs

Among the syndromic conditions, Majeed syndrome (OMIM #609628) is a rare autosomal recessive disorder that is hallmarked by chronic recurrent multifocal osteomyelitis, inflammatory dermatosis, and hypochromic microcytic anemia with dyserythropoiesis. The causative gene is *LPIN2* (18p11.31), which encodes a phosphatidate phosphatase that is important in lipid metabolism (Iolascon et al., 2020; Rao, Gopalakrishna, Bing, & Ferguson, 2016). More recently, a syndromic condition, known as early infantile epileptic encephalopathy-50 (OMIM #616457), has been associated to mild CDA-like anemia with marked anisopoikilocytosis and abnormal glycosylation of the erythrocyte proteins band-3 and RhAG (Koch et al., 2017; Ng et al., 2015). This syndrome is caused by bi-allelic mutations in the *CAD* gene (2p23.3), which encodes a trifunctional enzyme that catalyzes the first steps of *de-novo* pyrimidine biosynthesis. A unique study described two consanguineous families with homozygous mutation in the *COX4I2* gene (20q11.21), which encodes a subunit of cytochrome c oxidase, the terminal enzyme in the respiratory chain. In these patients, dyserythropoiesis is associated with exocrine pancreatic insufficiency. Finally, morphological erythroblast abnormalities like CDAII have also been reported in a case of mevalonate kinase

deficiency, caused by compound heterozygosity for two missense mutations in the *MVK* gene (12q24.11) (Samkari et al., 2010).

1.4 Hereditary hemolytic anemia due to Pyruvate Kinase Deficiency

Pyruvate kinase deficiency (PKD) is the most common enzyme-related glycolytic defect that results in red cell hemolysis, due to bi-allelic mutations in *PKLR* gene (1q21) (Secrest et al., 2020). This disorder is characterized by clinical heterogeneity, which results in a variable degree of hemolysis, causing irreversible cellular disruption (Radosinska & Vrbjar, 2016). Pyruvate kinase enzyme (PK) has a key role in glycolysis, converting phosphoenolpyruvate to pyruvate, upon which RBC metabolism completely hinges on. Lack of PK enzyme, with its levels of <25%, affects RBC ATP production, causing decreased RBC deformability, RBC damage, and the consequent trapping of defective RBCs by splenic and hepatic capillaries, causing hepatosplenomegaly.

Human PK comprises four isozymes (L, R, M1, M2). Among them, R-PK is the only found in normal mature red cells (Secrest et al., 2020). Hematopoietic stem cells and progenitor cells express M2-PK, which switches to R-PK in cells of erythroid lineage, where it is constantly synthesized during erythroid cell maturation (Takegawa, Fujii, & Miwa, 1983), suggesting that metabolic abnormalities in R-PK deficiency could alter the differentiation of erythroid progenitors into mature erythrocytes. Indeed, it was reported that hematopoiesis was enhanced in the spleen of PK-deficient patients (Aizawa et al., 2003). Intriguingly, in PK-1^{slc} deficient mice, the number of BFU-E in the BM was significantly increased, indicating enhanced erythropoiesis. Moreover, apoptotic cells of erythroid lineage were identified in the splenic red pulp of PK deficient mice, suggesting that the metabolic disturbance in PKD alters not only the survival of red cells but also the maturation of erythroid progenitors, resulting in ineffective erythropoiesis (Aizawa et al., 2005).

Concerning therapy and clinical management, folic acid supplementation and blood transfusion ameliorates anemia, and surgical management of massive splenomegaly may be required. Recently, it was demonstrated that AG-348, namely Mitapivat, an allosteric activator of both wild type and mutant PK enzymes, can increase enzymatic activity in patient erythrocytes (Kung, et al 2017). Indeed, a phase 2 study on PKD patients treated with Mitapivat has been recently completed (ClinicalTrials.gov: NCT02476916) (Grace et al., 2019; Yang et al., 2019), and a phase III trial is ongoing. These findings pointed out how the proper identification and diagnosis of rare conditions, as PKD, may be valuable for guiding treatment (Grace et al., 2018).

1.5 Differential diagnosis of CDAs and other hereditary hemolytic anemias

Although the diagnostic workflow for hereditary anemias is a normal clinical practice, differential diagnosis, classification, and patient stratification of HA are often exceedingly difficult. Among all HA subtypes, thalassemias and hemoglobinopathies are the only conditions that do not undergo misdiagnosis, thanks to the presence of microcytic anemia, pathological hemoglobin electrophoresis, and positive family history (Gambale et al., 2016). Concerning the other conditions, a wide range of unspecific and overlapping phenotypes can be observed in patients with different genetic backgrounds. For instance, CDAI/CDAII and Dehydrated Hereditary Stomatocytosis (DHS) share several clinical findings, thus are often misdiagnosed, as well as CDAII and Hereditary spherocytosis (HS) (Andolfo, Russo, Gambale, & Iolascon, 2018). The incorrect diagnosis could critically impact on follow-up and therapy of the patients. Splenectomy, which is the most effective surgical treatment for HS, could be unnecessary for a CDA patient or, even worst, contraindicated in DHS, due to the risk of severe thrombotic events (Iolascon et al., 2017; Stewart et al., 1996). Co-inheritance of multiple conditions or multiple disease-associated variants are further issues that increase the complex scenario of

hereditary anemias. For instance, even though ektacytometry is the gold standard for DHS diagnosis, the co-inheritance of beta-thalassemia trait and splenectomy may modify the ektacytometry curve shape thereby leading to a misdiagnosis (Lazarova, Gulbis, Oirschot, & van Wijk, 2017). In the past years, the conventional workflow for the diagnosis of these conditions started with positive familial history, complete blood count, and peripheral blood smear observation as first line of investigation; then, biochemical specialized tests and, eventually, bone marrow aspirate were required; finally, genetic testing served as a confirmatory test (Figure 1.8). Very often, no mutations in the candidate gene were identified by this approach, leading to a confusing or lacking molecular diagnosis. Nowadays, genetic testing is the first line of investigation for hereditary anemias (N. B. A. Roy & Babbs, 2019), especially when clinical data of the patients are not informative or when the patient is transfusion dependent. In this context, next-generation sequencing (NGS) has revolutionized the framework of diagnosis.

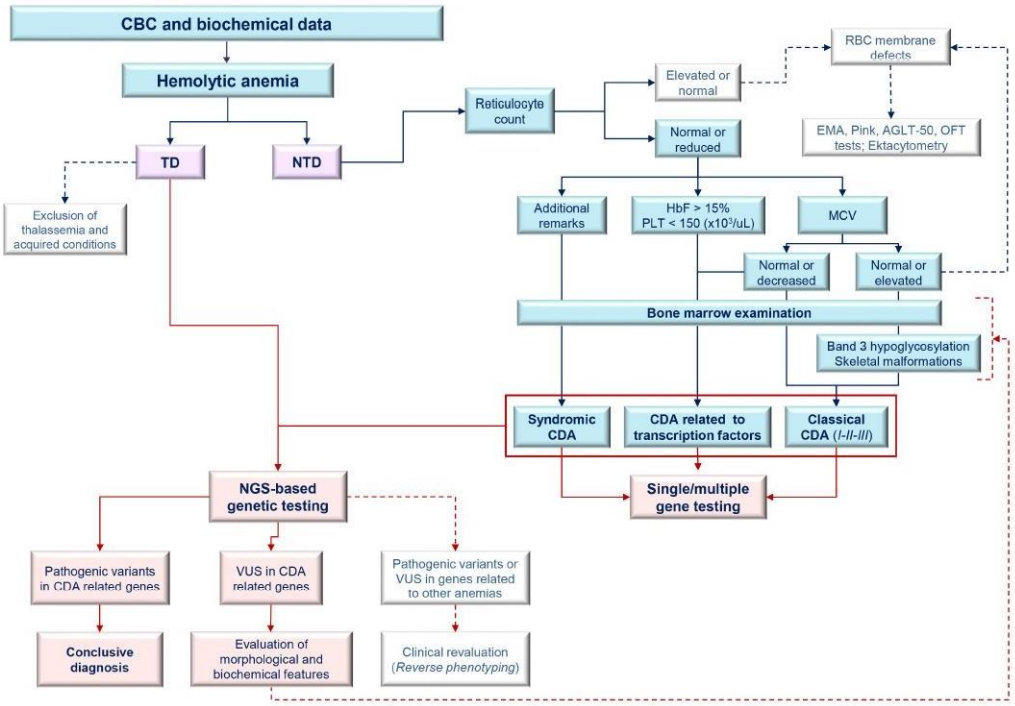


Figure 1.8. Comparison between the conventional pathway and an alternative NGS-based pathway for the diagnosis of inherited hemolytic anemias. In the classical pathway (blue arrows), the investigation is based on family history, and standard blood investigations. Specialized tests are then requested according to the suspected diagnosis, such as bone marrow examination, Band-3 hypoglycosylation, and presence of bone malformations. Genetic testing is then reserved as a confirmatory test. The alternative pathway is shown for transfusion dependent patients (red arrows). NGS is employed as first line investigation, obviating the need for bone marrow biopsies. This approach can cut the time to diagnosis, remove the need for some bone marrow biopsies, provide accurate diagnosis of cases and allow genetic counselling [Adapted from Iolascon et Al., Blood, 2020].

1.6 NGS as a diagnostic tool for hereditary anemias

So far, monogenic approaches based on Sanger sequencing, were used for diagnosis and identification of new causative genes of HAs. Nowadays, even though single gene testing is still suggested for patients with complete phenotyping, second-generation sequencing technologies, commonly referred to as next generation sequencing (NGS), allowed us to move from a monogenic approach to an oligo/multigenic approach (Iolascon et al., 2020; Russo et al., 2020) (Figure 1.9). NGS is a very versatile technology, applicable to various questions either in basic research or in clinical research. Indeed, over the past decade, NGS has led to an exponential increase in our understanding of the genetic basis of Mendelian diseases. This technology of high-throughput sequencing has provided a large impetus for *de novo* sequencing, resequencing, exome sequencing, transcriptome profiling, methylation profiling, and metagenomics studies (Desai & Jere, 2012). Each approach has its own advantages and limitations. In the NGS era, the genetic testing is going to move from few candidate genes to wider panels of genes, named targeted t-NGS, which targets a group of selected genes. T-NGS is faster and cheaper than other NGS technologies, leads to a higher sequencing coverage, and, therefore, highly accurate DNA variant calling for the region of interest. These features make it a reliable solution for the processing of large sample numbers in a diagnostic laboratory. The general principle of t-NGS provides for specific probes designed to target regions of interest relevant to genetically heterogeneous disease phenotypes, for which the differential diagnosis is essential. Gene panels often have a higher sequencing coverage and depth than whole exome sequencing (WES) or whole genome sequencing (WGS), resulting in a greater diagnostic yield (Xue et al., 2015). Moreover, the risk of incidental findings is reduced, being designed to only include a pre-selected subset of genes. On the other hand, custom panels need to be regularly updated to include newly discovered genes and the inherent requirement of a correctly interpreted clinical context (de Haan, Eijgelsheim, Vogt, Knoers, & de Borst,

2019). In the last few years, several custom gene panel for hereditary anemias have been developed, providing a diagnostic yield from 38% to 87%, to overcome the diagnostic limit for CDAs and related disorders (N. B. Roy et al., 2016; Russo et al., 2018; Shefer Averbuch et al., 2018). Of note, this diagnostic approach leads to modification of the original clinical diagnosis in 10% to 40% of the patients investigated (Iolascon et al., 2020). Moreover, gene panels also allowed the identification of polygenic conditions, or the presence of modifier variants associated with causative mutations (Andolfo et al., 2019).

In contrast with gene panels, WES sequences all the protein-coding genes of the genome, allowing a more flexible analysis compared to gene panels and the opportunity to identify new causative genes. Furthermore, WES data can be stored for future reanalysis as new genes are discovered and variants are reclassified (de Haan et al., 2019). Of note, it is possible to only target a specified subset of genes with an *in-silico* panel (targeted WES). By this approach, the original WES data can be accessible for further analysis if new causative genes are discovered or if a causative variant cannot be identified in the first analysis. Finally, WGS is also useful for the identification of noncoding causative mutations, which might account for the disruption of transcriptional factor occupancy sites and *cis*-regulatory elements. Indeed, the mutational disruption of transcription factor occupancy sites has been shown to be a pathogenic mechanism in several hematological disorders (Kaneko et al., 2014; Manco et al., 2000).

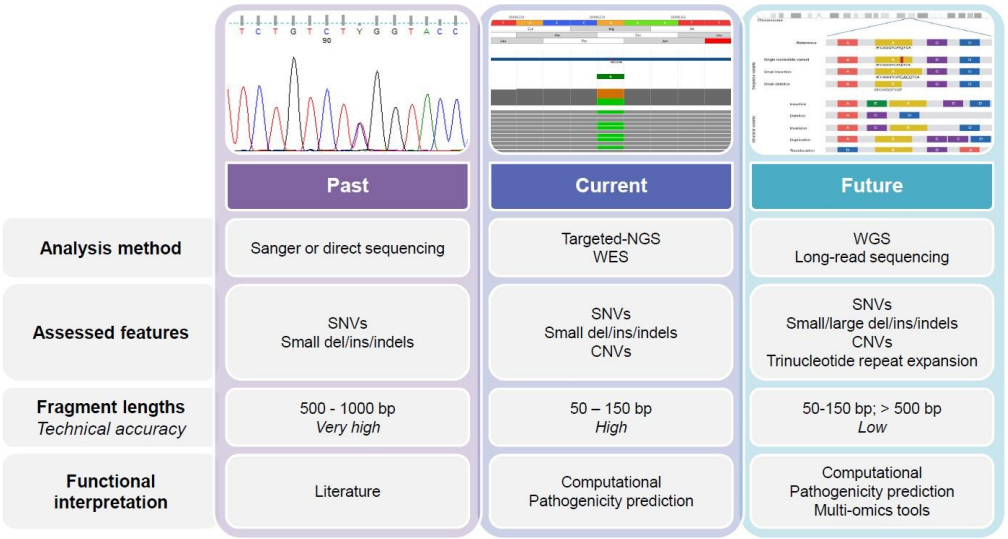


Figure 1.9. Past, present, and future of genetics and genomics technologies for the diagnosis of and research into hereditary anemias. The diagram shows the current analysis methods for DNA sequencing, while also highlighting the main differences among them. WES, whole-exome sequencing; WGS, whole-genome sequencing; SNV, single nucleotide variant; CNV, copy number variation. [Russo, Marra, et Al. 2020]

2. Aims of the study

CDAs are hereditary diseases that embrace a highly heterogeneous set of rare or very rare anemias that result from various kinds of abnormalities during late stages of erythropoiesis. Differential diagnosis, classification, and patient stratification among CDAs and clinically related hereditary anemias are often challenging. In the last few years, much effort on the genotype-phenotype correlation has been spent. In this context, the primary aim of this study was to establish an NGS-based diagnostic workflow for HAs and to demonstrate how it helped to unravel novel aspects of the molecular biology of erythropoiesis, not only in the field of CDAs but in the overall scenario of the hereditary anemias. Starting from the identification of the genetic background of our cohort of patients, the secondary aim of this study was to characterize novel causative variants not only to unravel their potential contribution in the pathogenesis of the disease, but also to assess the role of unexplored genes in physiological conditions.

3. Materials and Methods

3.1 Patients and genomic DNA preparation

Two hundred forty-four patients with clinical suspicion of different types of HA were included in the study. Clinical diagnosis was based on history, clinical findings, laboratory data, morphological analysis of peripheral blood and/or aspirated BM whenever available, and genetic testing. Local university ethical committees approved both the DNA sampling and the collection of patients' data from Medical Genetics Ambulatory in Naples (University Federico II, DAIMedLab). The blood samples of patients, after signed informed consent and according to the Declaration of Helsinki, have been processed to obtain the DNA necessary for the molecular analysis. Genomic DNA (gDNA) extraction was performed from peripheral blood using a Wizard Genomic DNA Purification Kit, following the manufacturer's instructions (Promega, Milan, Italy). Samples were quantified by NanoDrop machine (Thermo Scientific, Italy). Then, gDNA was loaded on 0,8% DNA agarose gel electrophoresis to check its integrity. Direct sequencing for *C15orf41*, and *SEC23B* genes was performed as previously described (Russo et al., 2013).

3.2 Targeted-NGS of HA patients

Libraries establishment

Genetic testing was achieved by targeted-NGS using three consecutive versions of a custom gene panel for hereditary RBC defects, namely RedPanel_rev.0, RedPanel_rev.1, and RedPanel_rev.3. In its current form, this panel is an updated version of a similar previously published one (Russo et al., 2018), and it is composed of 86 genes causative of CDAs, DBA, RBC

membrane defects, hemolytic anemias due to erythrocyte enzyme defects, anemias due to iron metabolism defects, hereditary hemochromatosis, and hereditary erythrocytosis. For the probe design, coding regions, 5′UTR, 3′UTR, 50 bp flanking splice junctions were selected as regions of interest. The probe design was performed by the web-based tool SureDesign (https://earray.chem.agilent.com/suredesign.htm, Agilent Technologies, USA). Sequence length was set at 150 x 2 nucleotides. The total probes' size was 298.393 kbp. Sample preparation was performed following the instruction's manufacturer for SureSelectQXT Target Enrichment for the Illumina Platform - SureSelect Custom Tier1 1-499 kb (Agilent Technologies).

Sequencing and data analysis

High-throughput sequencing was performed by Illumina MiSeq. The alignment of the sequencing reads to genomic locations, the quality control metrics, and the identification of variants have been achieved by Alissa Align and Call software (v1.1.2-2, Agilent Technologies). Variant annotation and analysis have been performed by Alissa Interpret software (v5.2.6, Agilent Technologies). According to the guidelines of the American College of Medical Genetics and Genomics (ACMG), we evaluated the pathogenicity of each variant by gathering evidence from various sources: population data, computational and predictive data, functional data, and segregation data. Due to the large range of prevalence in the population of these heterogeneous disorders, we selected both rare and low-frequency variants (MAF < 0.01 and 0.05, respectively), as reported by gnomAD browser (https://gnomad.broadinstitute.org/). InterVar (http://wintervar.wglab.org/) and Varsome (https://varsome.com/) web tools were used for the clinical interpretation of the new variants following the ACMG/AMP 2015 guideline (Richards et al., 2015). We selected and reported those variants with moderate pathogenic evidence, i.e. variants located in a mutational hot spot and/or critical and well-established functional domain. All the prioritized variants

were confirmed by Sanger sequencing and by the analysis of inheritance pattern, whenever possible. The validations were performed using 50 ng of genomic DNA. Custom primers were designed by the Primer3 program (Primer3 v. 0.4.0, freeware online).

Gene ranking

To rank the mutated genes based on their genic intolerance, we used the Residual Variation Intolerance Score (RVIS) percentile as retrieved from the database Genic Intolerance (http://genic-intolerance.org/). The RVIS is a gene-based score designed to rank genes in terms of whether they have more, or less common functional genetic variation relative to the genome-wide expectation given the amount of apparently neutral variation the gene has. A gene with a positive score has more common functional variation, and a gene with a negative score has less and is referred to as "intolerant" (Petrovski, Wang, Heinzen, Allen, & Goldstein, 2013).

3.3 Whole Exome Sequencing

Library preparation and sequencing

A total of 1.0µg of DNA per sample was used as input material for library preparation. Sequencing libraries were generated using Truseq Nano DNA HT Sample Preparation Kit (Illumina USA) following manufacturer's recommendations. gDNA was sonicated to a size of 350bp, and then fragments were end-polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further PCR amplification. At last, PCR products were purified (AMPure XP system) and libraries were analyzed for size distribution using the DNA Nano 6000 Assay Kit of Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) and quantified using real-time PCR.

The Whole Exome was captured with Agilent SureSelect Human All Exome V6 and the sequencing was performed on an Illumina HiSeq1500 platform.

Bioinformatic analysis of sequencing data

From raw reads, about 41 million per sample, sequencing artifacts were removed, including reads containing adapter contamination, low-quality nucleotides, and unrecognizable nucleotides. The cleaned data, the 98.41% of raw reads, were paired end reads of 150bp. We obtained high quality sequencing data. Indeed, the percentage of bases with quality scores above 20 and above 30 (Q20 and Q30) was 97.76 and 93.88, respectively. Mapping BAM files were obtained with BWA-mem (version 0.7.17) and SAMTools (version 1.8) by aligning the reads versus the GRCh37/hg19 reference genome assembly. On average, the 99.90% of reads were mapped and the 20.43% of duplicate reads were removed with Picard (version 2.18.9). On average, we covered the 99.63% of the target regions and average sequencing depth on target was 146.55x. The 95.00% of the target regions were covered with at least 20 reads which were sufficient for reliable variant calls. Single-nucleotide variants (SNVs) and small insertions and deletions (INDELs) were detected with GATK HaplotypeCaller. The functional annotation of variants was performed with ANNOVAR. We obtained 188,216 raw SNVs and 31,809 raw INDELs per sample. These variants were filtered to exclude common Single Nucleotide Polymorphisms (SNPs, with Allele Frequencies greater than 1%) in non-Finnish European populations of the 1000 Genomes Project, ExAC (v3) and GnomAD databases. To remove possible false positives, we eliminated variants falling in genomic duplicated regions. We removed off-target variants (eg.: intergenic, intronic, etc.). The set of exonic variants was then filtered to remove synonymous SNVs. Overall, about 500 variants per sample was kept for further analysis.

3.4 Plasma collection and ELISA assay

Plasma samples were collected from peripheral blood of patients and healthy controls (with informed consent, according to Declaration of Helsinki). Plasma levels of human ERFE and Hepcidin were quantified using ELISA kits (Cat. N° SKU #ERF-001, Intrinsic Erythroferrone IE, Cat. N° ICE-007, Intrinsic HEPCIDIN IDx^{TM}, Intrinsic Lifesciences, CA, USA) as previously described (Ganz et al., 2017). The ERFE and Hepcidin levels in each sample were determined through the fitting of a four-parameter logistic curve, according to the manufacturer protocol. Plasma levels of EPO and sTfR were quantified using ELISA kits (Quantikine IVD ELISA Human Erythropoietin, and Quantikine IVD ELISA Human sTfR, R&D System).

3.5 Cloning of C15orf41 coding sequence product into Expression Vector and Site-directed mutagenesis

cDNA encoding full-length wild-type (WT) C15orf41 sequence was cloned in the pCMV-Tag1 vector for mammalian cell expression (Invitrogen) in the BglII and XhoI sites, to obtain an N-terminal and C-terminal tagged protein, with FLAG and c-myc respectively. The point mutations c.281A > C, p.Tyr94Ser (Y94S), c.689A > C, p.His230Pro (H230P), and c.58 C > A, p.Pro20Thr (P20T) were introduced into the pCMV-Tag1 vector by using a QuikChange site-directed mutagenesis kit (Stratagene) as previously described (Russo et al., 2017). The coding sequence was analyzed by Sanger sequencing after mutagenesis.

3.6 Isolation of mononuclear cells, reticulocytes, and red blood cells membrane from peripheral blood

Peripheral blood was collected from patients, relatives, and healthy volunteers (with informed consent, according to Declaration of Helsinki) and

processed within 24 h. Peripheral blood mononuclear cells (PBMCs), reticulocytes, and RBC membrane were isolated using Ficoll-Hypaque, following the manufacturer's instructions (1.077-0.001 kg/L; Sigma-Aldrich, Milan, Italy).

3.7 Gene and protein expression analysis

RNA isolation, reverse transcription, and quantitative real-time PCR analysis

Total RNA was extracted either from peripheral blood leukocytes (PBLs), reticulocytes and from cell lines using TRIzol reagent (Life Technologies). Synthesis of cDNA from total RNA (2 mg) was performed using SensiFAST^{TM} cDNA Synthesis Kit (Bioline). Quantitative RT-PCR (qRT-PCR) of *C15orf41*, *FLAG-C15orf41*, and *CDAN1*, using Power SYBR Green PCR Master Mix (Applied Biosystems) was performed on Applied Biosystems 7900HT Sequence Detection System using standard cycling conditions. b-actin was used as internal control, while the Neomycin resistance gene was used as a control of transfection efficiency for K562 stable clones. Relative gene expression was calculated by using the $2^{-\Delta Ct}$ method, as described (Russo et al., 2013).

Protein isolation, subcellular fractionation, and western blotting analysis

Proteins were extracted from cell lines using RIPA lysis buffer containing protease inhibitor cocktail (1X). Subcellular fractionation in nuclear and cytoplasmic proteins was performed using NE-PER^{TM} Nuclear and Cytoplasmic Extraction Reagents (Thermo Fisher Scientific ^{TM}). Equal amounts of RBC membrane and protein from each lysate, as determined by a Bradford assay, were subjected to 12% SDS-PAGE and blotted onto polyvinylidene difluoride membranes (Biorad). Detection was performed with mouse anti-FLAG antibody (1:1000) (Sigma-Aldrich), rabbit anti-C15orf41

(1:500) (Atlas Antibodies, HPA061023), and rabbit anti-band-3 glycoprotein antibody (1:200) (Santa-Cruz Biotechnology). Since rabbit anti-C15orf41 was recommended for immunofluorescence (IF) we tested its specificity for western blotting (WB) by using C15orf41 over-expression cells as a positive control (Bordeaux et al., 2010). Mouse anti-TBP (TATA Binding Protein) (1:1000) (Sigma- Aldrich) and mouse anti-a-TUBULIN (1:5000) (Abcam) were used as a control for equal loading for cytosolic and nuclear proteins' extracts, respectively. Mouse anti-b-ACTIN (1:12000) (Sigma-Aldrich) and rabbit anti-GAPDH (1:1000) (Cell Signaling) was used as a loading control for total proteins' extracts and RBC membrane samples. Labeled bands were visualized and densitometric analysis performed with the BioRad Chemidoc using Quantity One software (BioRad) to obtain an integrated optical density (OD) value.

3.8 Cell cultures

Hek-293, HepG2, HuH7, MG-63, HEL, and K562 cells were obtained from American Type Culture Collection (ATCC, Manassas, VA, United States). Cells were maintained in Dulbecco's modified Eagle medium (DMEM) (Invitrogen) or RPMI 1640 medium (Invitrogen) supplemented with 10% fetal bovine serum (FBS)(Invitrogen), 100 U/mL penicillin (Invitrogen), and 100 mg/mL streptomycin (Invitrogen) in a humidified 5% CO2 atmosphere at 37 C, according to the manufacturer's instructions. Hek-293 cells (400×10^3) were transfected with pCMV-Tag1-C15orf41 plasmids (2.5 mg/well) using the DNA Transfection Reagent (TransFectin Lipid Reagent, Bio-Rad) according to the manufacturer's procedures. Cells were collected 16, 24, and 48 h after the transfection to perform RNA and protein extractions.

Stable clones production

For K562 stably over-expressing FLAG-C15orf41 gene, 1 x $10^{\wedge 6}$ cells were transfected with pCMV-Tag1-C15orf41 plasmids (2μg), using Hily Max DNA Transfection Reagent (Dojindo Laboratories). After 48 hours, Neomycin (0.6 mg/mL) was added as a selection marker. Clones were generated by plating on a 48-well an appropriate number of cells and diluting them to decrease the cell number (limiting dilution method) progressively. Resistant clones were expanded, and then only the FLAG C15orf41 over-expressing clones were selected for the following experiments. C15orf41 WT and mutant clones were also confirmed using *HinfI* or *FatI* enzymes, as predicted by Restriction of DNA sequences Tool (http://insilico.ehu.eus/restriction/two_seq/index.php, freeware online).

Drug treatment

Erythroid differentiation of K562-C15orf41 stable clones (2 x 10^5/mL) was performed adding 50 mM hemin (Sigma) to the culture medium, after 24 h of starvation by FBS depletion, as previously described (Andolfo et al., 2010). Cells were collected before hemin addition (0 days) and two days after hemin addition (2 days).

Inhibition of proteasome activities in K562-C15orf41 stable clones was performed by MG132 treatment. Briefly, cells were treated with 10 μM of the inhibitor MG132 (Sigma-Aldrich), and collected 2, 4, and 8 hours after MG132 administration. DMSO was used as vehicle.

Cell cycle synchronization in K562-C15orf41 stable clones was performed following a double thymidine block assay: cells were treated 19 hours with RPMI containing 3mM of thymidine (Sigma-Aldrich), then seeded 9 hours in standard growth medium, and finally re-treated 16 hours with RPMI supplemented with 2mM or 3mM of thymidine. Cells were collected 0, 4, 8, and 24 hours after the release from thymidine.

3.9 Flow cytometry

Determination of cell cycle distribution was performed on K562 stable clones by flow cytometric analysis. Cells were harvested by centrifugation, resuspended in PBS containing 3.75% Nonidet P-40, 100 mg/ml RNase A (Invitrogen) and 40 mg/ml propidium iodide (Sigma-Aldrich), and incubated at room temperature for 3 hours in the dark. The cell antigen profile was analyzed using CD71-APC antibody (Miltenyi Biotec), and CD235a (Glycophorin A)-PE-Vio770 antibody (Miltenyi Biotec) by flow cytometry. Samples were analyzed on a FACS flow cytometer (Becton Dickinson Immunocytometry Systems, BDIS).

3.10 Immunofluorescence analysis

For immunofluorescence (IF) analysis, 3 x 10^5 cells were fixed for 10 min in 4% Paraformaldehyde (PFA, Sigma) and washed in 50 mM PBS/NH4Cl. After washing in PBS 1x, cells were allowed on 35 mm IBIDI m-Dishes (Ibidi) coated with 0.05% poly-L-lysine (Sigma-Aldrich) to adhere. Permeabilization was performed with 0.2% Triton/PBS, followed by blocking with 1% BSA/PBS. The seeded cells were immunologically stained with rabbit anti- C15orf41 antibody (1:25) (Atlas Antibodies, HPA061023), mouse anti-NUCLEOPHOSMIN (1:200), and secondary antibodies (1:200) (Alexa Fluor 546 anti-rabbit, and Alexa Fluor 488 anti-mouse, Life Technologies). Nuclei were stained with 1 mg/ml DRAQ5 (Abcam) in PBS for 15 min at room temperature. Cells were preserved in PBS 1x and imaged using a LEICA TCS SP8 meta confocal microscope, equipped with an oil immersion plan Apochromat 63x objective 1.4 NA. The following settings were used: Green channel excitation of Alexa488 by the argon laser 488 nm line was detected with the 505–550 nm emission bandpass filter. Red channel excitation of Alexa546 by the Helium/Neon laser 543 nm line was detected with the 560–700 nm emission bandpass filter (using the Meta monochromator). Blue

channel excitation of DRAQ5 by the blue diode laser 647 nm and emission bandpass filter.

3.11 Statistical and bioinformatics analyses

Statistical significance of differences in protein and gene expression was determined using the Mann-Whitney test or Student's t-test. Correlation analysis of *C15orf41* with *CDAN1* gene expression was performed by Pearson correlation test. A two-sided p-value < 0.05 was considered statistically significant. For the *in-silico* correlation analysis between *C15orf41* and *CDAN1* gene expression in normal hematopoietic cell subpopulations we used the dataset "Normal Hematopoietic Subgroups – (GEO ID: gse19599)," stored in the R2: Genomics Analysis and Visualization Platform2, a biologist-friendly, web-based genomics analysis, and visualization application.

4. Results

4.1 NGS-based genetic testing improves diagnostic yield of HAs

4.1.1 NGS-based genetic testing for hereditary anemias

We collected 244 patients with different clinical suspicion. Patients were analyzed by three different versions of a custom gene panel: (1) RedPanel_rev.0, which contained 34 genes responsible for HAs, showed high sensitivity and specificity. Indeed, approximately 97% of analyzable target bases were covered by at least 20 reads and showed a diagnostic yield of 55.2%. (2) RedPanel_rev.1 analyzed 71 HAs genes, with a 20 reads-coverage of 93% and a diagnostic yield of 62.7%. (3) The last version, RedPanel_rev.2, contained 86 genes, increasing the diagnostic yield to 83.9% (Figure 4.1). Overall, we reached a final diagnosis in 193/244 patients (79.1%). All the undiagnosed patients (20.9%) were analyzed or recommended for WES (Figure 4.2).

Among all analyzed patients, 35.2% (86/244) presented a clinical suspicion of CDA. Interestingly, only 16.3% of them (14/86) showed a molecular defect compatible with the suspicion (matched phenotype-genotype). Indeed, the multi-gene approach modified the original diagnosis of CDA in 64% of patients (72/86) (Figure 4.3).

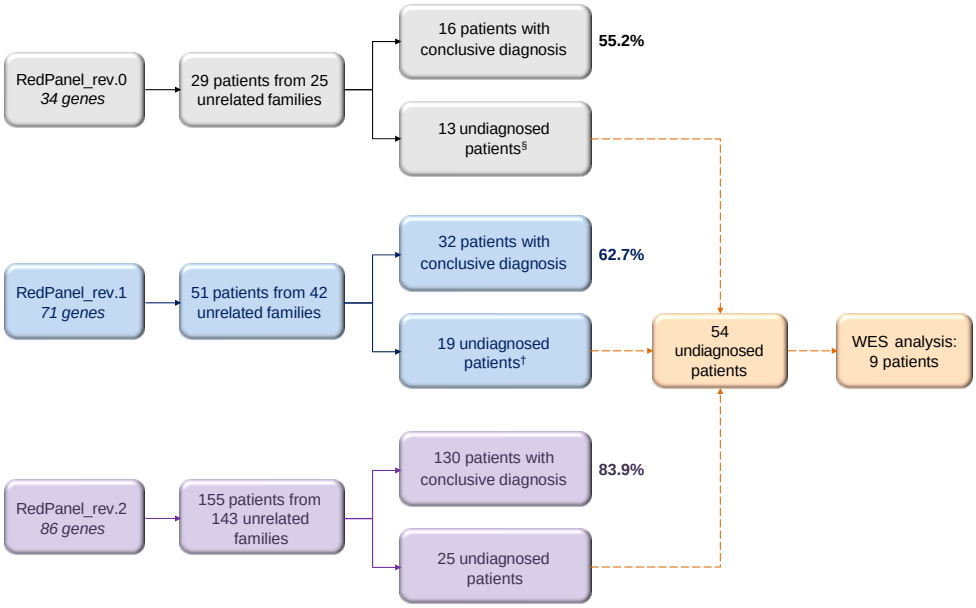


Figure 4.1. Diagnostic workflow for hereditary anemias. First t-NGS-analysis was performed on 29 patients from 25 unrelated families by RedPanel_rev.0, composed of 34 genes. 13 patients received no diagnosis. RedPanel_rev.1, composed of 71 genes, was used to analyse 51 patients from 42 unrelated families. Overall, 19 patients were undiagnosed. 155 patients from 143 unrelated families were analyzed by RedPanel_rev.2, composed of 86 genes. 25 patients received no diagnosis. Of 54 un-diagnosed patients, 9 of them underwent to WES analysis. Percentages refers to the diagnostic yield of each t-NGS panel. §Among the 13 undiagnosed patients included in the first analysis, six of them underwent the second round by RedPanel_rev.1: two patients obtained a conclusive diagnosis, while the remaining four probands resulted negative again. †Among the 19 undiagnosed patients analyzed by RedPlex_rev.1 we also counted the four undiagnosed patients previously analyzed.

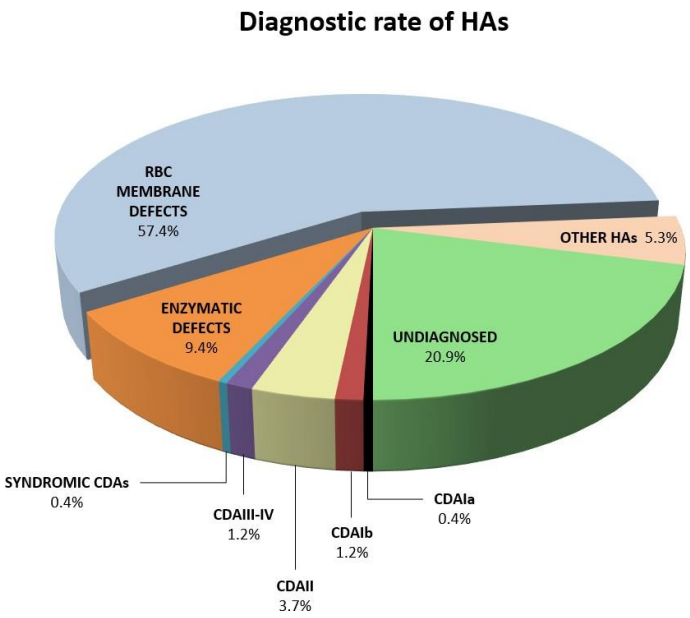


Figure 4.2. Diagnostic rate of HA patients. Pie chart showing the frequencies of the different HA subtypes diagnosed after genetic testing by RedPanel_Rev.0, _Rev.1, _Rev.2. The frequency of each condition was calculated as the ratio between the number of patients in each HA subtype and the overall count of patients tested (n = 244 patients). RBC Membrane defects include hereditary spherocytosis, hereditary elliptocytosis, hereditary dehydrated stomatocytosis, southeast asian ovalocytosis, cryohydrocytosis, and pseudohyperkalemia. Enzymatic defects refer to adenylate kinase deficiency, pyruvate kinase deficiency, hexokinase deficiency, glucose phosphate isomerase deficiency, and glucose-6-phosphate dehydrogenase deficiency. Syndromic CDAs refer to one patient with a mutation in the CAD gene. Other hereditary anemias (HA) include sideroblastic anemia, sitosterolemia, polycythemia, and iron-refractory iron deficiency anemia.

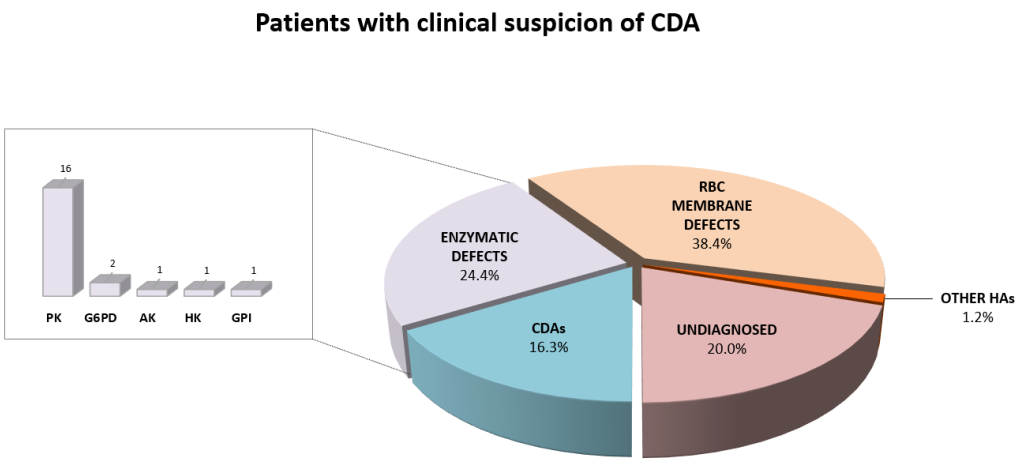


Figure 4.3. Molecular diagnosis of patients with clinical suspicion of CDA. Pie chart showing the frequencies of the different HA subtypes diagnosed in patients with a clinical suspicion of CDA. The frequency of each condition was calculated as the ratio between the number of patients in each HA subtype and the overall count of patients suspected of CDA (n = 86 patients). RBC Membrane defects include hereditary spherocytosis, hereditary dehydrated stomatocytosis, southeast asian ovalocytosis, and cryohydrocytosis. Other hereditary anemias (HA) include sideroblastic anemia. The histograms showed the enzymatic defects identified in 24.4 % of patients: pyruvate kinase deficiency (n=16 patients), glucose-6-phosphate dehydrogenase deficiency (n=1 patient), adenylate kinase deficiency (n=1 patient), hexokinase deficiency (n=1 patient), and glucose phosphate isomerase deficiency (n=1 patient).

4.1.2 Misdiagnosis of PKD patients as CDAs is due to PKD dyserythropoietic component

Among patients originally suspected of CDA, 24.4% of them (21/86) exhibited a final diagnosis of chronic anemia due to enzymatic defects (Figure 4.3). Interestingly, 16 patients clinically referred as CDA showed mutations in *PKLR*, the causative gene of PK deficiency. Of note, these patients presented clinical data and morphological features of the bone marrow mostly resembling those of both CDAI and CDAII. Intriguingly, almost all the patients originally classified as CDA and subsequently diagnosed as chronic hemolytic anemia due to enzyme defects were transfusion-dependent, with subsequently reduced reliability of the enzyme assays. Starting from this assumption, to investigate on dyserythropoietic component of PKD patients, we compared the levels of different markers of erythropoiesis of PKD patients to those of CDA patients, which are classically characterized by ineffective erythropoiesis and increased levels of the erythroid negative regulator of hepcidin, ERFE (Russo et al., 2016). As we expected, ELISA assay for EPO showed significantly increased levels in PKD patients (n=16) compared to the reference range (3.1 - 14.9 mlU/mL) and CDA patients (n=57) (Figure 4.4A). In addition, PKD patients showed significantly increased plasma levels of sTfR compared to the reference range (9.2 - 22.3 nmol/L) and CDA patients (n=41) (Figure 4.4B). We also measured ERFE concentration in the same cohort of PKD patients, which showed increased levels compared to the reference range (0.1 - 3.8 ng/mL) measured in plasma samples of healthy controls (n=24). Of note, PKD patients exhibited ERFE levels similar to those measured in CDA patients (n=60) (Figure 4.4C). Finally, we also evaluated hepcidin expression by ELISA assay. As expected, both CDA (n=21) and PKD (n=17) patients showed reduced levels of hepcidin compared to the median of the reference values in men (112 ng/mL) and women (65 ng/mL), even though PKD levels of hepcidin were significantly higher than those measured in CDA patients (Figure 4.4D).

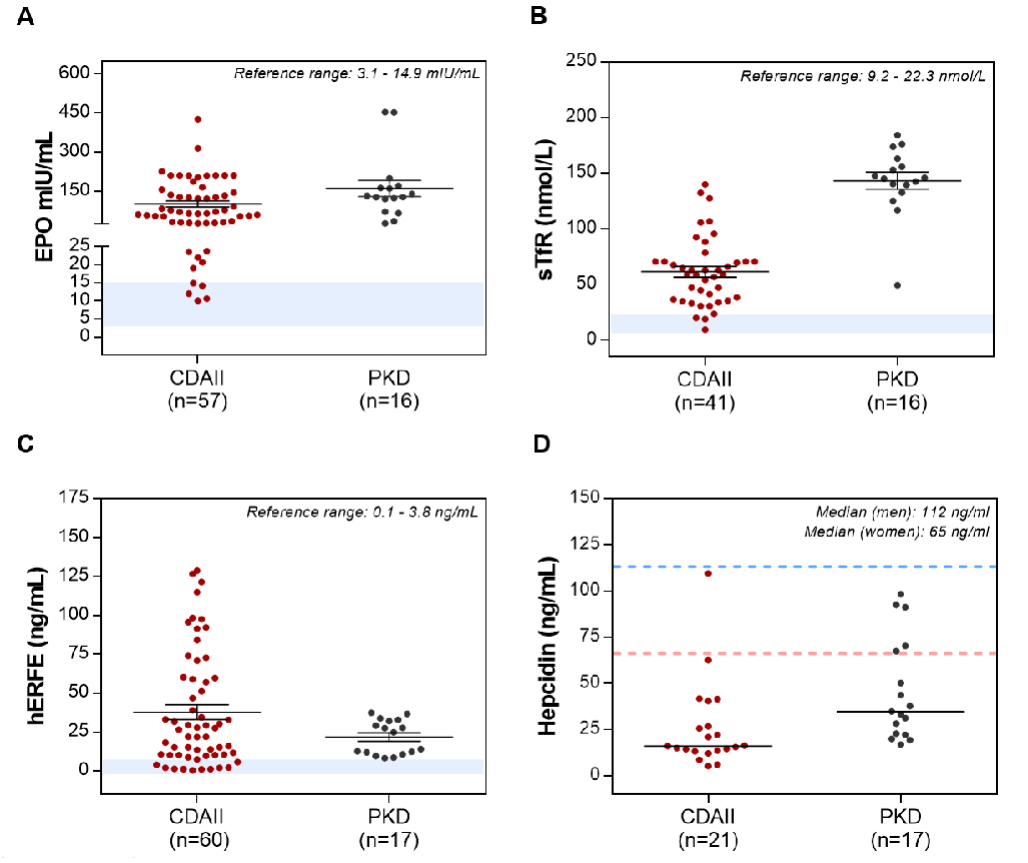


Figure 4.4. EPO, sTfR, ERFE, and hepcidin expression in plasma from patients with CDAII and PKD by ELISA assay. (A) Quantification of EPO levels in plasma from 57 patients with CDAII and 16 patients with PKD. Reference range (3.1 – 14.9 mIU/mL) is shown in light blue. Data are means ± standard error. P Value (P= 0.03) was calculated by student t test. **(B)** Quantification of sTfR levels in plasma from 41 patients with CDAII and 16 patients PKD. Reference range (9.2 – 22.3 nmol/L) is shown in light blue. Data are means ± standard error. P Value (P<0.0001) was calculated by student t test. **(C)** Quantification of ERFE protein levels in plasma from 60 patients with CDAII and 17 patients with PKD. Reference range (0.1 – 3.8 ng/mL), shown in light blue, is calculated on ERFE plasma levels of 24 healthy controls. Data are means ± standard error. P Value (P=0.08) was calculated by student t test. **(D)** Quantification of hepcidin-25 levels in plasma samples from 21 CDAII patients and 17 patients with PKD. Data are shown as median. Median reference values are indicated by light blue line (men, 112 ng/ml), and pink line (women, 65ng/ml). P value (P= 0.002) was calculated by Mann-Whitney Test.

4.2 Description of novel cases of CDA Ib and syndromic CDA

Within the case series of patients herein described, the present thesis focused on (i) a rare case of syndromic CDA due to a novel variant in the *CAD* gene and (ii) CDAI cases due to pathogenic variants in *C15orf41*.

4.2.1 Uridine treatment restores the CDAII-like hematological phenotype in a patient with homozygous mutation in the CAD gene

RP0_12/RP1_51 patient was a 3-year-old Cambodian male child from consanguineous parents (Figure 4.5A), presented to neurology with non-verbal autism and intractable seizures. Red cell macrocytosis and an abnormal peripheral blood film were reported. At 6.5 years of age, autism, developmental delay, and generalized epilepsy, with partial response to levetiracetam were present. He also exhibited mild to borderline anemia (hemoglobin [Hb], 9.8-10.5 g/dL), relative reticulocytopenia (reticulocyte absolute count, 39.0 × 103/µL), markedly increased red blood cell distribution width (19.0 %), and signs of hemolysis. He had not received any transfusions. Ferritin and transferrin saturation were within normal limits. Fetal Hb and HbA2 were increased (Table 4.1). The peripheral blood smear showed marked anisopoikolocytosis with macrocytes, occasional microcytes and spherocytes, teardrop cells, fragmented red cells, and rare nucleated erythrocytes. The bone marrow aspirate revealed moderate erythroid hyperplasia with dyserythropoiesis, 5% to 10% binucleated late erythroblasts, rare tri-nucleate forms, and prominent cytoplasmic bridging (Figure 4.5B, left panel). No abnormalities of the fingers, toes, or nails were seen. The genomic analysis of the proband, by our t-NGS RedPanel, highlighted the presence of the ultra-rare missense homozygous variant c.5366G>A, p.Arg1789Gln (rs1341652994, minor allele frequency A = 0.00000 [1/251066, GnomAD_exome]) in the *CAD* gene. SNP-array analysis excluded genomic imbalance associated with autism but showed a loss-of-heterozygosity region that encompassed the *CAD* gene

(Data not shown). These data were consistent with a recessive pattern of the inheritance. The same variant was present in both parents in the heterozygous state, with normal hematological phenotypes (Figure 4.5A). Biallelic mutations in the *CAD* gene have been reported, in only 4 additional families, to be causative of early infantile epileptic encephalopathy-50. Interestingly, all patients affected by this condition also show mild anemia with marked anisopoikolocytosis. In the first report describing the *CAD* mutation in early infantile epileptic encephalopathy-50, slightly abnormal migrations for the erythrocyte RhAG protein and band 3 glycoprotein were reported on polyacrylamide gel electrophoresis, which suggested abnormal glycosylation. Consistent with this finding, we performed western blotting analysis on red blood cell membrane samples isolated from the proband and parents. The immunoblot analysis of isolated erythrocyte membrane proteins from the patient (II.2) and father (I.1) revealed a slight reduction of protein values of band 3 compared to the mother (I.2) and a healthy control, although not statistically significant. However, there was no indication of narrower band size or different migration of band 3 in this patient, compared to a CDAII *SEC23B*-related patient (Figure 4.5C). At age 11, treatment with oral uridine supplementation was initiated. Follow-up of the patient after 22 months of uridine treatment demonstrated normal hematological and biochemical parameters, and normal iron balance (Table 4.1). The blood smear showed normocytic red cells with no significant anisopoikilocytosis (Figure 4.5B, right panel). A bone marrow aspirate showed absence of erythroid hyperplasia, binucleated erythroblasts, or cytoplasmic bridging (Figure 4.5B).

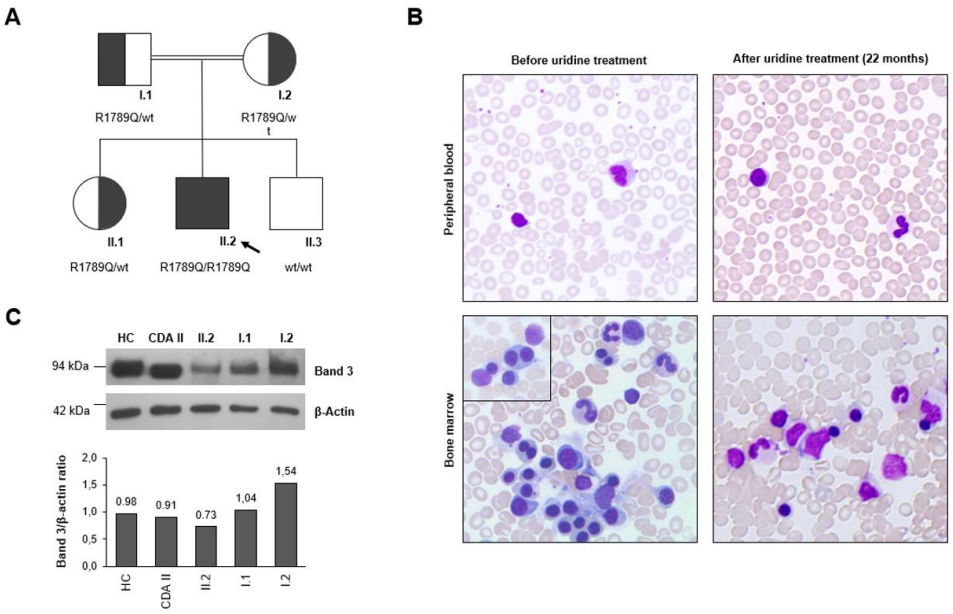


Figure 4.5. Characterization of a novel uridine-responsive CDAII-like patient with CAD deficiency. (**A**) Family pedigree of the patient. Squares, males; circles, females; solid symbols, affected persons; black arrow, the proband. According to the autosomal recessive inheritance pattern, consanguineous parents of the proband (II.2) are heterozygous for the variant c.5366G>A, p.Arg1789Gln in the *CAD* gene. (**B**) Peripheral blood and bone marrow smears before uridine treatment (left) and after 22 months of uridine treatment (right). (**C**) Top: Immunoblot of band 3 glycoprotein in the membrane protein lysate from red blood cells of a healthy control (HC), a CDAII SEC23B-related patient, the proband (II.2), and the parents of the proband (I.1, I.2). Bottom: Densitometric analysis of the blot.

Table 4.1. Clinical data of CAD patient before and after treatment

	Before uridine		**After uridine**		***Reference range***
	6.5-years-old	*11-years-old*	*6-months*	*22-months*	
Hematological data					
RBC ($x10^6$/µL)	3.8	4.1	5.2	5.4	4.0-5.2
Hb (g/dL)	11.3	12.3	15.0	16.2	11.5-15.5
Ht (%)	34.0	37.0	47.0	48.0	35-45
MCV (fL)	90.0	92.0	90.0	88.0	77-95
MCH (pg)	28.8	30.2	31.0	30.0	25-33
MCHC (g/dL)	32.0	32.8	34.5	34.0	32-36
RDW (%)	19.0	22.0	12.8	13.9	11.5-15
Retics abs count ($x10^3$/µL)	39.0	63.9	53.6	86.0	20-90
WBC ($x10^3$/ µL)	6.0	13.0	5.3	8.1	5-14.5
PLT ($x10^3$/ µL)	393	249	325	248	150-450
MPV (fL)	8.8	9.1	8.0	8.2	6.5-10
HbA2 (%)	5.0	5.5	3.1	3.0	2.5
HbF (%)	1.8	3.1	1.1	1.2	< 1.0
Biochemical data and iron balance					
Total bilirubin (mg/dL)	0.82	0.94	0.35	0.53	0.3 to 1.0
Unconj. bilirubin (mg/dL)	0.64	0.76	0.23	0.41	0.2 to 0.8
LDH (U/L)	176	176	200	242	115 - 211
Haptoglobin (g/L)	0.20	0.13	0.80	1.7	0.3-2.0
ALT (U/L)	16.0	24.0	16.0	38.0	< 30
Serum iron (µg/dL)	117.3	123	106	55.9	53-119
Ferritin (ng/mL)	70.0	43.0	51.0	199	10-55

RBC, red blood cells; Hb, hemoglobin; Ht: Hematocrit MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; RDW, red blood cells distribution width; Retics, reticulocytes; PLT, platelets; MPV: Mean platelet volume; LDH; lactate dehydrogenase; ALT, alanine transaminase.

4.2.2 Identification of novel causative mutations in the uncharacterized C15orf41 gene

RP1_3 patient (Family A) was a 7-years-old female, second child from healthy non-consanguineous parents of Italian origin (Sardinia). At birth, cholestatic hepatopathy, dysmorphic features (bilateral syndactyly of the IV–V toes), and severe anemia (Hb 5.5 gr/dl) were observed. Family history was not indicative of anemia. At diagnosis, the proband presented transfusion-dependent normocytic anemia with a blood transfusion frequency every 15–20 days, and low reticulocyte count (Table 4.2). BM analysis showed: erythroid hyperplasia with 6% of cells showing megaloblastic features, nuclear abnormalities, and nuclear/cytoplasmic maturation asynchrony; 4% of erythroblasts were bi- and tri-nucleated; the granulopoietic/erythropoietic ratio (G:E) = 0.53. A substantial percentage of erythroblasts showed inter-nuclear bridges (5%), a typical feature of CDAI. Accordingly, genetic testing RedPanel_rev.1 revealed the presence of the transversion c.281A>C in the homozygous state, resulting in a novel aminoacidic substitution p.Tyr94Ser (Y94S). It is an ultra-rare variant (rs587777101) with a minor allele frequency (MAF) C = 0.00001 in the ExAC database. In agreement with the recessive inheritance pattern, both parents were heterozygous (Figure 4.6A).

RP0_17 patient (Family B) was a 2.4-years-old male, born from 3rd degree consanguineous parents of Turkish origin. At birth, recurrent pneumonia, thoracic dysplasia, and short limbs were observed. Family history was negative for anemia or jaundice. The proband presented transfusion-dependent normocytic anemia (12 transfusions/year), low reticulocyte count, growth retardation, and increased ferritin level, suggesting an iron loading condition (Table 4.2). No splenomegaly was observed at physical examination and abdominal echography. BM analysis showed severe megaloblastic changes and normoblasts with double or multiple nuclei, a morphological feature suggestive of CDAII. Accordingly, we firstly performed Sanger sequencing analysis for

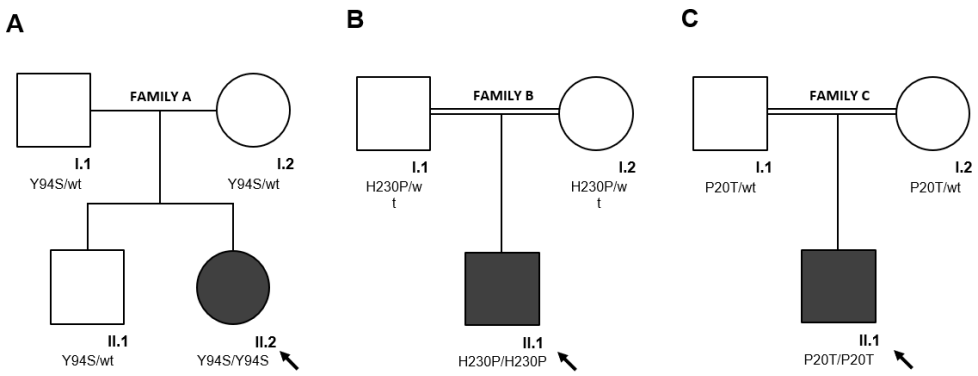
CDAII-disease gene *SEC23B*, finding no causative variants. Then, as a second-step analysis, we enrolled the patient in our multi-gene panel for hereditary anemias, identifying the transversion c.689A>C in *C15orf41* in the homozygous state, resulting in the amino acid substitution p.His230Pro (H230P). In agreement with the recessive inheritance pattern, both parents were heterozygous (Figure 4.6B).

RP0_22 patient (Family C) was an 8.4-years-old male, born from consanguineous parents of Turkish origin. The proband presented splenomegaly and macrocytic anemia, without any dysmorphic remarks (Table 4.2). He never underwent to transfusion regimen. BM analysis showed normoblasts with double or multiple nuclei, a morphological feature typical of CDA II, but Sanger sequencing analysis of *SEC23B* gene and RedPanel_rev.1 evidenced no causative mutations. Thus, we performed a WES analysis, identifying the transversion c.58C>A in *C15orf41* in the homozygous state, resulting in the amino acid substitution p.Pro20Thr (P20T). Both parents were heterozygous for the variant, in agreement with the recessive inheritance pattern, as is shown in the family pedigree (Figure 4.6C).

Table 4.2 Clinical features of the C15orf41-CDAI affected patients

	(A-II.2)	**(B-II.1)**	**(C-II.1)**	**Reference range‡**
Age at diagnosis	7 years	2.4 years	8.4 years	-
Transfusion regimen	Yes	Yes	No	-
Dysmorphic remarks	Toes syndactyly	Thoracic dysplasia; short limbs	-	-
Bone marrow features	Erythroblasts with inter-nuclear bridges	Normoblasts with double or multiple nuclei	Normoblasts with double nuclei	-
Complete blood count				
RBC ($x10^6/\mu L$)	2.72	3.67	3.71	3.9-5.6
Hb (g/dL)	7.8	10.6	11.6	11.0-16.0
Hct (%)	22.2	31.4	34.9	33.0-45.0
MCV (fL)	81.6	85.4	94	70.0-91.0
MCH (pg)	20.2	28.8	31.3	23.0-33.0
MCHC (g/dL)	24.8	33.8	33.3	23.0-33.0
Retics %	5.8	1.0	0.4	0.5-2.0
Retics count ($x10^3/\mu L$)	158000	36700	14840	-
PLT ($x10^3/\mu L$)	-	518.0	298.0	150.0-450.0
Biochemical, laboratory data and iron balance				
Total bilirubin (mg/dL)	1.90	1.46	1.34	0.2-1.2
LDH (U/L)	779	511	-	125.0-243.0
Ferritin (ng/mL)	825	1512	158	22.0-275.0
TSAT (%)	75	89	-	15.0-45.0

‡Reference ranges from AOU Federico II, University of Naples, Italy
RBC, red blood cells; Hb, hemoglobin; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; Retics, reticulocytes; PLT, platelets; LDH; lactate dehydrogenase; TSAT: Transferrin saturation.



D

C15orf41 variants			
Family	**Family A (RP1_3)**	**Family B (RP0_17)**	**Family C (RP0_22)**
HGVS (Coding)[a]	c.281A>C	c.689A>C	c.58C>A
HGVS (Protein)[b]	p.Tyr94Ser	p.His230Pro	p.Pro20Thr
RefSeq ID	rs587777101	-	-
MAF	C=0.00001	-	-
InterVar (evidence codes)[§]	Pathogenic (PS1, PS3, PM2, PP4)	Likely pathogenic (PS3,PM2,PP4)	Likely pathogenic (PM2, PM5, PP2, PP4)

Figure 4.6. Pedigrees of the C15orf41-CDAI families. Squares denote males, circle females, solid symbols affected persons. The black arrow indicates probands. (**A**) The pedigree of family A is shown. According to the autosomal recessive inheritance pattern, the parents of A-II.2 are heterozygous for the variant c.281A > C, p.Tyr94Ser. (**B**) The pedigree of family B is shown. According to the autosomal recessive inheritance pattern, consanguineous parents of B-II.1 are heterozygous for the variant c.689A > C, p.His230Pro. (**C**) The pedigree of family C is shown. According to the autosomal recessive inheritance pattern, consanguineous parents of C-II.1 are heterozygous for the variant c.58C>A, p.Pro20Thr. (**D**) Pathogenicity prediction of *C15orf41* mutations. Human Genome Variation Society (HGVS) coding (a): NM_001130010; HGVS protein (b): NP_001123482.

[§]InterVar evidence scores by the website http://wintervar.wglab.org/evds.php: PS1, same amino acid change as an established pathogenic variant; PS3, well-established functional studies show a deleterious effect; PM2, absent (or at extremely low frequency if recessive) in population databases; PP4, patient's phenotype is highly specific for a single gene etiology.

4.3 characterization of C15orf41 in physiological and pathological erythropoiesis

4.3.1 C15orf41 and CDAN1 gene expression are directly correlated

To evaluate the effect of the identified mutations on *C15orf41* gene expression, we initially analyzed *C15orf41* expression in PBLs isolated from the two probands and healthy controls (HCs). No difference in gene expression levels of the proband A-II.2 compared to those detected in HCs was observed, suggesting that the Y94S variant does not affect gene expression. Conversely, we found a marked downregulation of C15orf41-H230P in the second proband B-II.1 (Figure 4.7A). Likewise, we saw a similar trend of *CDAN1* expression in the two patients. Notably, the proband A-II.2 did not show any alterations of *CDAN1* expression compared to those seen in HCs, while the B-II.1 proband revealed a decrease of *CDAN1* expression level, although not statistically significant (Figure 4.7B). To note, we also assessed the *C15orf41* expression in reticulocytes isolated from 14 healthy controls (HC), and we observed a similar expression detected in PBLs (Figure 4.7C). Moreover, a direct correlation between *C15orf41* and *CDAN1* expression genes in healthy subjects was observed ($r = 0.62$, $p = 0.0006$) (Figure 4.7D). We achieved comparable results by gene expression profiling of different human cell lines (Hek-293, HepG2, HuH7, MG-63, HEL, and K562 cells), where a significant direct correlation between *C15orf41* and *CDAN1* expression was observed (Figure 4.7E). Finally, we confirmed the *ex vivo* data on *C15orf41-CDAN1* correlation by *in silico* analysis of the expression dataset for normal hematopoietic cell subpopulations, obtained by R2 database (Figure 4.7F).

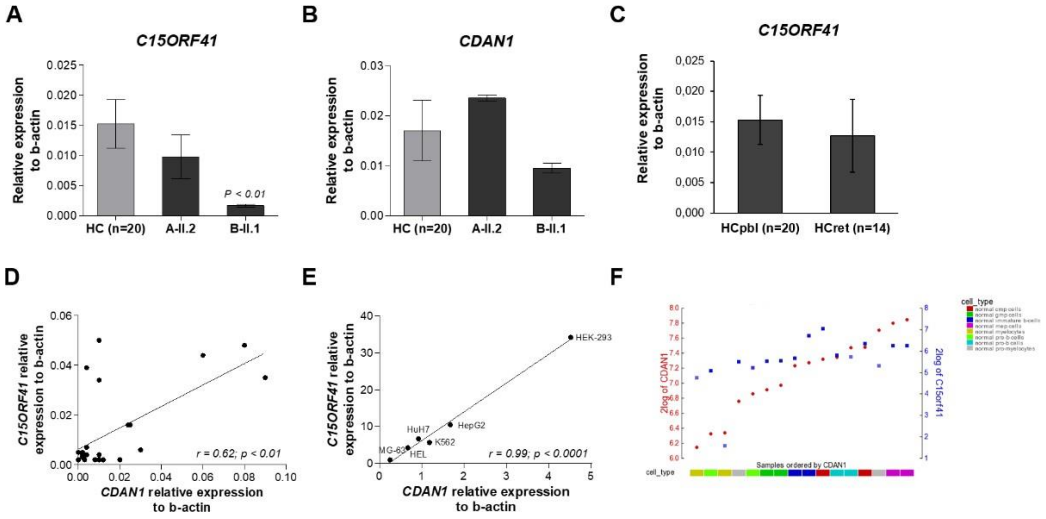


Figure 4.7. Analysis of the C15orf41 and CDAN1 expression. (**A**) *C15orf41* mRNA relative expression to b-actin of patients A-II.2, B-II.1, and HCs (n = 20) are shown. Data are presented as mean ± SE. P-value by Student's t-test. (**B**) *CDAN1* mRNA relative expression to b-actin of patients A-II.2, B-II.1 and HCs (n = 20) is shown. Data are presented as mean ± SE. (**C**) *C15orf41* mRNA relative expression to β-actin in both peripheral blood leukocytes (pbl) and reticulocytes (ret) from healthy controls (HC) is shown. Data are presented as mean ± SE. (**D**) Correlation analysis between *C15orf41* and *CDAN1* gene expression, performed on 20 HCs and two probands, showed a direct correlation between the two genes (Pearson correlation r = 0.62, p = 0.01). (**E**) Correlation analysis between *C15orf41* and *CDAN1* expression, performed in MG-63, HEL, K562, HuH7, HepG2, and Hek-293 cell lines. MG-63, bone osteosarcoma cells; HEL, human erythroblasts; HuH7, hepatocellular carcinoma cells; HepG2, hepatocellular carcinoma cells (Pearson correlation r = 0.99, p < 0.0001). (**F**) YY-plot of correlation analysis performed by R2 database to investigate the expression level of C15orf41 and CDAN1 genes in the expression dataset for normal flow-sorted hematopoietic cell subpopulation (GEO ID: gse19599); CMP, common myeloid progenitor; GMP, granulocyte-macrophage progenitor. Pearson correlation r = 0.64, p = 0.008.

4.3.3 C15orf41 localization into nuclear and cytosolic compartments

We carried out an in vitro study by cloning the cDNA sequence of *C15orf41* into the pCMV-tag1 expression vector, to obtain the protein of interest tagged with both FLAG and C-MYC tags. Then, Y94S, H230P, and P20T point mutations were introduced into the pCMV-tag1-C15orf41 wild type (WT) plasmid by site-directed mutagenesis. To investigate the expression and subcellular localization of C15orf41, we transiently transfected Hek293 cells with pCMV-tag1-C15orf41-WT. Time-course analysis showed a gradual increase of *C15orf41* gene expression in cells transfected with WT clone at 16, 24, and 48 h compared to those transfected with empty vector (EV) (Figure 4.8A). Conversely, WB analysis on the same harvested cells revealed a marked increase of C15orf41 protein level at 16 h after transfection, with a progressive decrease of the C15orf41-FLAG signal, which resulted highly down-regulated at 48 h after the transfection (Figure 4.8B). Moreover, we assessed the endogenous protein levels and localization of the protein by both WB and IF on a nuclear and a cytosolic fraction of Hek-293 cells. Both analyses confirmed that the protein was mainly expressed in the nucleus, but also in the cytosol compartment, even if in a smaller amount, suggesting a role of the protein in these two cellular compartments (Figures 4.8C-D). No co-localization of C15orf41 with nucleoli was observed (Figure 4.9).

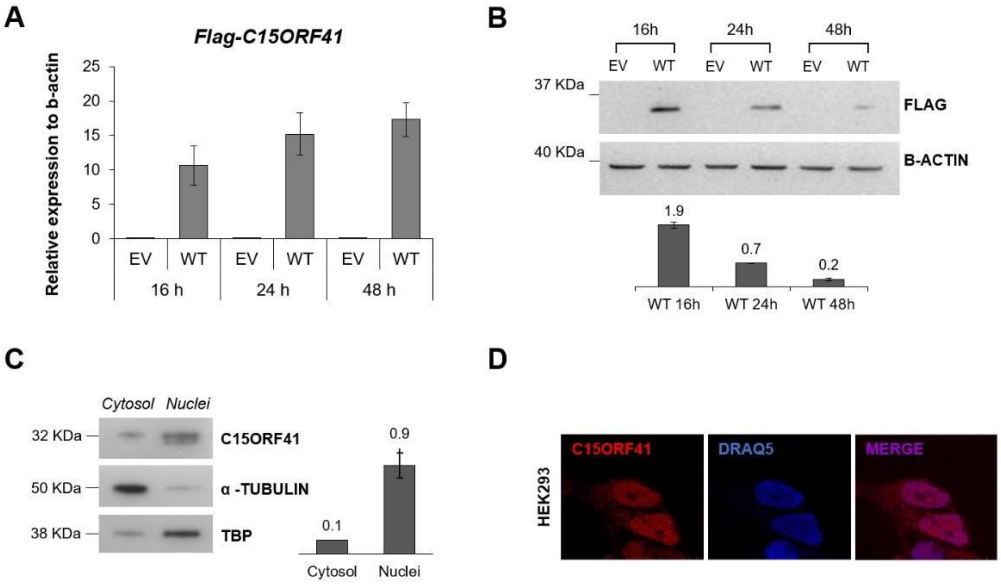


Figure 4.8. C15orf41-WT expression and subcellular localization. (**A**) The panel shows FLAG-C15orf41 mRNA relative expression to b-actin of Hek-293 cells over-expressing pCMV-tag1-C15orf41 WT compared to those transfected with empty vector (EV) at 16, 24, and 48 h after the transfection. Data from two different transfections are presented as mean ± SD. (**B**) The panel shows WB analysis of Hek-293 cells over-expressing pCMV-tag1-C15orf41 WT compared to those transfected with EV at 16, 24, and 48 h after the transfection. b-actin is loading control. Sizes (in kDa) are on the left. The histogram shows the densitometric quantification based on b-actin amount. Data derived from two experiments are presented as mean ± SD. (**C**) WB on cytosolic and nuclear fractions of Hek-293 cells showing C15orf41 expression. TBP and a- TUBULIN are shown as a loading control of nuclear and cytosolic compartments, respectively. The histogram shows the densitometric quantification based on TBP and a- TUBULIN amounts. Data derived from three experiments are presented as mean ± SD. Sizes (in kDa) are on the left. (**D**) Immunofluorescence analysis of Hek-293 cells is shown. Rabbit anti-C15orf41 antibody was used to stain C15orf41 protein. DRAQ5 was used as a nuclear marker. Overlapping of both signals (MERGE) is shown on the right.

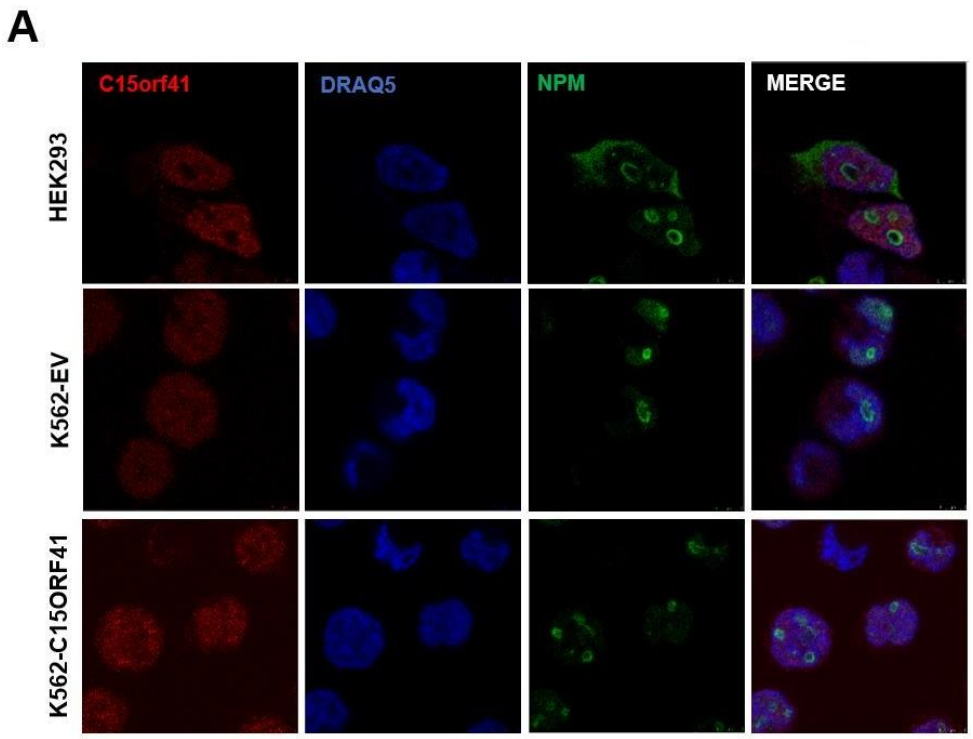


Figure 4.9. C15orf41 localization in nucleoli. (**A**) Immunofluorescence analysis of Hek-293 and K562 cells is shown. Rabbit anti-C15orf41 antibody was used to stain C15orf41 protein. DRAQ5 was used as a nuclear marker. Nucleophosmin (NPM) was used as nucleolar marker. Overlapping of both signals (MERGE) is shown on the right.

4.3.4 Establishment of K562 C15orf41-WT and mutant stable clones

For further investigations and characterization of C15orf41 WT and mutants we needed a stable cellular model that allowed long-term as well as defined and reproducible expression of the gene of interest. Thus, we proceeded with the establishment of K562 cells stably overexpressing pCMV-tag1-C15orf41-WT, pCMV-tag1-C15orf41-Y94S, and pCMV-tag1-C15orf41-H230P plasmids. We chose K562 cell line, as this cellular model is widely used as erythroid model, able to be induced to erythroid differentiation by hemin treatment. Selection and expansion of clones is described in materials and methods section. Each K562 clone was screened by qRT-PCR analysis to evaluate Neomycin relative expression to β-actin. Clones that showed similar expression levels of Neomycin, underwent to further analysis. Thus, we selected two clones for each plasmid type: empty vector (EV#3 - #7), wild-type (WT#2 - #3), H230P (H230P #10 - #3), and Y94S (Y94S#5 - #8) (Figure 4.10A). To validate the presence of both mutations in homozygous state in K562 stable clones, the sequence of each clone was investigated by digestion with restriction enzymes, capable to produce different pattern to distinguish WT clones from mutant ones. The choice of the proper restriction enzymes was performed with the Restriction of DNA sequences Tool (see Materials and Methods). gDNA from healthy controls and patients carrying the selected mutations were used as positive controls of the digestion (Figure 4.10B).

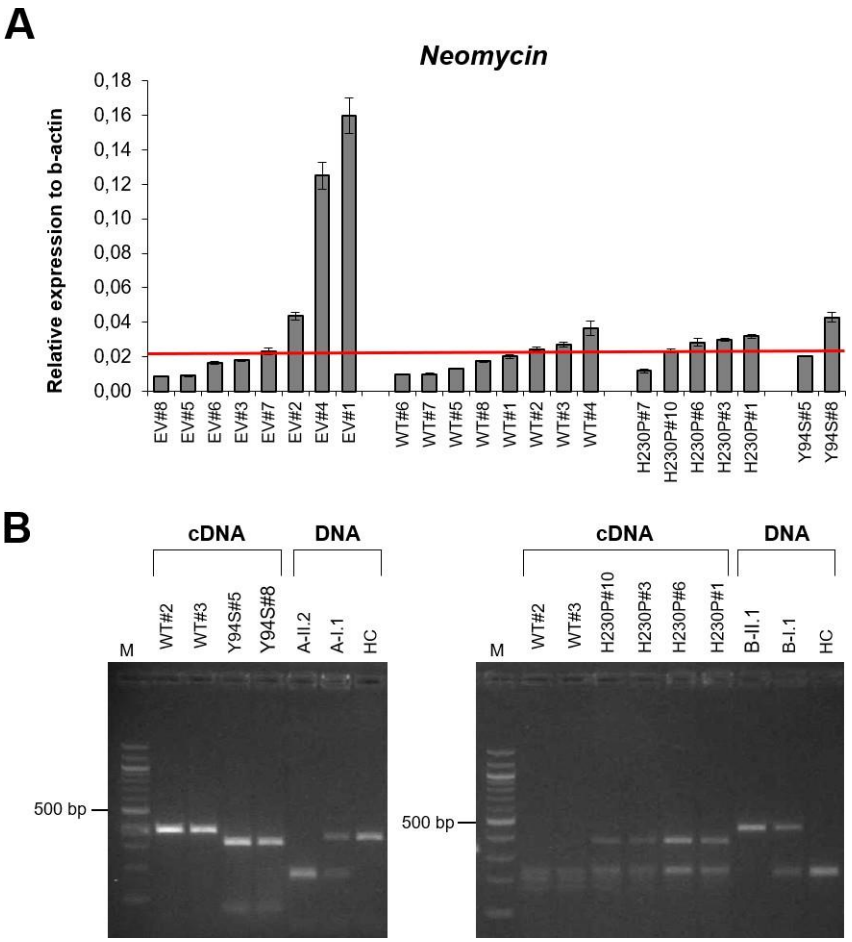


Figure 4.10. Selection of K562 stable clones for C15ORF41-WT, C15ORF41-Y94S, and C15ORF41-H230P. (A) Neomycin relative expression to b-actin of K562 clones over-expressing pCMV-tag1-C15orf41-WT, pCMV-tag1-C15orf41-H230P, and pCMV-tag1-C15orf41-Y94S. Data are presented as mean ± SD of three replicates. Clones that showed similar expression levels, whose levels are along the median value of Neomycin expression (red line), underwent to further analysis. (**B**) Left panel: digestion pattern of C15orf41-Y94S over-expressing clones and family A carrying the Y94S variant by *HinfI* enzyme. The Y94S variant creates a new restriction site for *HinfI* enzyme, generating a 298 bp-fragment at cDNA level and a 164 bp-fragment at DNA level, while no restriction products are obtained in the presence of WT genotype. On the right, digestion pattern of C15orf41-H230P over-expressing clones and family B, carrying the H230P variant, by *FatI* enzyme. The H230P variant abrogates the restriction site for *FatI* enzyme, generating the 357 bp- and 222 bp-fragments at cDNA level, while no restriction products were obtained in the presence of the mutated base at DNA level. M=marker 100bp; HC, healthy control.

4.3.5 C15orf41 is regulated by proteasome degradation

To further investigate on the presence of both C15orf41 isoforms in our cellular system during cell cycle, we compared protein sequences of C15orf41 isoform 1 and 2 by the Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi). We observed that, even though the isoform 2 was missing of the first 99 amino acids, both isoforms preserved the predicted K176 ubiquitination site (Figure 4.11A). To demonstrate the involvement of the ubiquitin-proteasome pathway in the degradation of C15orf41 isoforms, we treated K562 C15orf41-WT cells with MG132 proteasome inhibitor. Then, we tested the efficacy of the treatment on C15orf41 protein expression by WB analysis on cells collected at different time points after MG132 administration. We observed the gradual accumulation of the protein, from 0 to 24 hours, in treated cells compared to the untreated one, demonstrating that C15orf41 is degraded via the ubiquitin-proteasome pathway (Figure 4.11B).

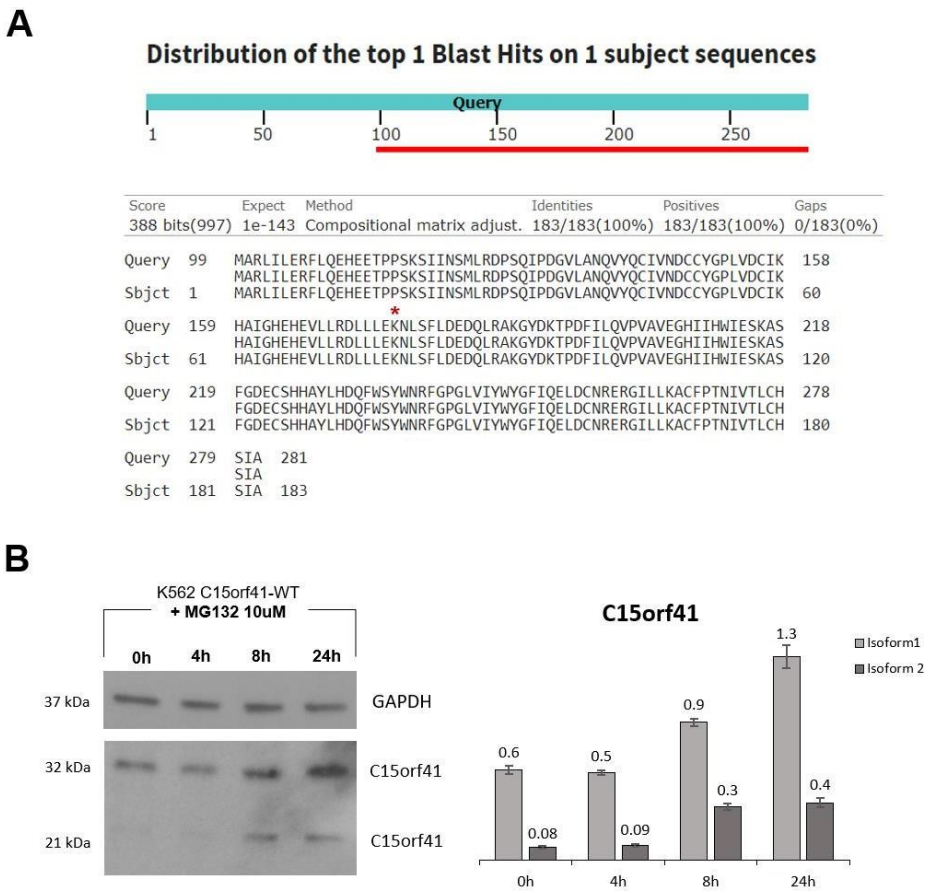
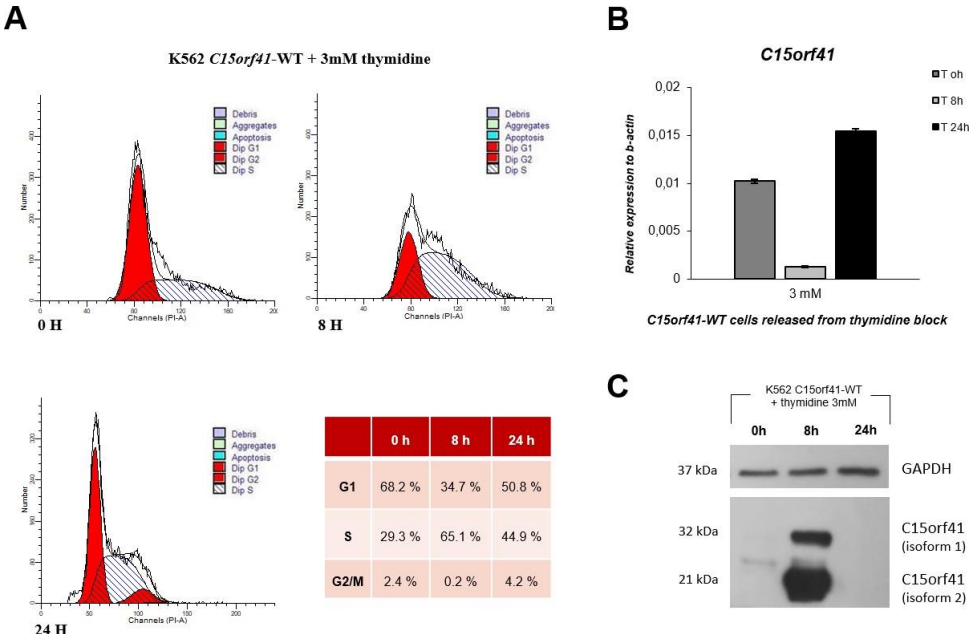


Figure 4.11. C15orf41 degradation by ubiquitin-proteasome pathway. (**A**) Comparison between protein sequence of the two C15orf41 isoforms by BLAST online tool. Query sequence correspond to the sequence of the isoform 1 (281 aa), while isoform 2 (183 aa) was entered as subject sequence. 100% of identity is shown for the sequence shared between the two isoforms. Red asterisk highlighted the lysine (K) residue predicted as ubiquitination site of the canonical isoform 1. (**B**) WB on K562 C15orf41-WT cells 0, 4, 8, and 24 hours after MG132 treatment showing C15orf41 expression. GAPDH is shown as a loading control. The image shows C15orf41 isoform 1, with the predicted molecular weight of 32 KDa, and the isoform 2 with the predicted molecular weight of 21 KDa.

4.3.6 C15orf41 is expressed during G1/S phase of cell-cycle

To investigate on C15orf41 modulation during cell cycle, we synchronized K562 cells stably overexpressing C15orf41-WT through a double thymidine block assay. Treatment with double thymidine results in a G1-phase arrested cell population. We monitored progression of the cells through cell cycle, particularly 0, 8, and 24 hours after the synchronization with thymidine (Figure 4.12A). Then, at each time point, we measured both C15orf41 gene and protein expression levels. We observed that *C15orf41* gene expression was enhanced 0 and 24 hours after the release from thymidine when compared to the expression of the same cells measured 8 hours after the treatment (Figure 4.12B). On the contrary, C15orf41 protein amount was increased 8 hours after the release from thymidine respect to the expression observed at the other time points (0 and 24 hours) (Figure 4.12C).



	0 h	8 h	24 h
G1	68.2 %	34.7 %	50.8 %
S	29.3 %	65.1 %	44.9 %
G2/M	2.4 %	0.2 %	4.2 %

Figure 4.12. C15orf41 expression during cell-cycle. (**A**) Determination of cell cycle distribution by flow cytometric analysis. Histograms representing PI-A staining 0, 8, and 24 hours after the release from thymidine block of K562 C15orf41-WT treated with 3mM thymidine. At the bottom, the table resumes the percentage of cells for each phase of the cell cycle, at each time point. (**B**) C15orf41 relative expression to b-actin of K562 C15orf41-WT cells after 0, 8, and 24 hours from thymidine block. Data are presented as mean ± SD of three replicates. (**C**) WB on K562 C15orf41-WT cells 0, 8, and 24 hours after thymidine block showing C15orf41 expression. GAPDH is shown as a loading control. The image shows C15orf41 isoform 1, with the predicted molecular weight of 32 KDa, and the isoform 2 with the predicted molecular weight of 21 KDa.

4.3.7 Characterization of C15orf41 expression in Y94S, H230P, and -P20T mutants

To study in vitro the pathogenetic effect of the variants identified in patients, we evaluated gene expression and protein level of C15orf41-H230P, Y94S, and P20T transient and stable mutants. As we already demonstrated with our *ex vivo* analysis (Figure 4.6A), we observed a sharp decrease of both FLAG-C15orf41 gene expression and protein levels in Hek-293 cells over-expressing C15orf41-H230P mutant compared to C15orf41-WT ones. Conversely, only a slight reduction in gene expression and protein level in Hek-293 cells overexpressing C15orf41-Y94S was observed (Figures 4.13A-B). We confirmed these data also in K562 C15orf41-H230P and -Y94S mutants (Figure 4.13C-D). Moreover, the analysis of C15orf41 gene and protein expression on Hek293 cells transfected with C15orf41-WT and C15orf41-P20T, showed no impairment of mRNA expression in P20T mutant compared to C15orf41-WT (Figure 4.13E-F). Overall, we observed that H230P mutation, which occurs in the PD-(D/E) XK nuclease domain of the protein (161-259 aa), affect C15orf41 mRNA stability and protein amount, while Y94S and P20T mutations, which occur in the DNA-binding domain of the protein (4-129 aa), preserve C15orf41 expression.

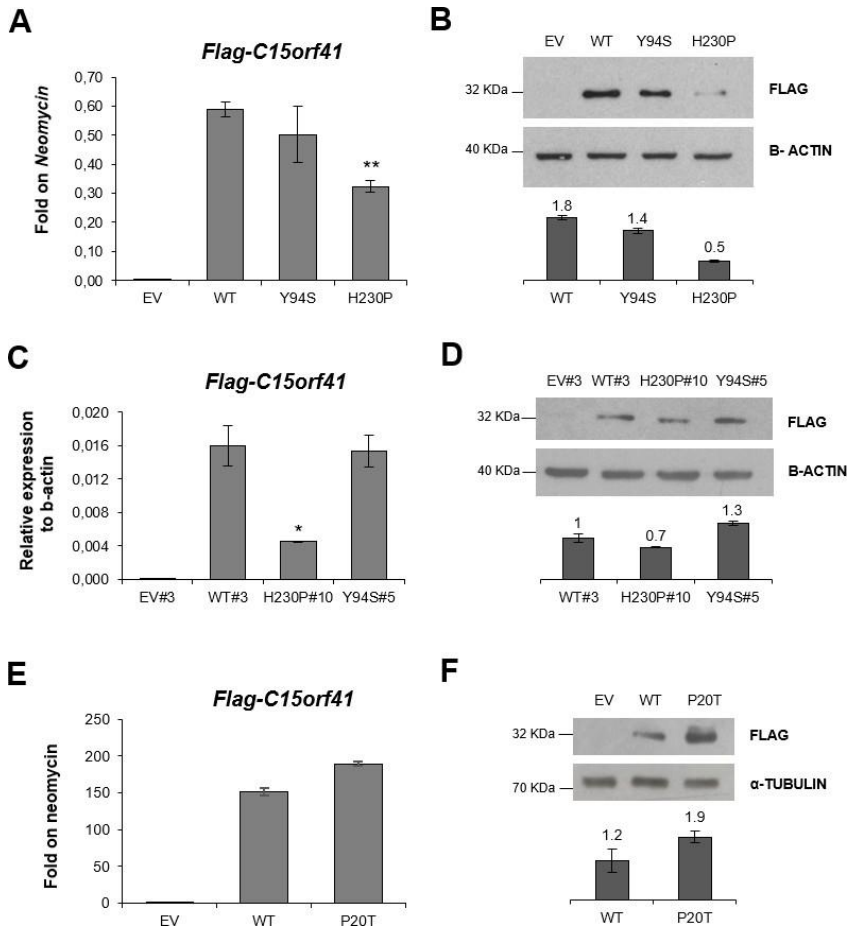


Figure 4.13. In vitro characterization of C15orf41 mutants. (A) FLAG-C15orf41 relative expression to β-actin of HEK-293 cells over-expressing pCMV-tag1-C15ORF41 WT, pCMV-tag1-C15orf41-Y94S, and pCMV-tag1-C15orf41-H230P compared to those transfected with EV at 16h after the transfection. P value has been calculated by Student t-test. *P = 0.03. **(B)** WB analysis of HEK-293 cells over-expressing pCMV-tag1-C15orf41 WT, pCMV-tag1-C15orf41-Y94S, and pCMV-tag1-C15orf41-H230P compared to those transfected with EV at 16h after the transfection. The histograms show the densitometric quantification based on the β-actin amount as a standard control. Sizes (in kDa) are on the left. **(C)** FLAG-C15orf41 mRNA relative expression to b-actin of K562 cells stably over-expressing C15orf41-WT, -Y94S, and -H230P compared to the EV. Data derived from two experiments are presented as mean SD. P value by Student's t-test. $P < 0.05$. **(D)** WB analysis of K562 cells stably over-expressing C15orf41-WT, -H230P, and -Y94S compared to the EV. The histogram shows the densitometric quantification based on the b-actin. Data derived from two experiments are presented as mean SD. Sizes (in kDa) are on the left. **(E)** FLAG-C15orf41 relative expression to β-actin of HEK-293 cells over-expressing pCMV-tag1-C15ORF41 WT and pCMV-tag1-C15orf41-P20T compared to those transfected with EV at 16h after the transfection. **(F)** WB analysis of HEK-293 cells over-expressing pCMV-tag1-C15orf41 WT and pCMV-tag1-C15orf41-P20T compared to those transfected with EV at 16h after the transfection. The histograms show the densitometric quantification based on the α-tubulin amount as a standard control. Sizes (in kDa) are on the left.

4.3.8. C15orf41-Y94S and -H230P mutations affect erythroid differentiation and cell cycle dynamics

To investigate if the presence of the mutations could affect erythroid differentiation, we treated K562 C15orf41-WT, -Y94S, and -H230P cells with hemin. Evaluation of CD71 and CD235 differentiation markers showed a statistically significant decreased percentage of $CD71^{+}/CD235^{+}$ cells in both Y94S and H230P clones compared to the WT one (Figure 4.14A). Moreover, we observed a slight increase of the rate of S-phase at cell cycle analysis in K562 cells over-expressing Y94S and H230P mutants compared to WT, although not statistically significant (Figure 4.14B-C).

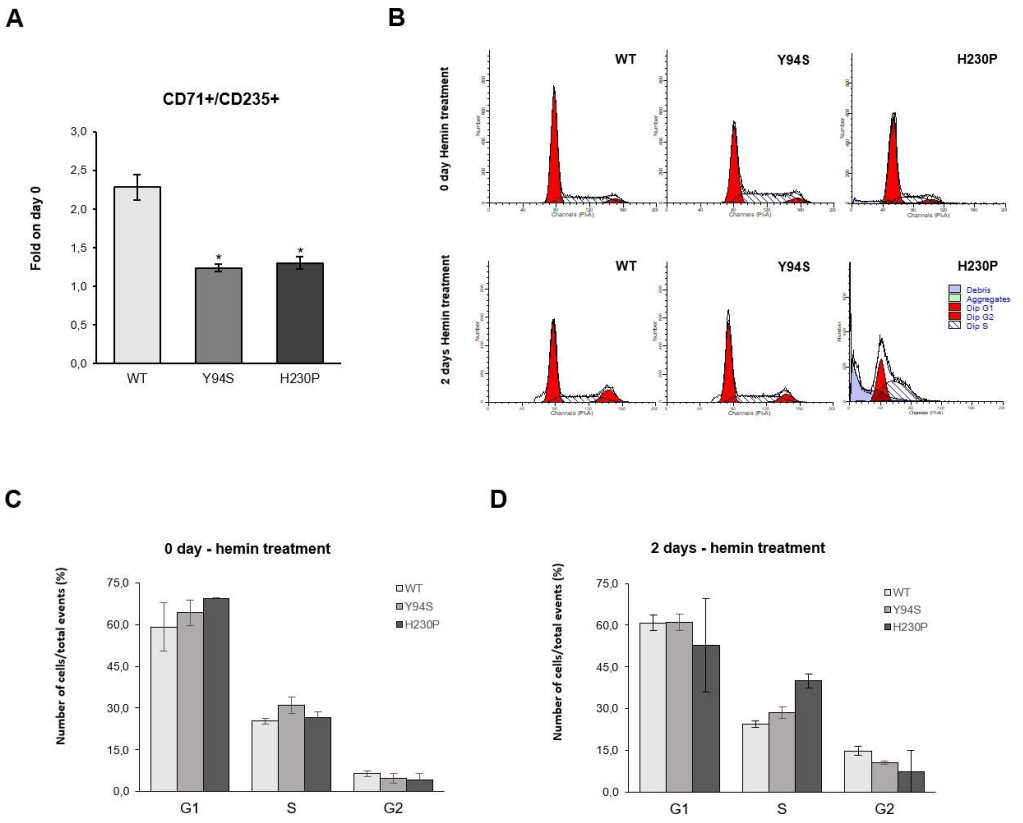


Figure 4.14. Analysis of K562-C15orf41 over-expressing clones during hemin-induced erythroid differentiation. (**A**) Erythroid differentiation markers of C15orf41-K562 stable clones. The histogram shows the percentage of $CD71^{+}/CD125^{+}$ cells at two days of hemin treatment normalized on untreated cells (0 days). Data derived from two experiments are presented as mean ± SD. P value by Student's t-test. $^{*}P < 0.05$. (**B**) Cell cycle distribution of C15orf41-K562 stable clones at 0 and two days of hemin treatment. Cell cycle was analyzed using FACS analysis of propidium iodide-stained cells. (**C**) The histograms show the number of K562 over-expressing pCMV-tag1-C15orf41 WT, -Y94S, and -H230P untreated cells on total events (%) in G1, S, and G2 phases of the cell cycle. Data derived from two experiments are presented as mean ± SD. (**D**) The histograms show the number of K562 over-expressing pCMV-tag1-C15orf41 WT, -Y94S, and -H230P cells on total events (%) in G1, S, and G2 phases of the cell cycle, two days after hemin treatment. Data derived from two experiments are presented as mean ± SD.

5. Discussion

The hereditary anemias represent a heterogeneous group of disorders with rare to low frequency that are characterized by complex genotype-phenotype correlations. This group includes conditions with quite different phenotypes, ranging from (1) hyporegenerative anemias, as CDAs, (2) hemolytic anemias due to RBC membrane defects, as HS and HSt, (3) hemolytic anemias due to enzymatic deficits, as PKD. In recent years, major advances have been made to understand the genetic basis and the pathophysiology of HAs. Indeed, more than 80 genes involved in RBC physiology have been identified as causative of HAs. Although the diagnostic workflow of these conditions is a normal clinical practice, differential diagnosis, classification, and patient stratification among HA are often difficult. Indeed, the variety of unspecific and overlapping phenotypes often hampers a correct clinical management of the patients. In our experience, although Sanger sequencing is still a valuable starting point when clinical features of the patient are typical of a specific HA, second-generation sequencing (i.e., NGS) has revolutionized the identification of such rare disorders. Indeed, in the last decade, several custom gene panels have been developed for the diagnosis of HAs, providing a diagnostic yield from 38% to 87% (Russo et al., 2020).

In this study, we collected 244 patients with clinical suspicion of CDAs admitted to our Medical Genetics Unit from 2018 to 2020. All the cases followed a diagnostic workflow based on different versions of a custom targeted-NGS panel, named RedPanel (Russo et al., 2018) that, in its last version, is composed of 86 genes causative of hereditary erythrocyte defects. Overall, we obtained a diagnostic yield ranging from 55.2 to 83.9%, in agreement with the different version of RedPanel. According to the frequency of the disease and the clinical suspicion, we found 16.3% of patients originally

suspected of CDA carrying mutations in CDA-related genes. Of note, our data pointed out the problem of overlapping phenotypes. Indeed, 64% of patients (72/86) initially suspected for CDA were diagnosed as other HAs, which included HS, DHS, red-cell enzymatic defects (i.e., PKD), and sideroblastic anemia. Among them, 16 patients carried mutations in *PKLR* gene, suggesting that misdiagnosis of PKD with CDA could be due to shared disease mechanisms between the two conditions. Several studies suggested that metabolic abnormalities in R-PK deficiency could alter the differentiation of erythroid progenitors and are associated to enhanced and ineffective erythropoiesis (Aizawa et al., 2005; Aizawa et al., 2003). Thus, to demonstrate the occurrence of dyserythropoiesis in PKD, we measured the levels of different markers in our cohort of patients (n=17). As we expected, we found significantly increased levels of EPO in PKD patients compared to reference values in healthy subjects (EPO= 3.1 - 14.9 mlU/mL). Of note, even with a slight difference, EPO levels were also significantly increased when compared to those of CDA patients, probably due to a greater hemolytic component and the higher degree of anemia in PKD than CDA. It is well known that the concentration of sTfR1 reflects erythropoiesis rate (Kohgo, Torimoto, & Kato, 2002). Expansion of erythroid precursors to compensate a loss of RBCs lead to an increased level of sTfR1 due to an increase of highly expressing $TfR1^{+}$ cells (Richard & Verdier, 2020). Indeed, we observed increased levels of sTfR in PKD patients compared not only to the reference range in healthy subjects (sTfR= 9.17 - 22.35 nmol/L) but also to sTfR levels measured in CDA patients. It was already demonstrated that ERFE levels, which is produced by erythroblasts in response to EPO, are increased in both CDAII and β-thalassemia patients (Kautz et al., 2015; Russo et al., 2016) due to ineffective erythropoiesis. Thus, we measured ERFE plasma levels of our cohort of PKD patients and compared their levels to those of CDA. Similarly, PKD patients showed increased levels of ERFE compared to those of healthy controls, but comparable values compared to those of CDAII patients. Several studies

demonstrated that hepcidin is suppressed by increased erythropoietic iron demand (Delaby et al., 2005; Nemeth et al., 2004), and that the erythroblast-derived hormone ERFE is responsible for hepcidin suppression in conditions of ineffective erythropoiesis. Thus, we also measured hepcidin plasma levels of our cohort of PKD patients, observing significant decreased values compared to the reference range in healthy controls. On the other hand, we observed that suppression of hepcidin in PKD patients was not comparable to those affected by CDA, even though they showed higher levels of ERFE, suggesting that there is no direct correlation between their concentration. These data corroborate the fact that ERFE is not the only erythroid factor that contribute to hepcidin suppression. Thus, further investigations in our cohort of patients should be carried out to evaluate the contribute of other known and putative hepcidin suppressors. Overall, our analysis provides an efficient approach to correctly identify and diagnose PKD in patients, frequently misdiagnosed as CDAI-II. Indeed, our data demonstrated the occurrence of ineffective erythropoiesis in these patients, even though further investigations are needed to unravel the mechanism by which this process is established. Additionally, our findings could have a crucial impact on clinical management and therapeutic approach of these patients. Indeed, novel targeted treatments for PKD, as mitapivat (AG-348), an oral small molecule allosteric activator of PK, have been developed and are currently in phase III clinical trials in adults (Grace et al., 2019; Kung et al., 2017; Yang et al., 2019).

The evidence that our diagnostic approach is a valuable starting point for precise medicine is a paradigmatic case of syndromic CDA that we identified by our t-NGS panel (RedPanel_rev1). The analysis of the proband revealed the presence of an ultra-rare missense homozygous variant c.5366G>A, p.Arg1789Gln in the *CAD* gene. It encodes a multifunctional enzyme complex that catalyzes the first steps of de novo pyrimidine biosynthesis. Of note, biallelic mutations in this gene have been associated to early infantile epileptic encephalopathy-50 in five patients, a severe neurodegenerative disease which

can be lethal in childhood. Interestingly, all patients showed also mild CDA-like anemia with marked anisopoikolocytosis (Koch et al., 2017; Ng et al., 2015). Consistent with the other described cases, our patient showed non-verbal autism, intractable seizures, developmental delay, and generalized epilepsy. He also presented mild to borderline macrocytic anemia, relative reticulocytopenia, markedly increased red blood cell distribution width, and signs of hemolysis. We originally included *CAD* gene in the list of candidate loci for CDA in the second version of our panel, RedPanel_rev.1, since it was firstly reported to be a candidate gene for congenital disorders of glycosylation affecting the erythrocyte RhAG protein and band 3 glycoprotein5 (Ng et al., 2015). However, the immunoblot analysis of isolated erythrocyte membrane proteins from the patient did not reveal narrower band size or different migration of band 3 compared to a CDAII *SEC23B*-related patient, but only reduced protein values of band 3. This result was not surprising, as reduced expression of band 3 has also been frequently reported for patients with CDAII (Hunt et al., 2015). Since supplementation with oral uridine had been suggested as treatment for CAD-deficient patients (Koch et al., 2017), treatment with oral uridine supplementation was initiated. Importantly, the patient was weaned off all antiepileptic drugs, and the follow-up of the patient after 22 months of uridine treatment demonstrated normal hematological and biochemical parameters, and normal iron balance. These paradigmatic cases pointed out even more clearly that molecular diagnosis is pivotal in the diagnostic workflow of these disorders. Thus, beyond achieving a definitive diagnosis, knowing the genetic basis of these patients is valuable also for guiding treatment.

Besides the therapeutic implications of this diagnostic approach, our t-NGS panel for HAs resulted valuable to various questions, not only in identifying novel causative variants, but also in understanding the molecular mechanisms underlying these rare and unexplored diseases, such as CDA type Ib.

C15orf41, recently named *CDIN1*, is the gene identified as causative of this condition (Babbs et al., 2013). We herein identified and described three unrelated cases of *C15orf41*-CDAI. The first patient (RP1_3) presented clinical characteristics, hematological status, and morphological features of erythroblasts compatible with a suspicion of CDAI. Particularly, BM analysis highlighted the presence of a substantial amount of inter-nuclear bridges between erythroblasts, the typical feature of CDAI. Indeed, t-NGS analysis by RedPanel_rev.1 of the patient revealed the presence of the novel variant c.281A>C, p.Tyr94Ser (Y94S) in *C15orf41*. This variant resulted annotated on public databases as ultra-rare single nucleotide variant. The second patient (RP0_17) was initially suspected of suffering from CDAII, since he presented normocytic anemia and non-specific morphological erythroblast features, such as the presence of bi- and multi-nuclearity, megaloblastic changes, but no inter-nuclear bridges. First genetic testing for *SEC23B* revealed no mutations in this gene. Thus, the analysis of the patient by RedPanel_rev.0 allowed us the identification of the homozygous missense variant c.689A>C, p.His230Pro (H230P) in *C15orf41*. Similarly, BM analysis of the third patient (RP0_22) showed non-specific morphological features, as normoblasts with double or multiple nuclei. Originally, the t-NGS analysis by RedPanel_rev.0 failed to diagnose this patient because of a low coverage of the genomic region encompassing the exon 1 of *C15orf41* (data not shown). Thus, we analyzed the patient by WES, identifying the novel variant c.58C>A, p.Pro20Thr (P20T). Of note, we observed that Y94S and P20T variants are novel missense changes at an amino acid residue where different pathogenic missense changes, Y94C and P20R, have been previously described (Babbs et al., 2013; Palmblad, Sander, Bain, Klimkowska, & Bjorck, 2018). According to the guidelines of the American College of Medical Genetics and Genomics, this is a strong criterion (i.e., PS1) to identify a variant as pathogenic. Furthermore, *ex vivo* and *in vitro* analyses of C15orf41 gene and protein expression of Y94S, H230P, and P20T mutants showed that H230P variant was associated to a decrease in both gene

and protein expression, while Y94S and P20T variants did not affect C15orf41 levels. Of note, Y94S and P20T variants are in the two turn-helix-turn DNA binding domains (DBD) of the protein, together with the previously identified Y94C and P20R mutations (Babbs et al., 2013; Palmblad et al., 2018). On the contrary, H230P variant occur in the PD-(D/E) XK nuclease domain, as well as other two causative mutations, L178Q and Y238C, previously identified (Babbs et al., 2013; Palmblad et al., 2018). Therefore, we might assume that variants could have a different effect on both the protein function and the pathogenetic mechanism of the disease, depending on their localization within the sequence. Since no impaired expression of C15orf41-Y94S and -P20T was observed, we speculate that these mutations could affect the three-dimensional structure of the protein and, thus, undermine the binding to the DNA. Of course, functional studies, also extended to the other known mutations in *C15orf41*, are needed to strengthen this hypothesis.

Since CDAI mutated proteins affect mainly the erythroid lineage, we induced erythroid differentiation by hemin treatment on the developed K562 cells stably over-expressing C15orf41 WT, -Y94S, and -H230P mutants. This cellular model allowed us to demonstrate that H230P and Y94S mutant clones showed impaired erythroid differentiation, exhibiting a decreased percentage of $CD71^{+}/CD235^{+}$ cells at two days of hemin treatment. Moreover, both clones were retained in the S phase of the cell cycle during differentiation, although with a different degree. It has been already demonstrated that there is an interdependence between S-phase progression and an essential commitment step during erythroid differentiation in which, within few hours, cells become dependent on the hormone EPO, undergo activating changes in chromatin of red cell genes, and activate GATA-1, the erythroid master transcriptional regulator. Arresting S-phase progression at this point prevents the execution of this commitment step and subsequent induction of red cell genes (Pop et al., 2010). Of note, CDAN1-CDAIa cultured erythroblasts showed an increase in S-phase cells, suggesting a cell cycle arrest (Tamary et al., 1996). Nevertheless,

based on the present data, we are not able to establish if the increased number of cells in S-phase represents faster cycling cells or a block in S-phase. Moreover, we aimed to understand whether also P20T variant, together with Y94S and H230P mutations, affects cell-cycle dynamics and impairs erythroid differentiation.

To investigate the expression and subcellular localization of C15orf41, we expressed the full-length WT protein fused to a FLAG-tag. Time-course analysis evidenced an indirect correlation between gene expression and protein levels, suggesting a rapid turnover of the protein. It was recently found that C15orf41 has at least three post-translational modification sites, such as K50 (Acetylation), T114 (Phosphorylation) and K176 (Ubiquitination) (Ahmed et al., 2018). Since that ubiquitination is one of the most common signals for proteasome-mediated degradation (Hershko & Ciechanover, 1998), we hypothesized that C15orf41 is degraded via proteasome during the cell cycle. Indeed, treatment of K562 C15orf41-WT cells with MG132, a proteasome inhibitor, resulted in the accumulation of C15orf41 protein. Two isoforms of C15orf41 protein are described in the UniProt database (UniProt, 2019): (1) the full-length isoform, of 281 amino acids (32kDa), and (2) the non-canonical isoform, missing of the first 98 amino acids (21kDa). Of note, K176 ubiquitination site is conserved in the sequence of both isoforms. Indeed, our data showed the accumulation of both isoforms after proteasome inhibitor treatment. The ubiquitin-proteasome system has emerged as an important mechanism to target cell cycle proteins for degradation (Koepp, 2014). Thus, to investigate on C15orf41 modulation during cell cycle, we synchronized cells on G1/S phase with a double thymidine block on K562 cells stably overexpressing C15orf41-WT. Time course analysis on cells released from thymidine block revealed an increased expression of C15orf41 during G1/S phase, suggesting that it could have a role restricted to this phase of the cell cycle. Intriguingly, it was already demonstrated that Codanin-1, the protein

encoded by the causative gene of CDAIa, is a cell cycle-regulated protein and is preferentially expressed during S-phase of the cell cycle (Noy-Lotan et al., 2009). Thus, even though the molecular pathogenic mechanism causing CDAIa and CDAIb is still unknown, we can state that both proteins participate to the same biological process. As a matter of fact, our *ex vivo* and *in vitro* analyses demonstrated that C15orf41 and CDAN1 gene expression levels were directly correlated in patients, healthy controls, and different cell lines, suggesting a relationship between the two proteins. Indeed, very recently, their interaction was finally demonstrated (Shroff et al., 2020; Swickley et al., 2020). The analysis of cytosolic and nuclear fractions on Hek-293 cells demonstrated that C15orf41 endogenous protein exhibits mainly nuclear localization. Accordingly, nuclear localization signals and nuclear export signals were predicted in the amino acid sequence, confirming that the protein is exported from the nucleus to the cytoplasm and vice-versa. Indeed, recent studies demonstrated that Codanin-1 overexpression shifts C15orf41 localization from the nucleus to the cytoplasm (Shroff, Knebel et al. 2020, Swickley, Bloch et al. 2020). Once again, changes in C15orf41 nuclear-cytoplasmic localization could represent a mechanism, or the effect, of its regulation. According to the predicted Holliday junction resolvase function of C15orf41 and its potential role in DNA repair machinery as guardians of genome integrity and viability, we initially hypothesized that C15orf41 could localize in the nucleoli. Indeed, it was recently demonstrated that the nucleolus, long regarded as a mere ribosome producing factory, plays a crucial role in monitoring and responding to cellular stress, as well as in DNA repair mechanisms (Mayer & Grummt, 2005; Ogawa & Baserga, 2017). However, our immunolocalization data did not support this hypothesis. Overall, our data increase the current knowledge on C15orf41 variants and improve our understanding on the role of this protein in physiological and pathological conditions.

6. Conclusions

Hereditary anemias are chronic disorders with a highly variable phenotypic manifestation, which frequently leads to misdiagnosis due to unspecific or overlapping phenotypes in patients with different genetic backgrounds. In this context, in the last decade, NGS-based genetic testing represented a breakthrough. Indeed, our analyses demonstrated that RedPanel, which we designed and use in our clinical practice, not only represents a reliable diagnostic/prognostic tool to achieve a correct and definitive diagnosis but is also a valuable starting point to elucidate the underlying pathogenic mechanisms of these rare disorders. Indeed, the identification of PKD patients, initially misdiagnosed as suffering of CDAI or II, allowed us to demonstrate that stress erythropoiesis, classically ascribed to CDAs, is also a matter of fact in PKD, and need to be further investigated. Even more importantly, our diagnostic approach is worthwhile for guiding management and treatment of HAs patients. Currently, novel therapeutic options are available for PKD patients, demonstrating that providing the proper diagnosis could be crucial to improve the quality of life of patients. A paradigmatic case of this important implication of our approach was the identification of the causative genotype of a patient with early infantile epileptic encephalopathy-50 associated to CDA-like anemia, due to the c.5366G>A, p.Arg1789Gln variant in *CAD* gene. This finding allowed us to direct the treatment of the patient to oral uridine supplementation, supporting previous data on the beneficial effects of this therapeutic approach in CAD-deficient patients. Lastly, we pointed out how the multi-target sequencing diagnosis has deepened the current knowledge of pathological and physiological role of uncharacterized genes, such as *C15orf41*. Indeed, our study represents the first investigation of the expression, the localization, and the regulation of C15orf41. Moreover, functional characterization of the identified variants demonstrated an impaired erythroid

maturation and suggested the block of cell cycle dynamics as a putative pathogenic mechanism for *C15orf41*-related CDAI.

7. Acknowledgments

I would like to express my gratitude to the Prof. Achille Iolascon for the support of my PhD study and related research, for his guidance, and knowledge. His supervision helped me in all the time of research.

I am particularly grateful to my co-tutor, Dr. Roberta Russo, for helping me every time I needed. Her advices and supervision have been crucial for my working and personal formation.

I would like to acknowledge Dr. Immacolata Andolfo for her continuous support and motivation. Her contribution helped me in all the time of my PhD.

A special thanks to Barbara (Eleny), who has been for me a shoulder to cry on. Without her participation, encouragement, criticism, and input, I would not be here. Working with her meant trying to improve myself day after day.

8. List of publications

1. Russo R, **Marra R**, Rosato BE, Iolascon A, Andolfo I. Genetics and Genomics Approaches for Diagnosis and Research into Hereditary Anemias. Front Physiol. 2020 Dec 22; 11:613559. doi: 10.3389/fphys.2020.613559. PMID: 33414725; PMCID: PMC7783452.
2. Russo R, **Marra R**, Andolfo I, Manna F, De Rosa G, Rosato BE, Radhakrishnan K,Fahey M, Iolascon A. Uridine treatment normalizes the congenital dyserythropoietic anemia type II-like hematological phenotype in a patient with homozygous mutation in the CAD gene. Am J Hematol. 2020 Nov;95(11): 1423-1426.doi: 10.1002/ajh.25946. Epub 2020 Aug 19. PMID: 32720728.
3. Russo R*, **Marra R***, Andolfo I, De Rosa G, Rosato BE, Manna F, Gambale A, Raia M, Unal S, Barella S, Iolascon A. Front Physiol. 2019 May 22;10:621. doi: 10.3389/fphys.2019.00621. Characterization of two cases of Congenital Dyserythropoietic Anemia Type I shed light on the uncharacterized C15orf41 protein.
4. De Rosa G, Andolfo I, **Marra R**, Manna F, Rosato BE, Iolascon A, Russo R. RAP-011 Rescues the Disease Phenotype in a Cellular Model of Congenital Dyserythropoietic Anemia Type II by Inhibiting the SMAD2-3 Pathway. Int J Mol Sci. 2020 Aug 4;21(15):5577. doi: 10.3390/ijms21155577. PMID: 32759740; PMCID: PMC7432210.
5. Andolfo I, Rosato BE, Manna F, De Rosa G, **Marra R**, Gambale A, Girelli D,Russo R, Iolascon A. Gain-of-function mutations in PIEZO1 directly impair hepatic iron metabolism via the inhibition of the BMP/SMADs pathway. Am JHematol. 2020 Feb;95(2):188-197. doi: 10.1002/ajh.25683. Epub 2019 Dec 9. PMID:31737919.
6. Andolfo I, Rosato BE, **Marra R**, De Rosa G, Manna F, Gambale A, Iolascon A,Russo R. The BMP-SMAD pathway mediates the impaired hepatic iron metabolism associated with the ERFE-A260S variant. Am J Hematol.

2019 Nov;94(11): 1227-1235.doi: 10.1002/ajh.25613. Epub 2019 Aug 30. PMID: 31400017.

7. Benítez J., **Marra R.**, Reyes J., Calvete O. Gastric Cancer. 2019 Jun 27. doi: 10.1007/s10120-019-00982-4. A genetic origin for acid-base imbalance triggers the mitochondrial damage that explains the autoimmune response and drives to gastric neuroendocrine tumors.

8. Russo R, Andolfo I, Manna F, Gambale A, **Marra R.**, Rosato BE, Caforio P, Pinto V, Pignataro P, Radhakrishnan K, Unal S, Tomaiuolo G, Forni GL, Iolascon A. Multi-gene panel testing improves diagnosis and management of patients with hereditary anemias. Am J Hematol. 2018 May;93(5):672-682. doi: 10.1002/ajh.25058. Epub 2018 Feb 24. PMID: 29396846.

9. Andolfo I, Manna F, De Rosa G, Rosato BE, Gambale A, Tomaiuolo G, Carciati A, **Marra R.**, De Franceschi L, Iolascon A, Russo R. PIEZO1-R1864H rare variant accounts for a genetic phenotype-modifier role in dehydrated hereditary stomatocytosis. Haematologica. 2018 Mar;103(3):e94-e97. doi: 10.3324/haematol.2017.180687. Epub 2017 Nov 30. PMID: 29191841; PMCID: PMC5830381.

Submitted

1. Andolfo I, Martone S, Rosato BE, **Marra R**, Gambale A, Forni G, Pinto V, Göransson M, Vasiliki P, Gavillet M, Elalfy M, Panarelli A, Tomaiuolo G, Iolasco A and Russo R. Complex modes of inheritance in hereditary red blood cell disorders: a case series study of 155 patients. Am J Hematol.

2. Roberta Russo, Immacolata Andolfo, Alessandro Vito Lasorsa, Sueva Cantalupo, **Roberta Marra**, Giulia Frisso, Abete Pasquale, Gian Marco Cassese, Giuseppe Servillo, Gabriella Esposito, Ivan Gentile, Carmelo Piscopo, Matteo Della Monica, Giuseppe Fiorentino, Giuseppe Russo, Rino Cerino, Antonio Limone, Massimo Zollo, Achille Iolascon, Mario Capasso. The

TNFRSF13C H159Y variant is associated with severe COVID-19: a retrospective study of 500 patients from Southern Italy. Genes & Immunity.

In preparation

1. SEC23B loss-of-function impairs iron metabolism by affecting glycosylation pathway in hepatic cells. Rosato BE, Andolfo I, **Marra R**, Manna F, Della Monica S, Martone S, Russo R and Iolascon A.

2. A knock-in PIEZO1-R2456H engineered hepatic cell line for the identification of novel players in iron overload. Rosato BE*, **Marra R***, Russo R, Iolascon A. and Andolfo I.

9. References

Ahmed, M. S., Shahjaman, M., Kabir, E., & Kamruzzaman, M. (2018). Structure modeling to function prediction of Uncharacterized Human Protein C15orf41. *Bioinformation, 14*(5), 206-212. doi:10.6026/97320630014206

Aizawa, S., Harada, T., Kanbe, E., Tsuboi, I., Aisaki, K., Fujii, H., & Kanno, H. (2005). Ineffective erythropoiesis in mutant mice with deficient pyruvate kinase activity. *Exp Hematol, 33*(11), 1292-1298. doi:10.1016/j.exphem.2005.07.008

Aizawa, S., Kohdera, U., Hiramoto, M., Kawakami, Y., Aisaki, K., Kobayashi, Y.,..... Kanno, H. (2003). Ineffective erythropoiesis in the spleen of a patient with pyruvate kinase deficiency. *Am J Hematol, 74*(1), 68-72. doi:10.1002/ajh.10380

Amer, J., Goldfarb, A., & Fibach, E. (2003). Flow cytometric measurement of reactive oxygen species production by normal and thalassaemic red blood cells. *Eur J Haematol, 70*(2), 84-90. doi:10.1034/j.1600-0609.2003.00011.x

Andolfo, I., De Falco, L., Asci, R., Russo, R., Colucci, S., Gorrese, M., . . . Iolascon, A. (2010). Regulation of divalent metal transporter 1 (DMT1) non-IRE isoform by the microRNA Let-7d in erythroid cells. *Haematologica, 95*(8), 1244-1252. doi:10.3324/haematol.2009.020685

Andolfo, I., Rosato, B. E., Marra, R., De Rosa, G., Manna, F., Gambale, A., . . . Russo, R. (2019). The BMP-SMAD pathway mediates the impaired hepatic iron metabolism associated with the ERFE-A260S variant. *Am J Hematol, 94*(11), 1227-1235. doi:10.1002/ajh.25613

Andolfo, I., Russo, R., Gambale, A., & Iolascon, A. (2018). Hereditary stomatocytosis: An underdiagnosed condition. *Am J Hematol, 93*(1), 107-121. doi:10.1002/ajh.24929

Angelucci, E., Barosi, G., Camaschella, C., Cappellini, M. D., Cazzola, M., Galanello, R., . . . Tura, S. (2008). Italian Society of Hematology practice guidelines for the management of iron overload in thalassemia major and related disorders. *Haematologica, 93*(5), 741-752. doi:10.3324/haematol.12413

Ask, K., Jasencakova, Z., Menard, P., Feng, Y., Almouzni, G., & Groth, A. (2012). Codanin-1, mutated in the anaemic disease CDAI, regulates Asf1 function in S-phase histone supply. *EMBO J, 31*(8), 2013-2023. doi:10.1038/emboj.2012.55

Ayas, M., al-Jefri, A., Baothman, A., al-Mahr, M., Mustafa, M. M., Khalil, S., . . . Solh, H. (2002). Transfusion-dependent congenital dyserythropoietic anemia type I successfully treated with allogeneic stem cell transplantation. *Bone Marrow Transplant, 29*(8), 681-682. doi:10.1038/sj.bmt.1703526

Babbs, C., Roberts, N. A., Sanchez-Pulido, L., McGowan, S. J., Ahmed, M. R., Brown, J. M., . . . Buckle, V. J. (2013). Homozygous mutations in a predicted endonuclease are a novel cause of congenital dyserythropoietic anemia type I. *Haematologica, 98*(9), 1383-1387. doi:10.3324/haematol.2013.089490

Beguin, Y. (2003). Soluble transferrin receptor for the evaluation of erythropoiesis and iron status. *Clin Chim Acta, 329*(1-2), 9-22. doi:10.1016/s0009-8981(03)00005-6

Bennett, G. D., & Kay, M. M. (1981). Homeostatic removal of senescent murine erythrocytes by splenic macrophages. *Exp Hematol, 9*(3), 297-307.

Bianchi, P., Fermo, E., Vercellati, C., Boschetti, C., Barcellini, W., Iurlo, A., . . . Zanella, A. (2009). Congenital dyserythropoietic anemia type II (CDAII) is caused by mutations in the SEC23B gene. *Hum Mutat, 30*(9), 1292-1298. doi:10.1002/humu.21077

Bordeaux, J., Welsh, A., Agarwal, S., Killiam, E., Baquero, M., Hanna, J., . . . Rimm, D. (2010). Antibody validation. *Biotechniques, 48*(3), 197-209. doi:10.2144/000113382

Camaschella, C., & Nai, A. (2016). Ineffective erythropoiesis and regulation of iron status in iron loading anaemias. *Br J Haematol, 172*(4), 512-523. doi:10.1111/bjh.13820

Cappellini, M. D., Porter, J., Origa, R., Forni, G. L., Voskaridou, E., Galacteros, F., . . . Hermine, O. (2019). Sotatercept, a novel transforming growth factor beta ligand trap, improves anemia in beta-thalassemia: a phase II, open-label, dose-finding study. *Haematologica, 104*(3), 477-484. doi:10.3324/haematol.2018.198887

Casanovas, G., Vujic Spasic, M., Casu, C., Rivella, S., Strelau, J., Unsicker, K., & Muckenthaler, M. U. (2013). The murine growth differentiation factor 15 is not essential for systemic iron homeostasis in phlebotomized mice. *Haematologica, 98*(3), 444-447. doi:10.3324/haematol.2012.069807

Chasis, J. A., & Mohandas, N. (2008). Erythroblastic islands: niches for erythropoiesis. *Blood, 112*(3), 470-478. doi:10.1182/blood-2008-03-077883

Crispino, J. D., & Horwitz, M. S. (2017). GATA factor mutations in hematologic disease. *Blood, 129*(15), 2103-2110. doi:10.1182/blood-2016-09-687889

Da Costa, L., Narla, A., & Mohandas, N. (2018). An update on the pathogenesis and diagnosis of Diamond-Blackfan anemia. *F1000Res, 7*. doi:10.12688/f1000research.15542.1

Danise, P., Amendola, G., Di Concilio, R., Cillari, E., Gioia, M., Di Palma, A., . . . Maggio, A. (2009). Nucleated red blood cells and soluble transferrin receptor in thalassemia syndromes: relationship with global and ineffective erythropoiesis. *Clin Chem Lab Med, 47*(12), 1539-1542. doi:10.1515/CCLM.2009.340

de Haan, A., Eijgelsheim, M., Vogt, L., Knoers, N., & de Borst, M. H. (2019). Diagnostic Yield of Next-Generation Sequencing in Patients With Chronic Kidney Disease of Unknown Etiology. *Front Genet, 10*, 1264. doi:10.3389/fgene.2019.01264

De Rosa, G., Andolfo, I., Marra, R., Manna, F., Rosato, B. E., Iolascon, A., & Russo, R. (2020). RAP-011 Rescues the Disease Phenotype in a Cellular Model of Congenital Dyserythropoietic Anemia Type II by Inhibiting the SMAD2-3 Pathway. *Int J Mol Sci, 21*(15). doi:10.3390/ijms21155577

Delaby, C., Pilard, N., Goncalves, A. S., Beaumont, C., & Canonne-Hergaux, F. (2005). Presence of the iron exporter ferroportin at the plasma membrane of macrophages is enhanced by iron loading and down-regulated by hepcidin. *Blood, 106*(12), 3979-3984. doi:10.1182/blood-2005-06-2398

Desai, A. N., & Jere, A. (2012). Next-generation sequencing: ready for the clinics? *Clin Genet, 81*(6), 503-510. doi:10.1111/j.1399-0004.2012.01865.x

Dgany, O., Avidan, N., Delaunay, J., Krasnov, T., Shalmon, L., Shalev, H., . . . Tamary, H. (2002). Congenital dyserythropoietic anemia type I is caused by mutations in codanin-1. *Am J Hum Genet, 71*(6), 1467-1474. doi:10.1086/344781

Dussiot, M., Maciel, T. T., Fricot, A., Chartier, C., Negre, O., Veiga, J., . . . Moura, I. C. (2014). An activin receptor IIA ligand trap corrects ineffective erythropoiesis in beta-thalassemia. *Nat Med, 20*(4), 398-407. doi:10.1038/nm.3468

Forster, L., Cornwall, S., Finlayson, J., & Ghassemifar, R. (2016). Cell cycle, proliferation and apoptosis in erythroblasts cultured from patients with beta-thalassaemia major. *Br J Haematol, 175*(3), 539-542. doi:10.1111/bjh.13875

Gambale, A., Iolascon, A., Andolfo, I., & Russo, R. (2016). Diagnosis and management of congenital dyserythropoietic anemias. *Expert Rev Hematol, 9*(3), 283-296. doi:10.1586/17474086.2016.1131608

Ganz, T., Jung, G., Naeim, A., Ginzburg, Y., Pakbaz, Z., Walter, P. B., . . . Nemeth, E. (2017). Immunoassay for human serum erythroferrone. *Blood, 130*(10), 1243-1246. doi:10.1182/blood-2017-04-777987

Geiduschek, J. B., & Singer, S. J. (1979). Molecular changes in the membranes of mouse erythroid cells accompanying differentiation. *Cell, 16*(1), 149-163. doi:10.1016/0092-8674(79)90196-x

Grace, R. F., Bianchi, P., van Beers, E. J., Eber, S. W., Glader, B., Yaish, H. M., . . . Barcellini, W. (2018). Clinical spectrum of pyruvate kinase deficiency: data from the Pyruvate Kinase Deficiency Natural History Study. *Blood, 131*(20), 2183-2192. doi:10.1182/blood-2017-10-810796

Grace, R. F., Rose, C., Layton, D. M., Galacteros, F., Barcellini, W., Morton, D. H., . . . Glader, B. (2019). Safety and Efficacy of Mitapivat in Pyruvate Kinase Deficiency. *N Engl J Med, 381*(10), 933-944. doi:10.1056/NEJMoa1902678

Heimpel, H., Matuschek, A., Ahmed, M., Bader-Meunier, B., Colita, A., Delaunay, J., . . . Iolascon, A. (2010). Frequency of congenital dyserythropoietic anemias in Europe. *Eur J Haematol, 85*(1), 20-25.

Heimpel, H., Schwarz, K., Ebnother, M., Goede, J. S., Heydrich, D., Kamp, T., . . . Kohne, E. (2006). Congenital dyserythropoietic anemia type I (CDA I): molecular genetics, clinical appearance, and prognosis based on long-term observation. *Blood, 107*(1), 334-340. doi:10.1182/blood-2005-01-0421

Hershko, A., & Ciechanover, A. (1998). The ubiquitin system. *Annu Rev Biochem, 67*, 425-479. doi:10.1146/annurev.biochem.67.1.425

Higgs, D. R., Engel, J. D., & Stamatoyannopoulos, G. (2012). Thalassaemia. *Lancet, 379*(9813), 373-383. doi:10.1016/S0140-6736(11)60283-3

Hunt, L., Greenwood, D., Heimpel, H., Noel, N., Whiteway, A., & King, M. J. (2015). Toward the harmonization of result presentation for the eosin-5'-maleimide binding test in the diagnosis of hereditary spherocytosis. *Cytometry B Clin Cytom, 88*(1), 50-57. doi:10.1002/cyto.b.21187

Iolascon, A., Andolfo, I., Barcellini, W., Corcione, F., Garcon, L., De Franceschi, L., . . . Iron of the, E. H. A. (2017). Recommendations regarding

splenectomy in hereditary hemolytic anemias. *Haematologica, 102*(8), 1304-1313. doi:10.3324/haematol.2016.161166

Iolascon, A., Andolfo, I., & Russo, R. (2020). Congenital dyserythropoietic anemias. *Blood, 136*(11), 1274-1283. doi:10.1182/blood.2019000948

Iolascon, A., Esposito, M. R., & Russo, R. (2012). Clinical aspects and pathogenesis of congenital dyserythropoietic anemias: from morphology to molecular approach. *Haematologica, 97*(12), 1786-1794. doi:10.3324/haematol.2012.072207

Iolascon, A., Heimpel, H., Wahlin, A., & Tamary, H. (2013). Congenital dyserythropoietic anemias: molecular insights and diagnostic approach. *Blood, 122*(13), 2162-2166. doi:10.1182/blood-2013-05-468223

Iolascon, A., Russo, R., Esposito, M. R., Piscopo, C., Asci, R., De Falco, L., & Di Noce, F. (2011). Congenital dyserythropoietic anaemias: new acquisitions. *Blood Transfus, 9*(3), 278-280. doi:10.2450/2010.0085-10

Johnstone, R. M. (1992). The Jeanne Manery-Fisher Memorial Lecture 1991. Maturation of reticulocytes: formation of exosomes as a mechanism for shedding membrane proteins. *Biochem Cell Biol, 70*(3-4), 179-190. doi:10.1139/o92-028

Kaneko, K., Furuyama, K., Fujiwara, T., Kobayashi, R., Ishida, H., Harigae, H., & Shibahara, S. (2014). Identification of a novel erythroid-specific enhancer for the ALAS2 gene and its loss-of-function mutation which is associated with congenital sideroblastic anemia. *Haematologica, 99*(2), 252-261. doi:10.3324/haematol.2013.085449

Kato, K., Sugitani, M., Kawataki, M., Ohyama, M., Aida, N., Koga, N., . . . Itani, Y. (2001). Congenital dyserythropoietic anemia type 1 with fetal onset of severe anemia. *J Pediatr Hematol Oncol, 23*(1), 63-66. doi:10.1097/00043426-200101000-00016

Kautz, L., Jung, G., Du, X., Gabayan, V., Chapman, J., Nasoff, M., Ganz, T. (2015). Erythroferrone contributes to hepcidin suppression and iron overload in a mouse model of beta-thalassemia. *Blood, 126*(17), 2031-2037. doi:10.1182/blood-2015-07-658419

Koch, J., Mayr, J. A., Alhaddad, B., Rauscher, C., Bierau, J., Kovacs-Nagy, R., . . . Haack, T. B. (2017). CAD mutations and uridine-responsive epileptic encephalopathy. *Brain, 140*(2), 279-286. doi:10.1093/brain/aww300

Koepp, D. M. (2014). Cell cycle regulation by protein degradation. *Methods Mol Biol, 1170*, 61-73. doi:10.1007/978-1-4939-0888-2_4

Kohgo, Y., Torimoto, Y., & Kato, J. (2002). Transferrin receptor in tissue and serum: updated clinical significance of soluble receptor. *Int J Hematol, 76*(3), 213-218. doi:10.1007/BF02982790

Koury, M. J., & Bondurant, M. C. (1990). Erythropoietin retards DNA breakdown and prevents programmed death in erythroid progenitor cells. *Science, 248*(4953), 378-381. doi:10.1126/science.2326648

Koury, M. J., & Bondurant, M. C. (1992). The molecular mechanism of erythropoietin action. *Eur J Biochem, 210*(3), 649-663. doi:10.1111/j.1432-1033.1992.tb17466.x

Koury, S. T., Koury, M. J., & Bondurant, M. C. (1989). Cytoskeletal distribution and function during the maturation and enucleation of mammalian erythroblasts. *J Cell Biol, 109*(6 Pt 1), 3005-3013. doi:10.1083/jcb.109.6.3005

Kung, C., Hixon, J., Kosinski, P. A., Cianchetta, G., Histen, G., Chen, Y., . . . Dang, L. (2017). AG-348 enhances pyruvate kinase activity in red blood cells from patients with pyruvate kinase deficiency. *Blood, 130*(11), 1347-1356. doi:10.1182/blood-2016-11-753525

Lavabre-Bertrand, T., Blanc, P., Navarro, R., Saghroun, M., Vannereau, H., Braun, M., . . . Navarro, M. (1995). alpha-Interferon therapy for congenital dyserythropoiesis type I. *Br J Haematol, 89*(4), 929-932. doi:10.1111/j.1365-2141.1995.tb08442.x

Lazarova, E., Gulbis, B., Oirschot, B. V., & van Wijk, R. (2017). Next-generation osmotic gradient ektacytometry for the diagnosis of hereditary spherocytosis: interlaboratory method validation and experience. *Clin Chem Lab Med, 55*(3), 394-402. doi:10.1515/cclm-2016-0290

Lee, J. C., Gimm, J. A., Lo, A. J., Koury, M. J., Krauss, S. W., Mohandas, N., & Chasis, J. A. (2004). Mechanism of protein sorting during erythroblast enucleation: role of cytoskeletal connectivity. *Blood, 103*(5), 1912-1919. doi:10.1182/blood-2003-03-0928

Libani, I. V., Guy, E. C., Melchiori, L., Schiro, R., Ramos, P., Breda, L., . . . Rivella, S. (2008). Decreased differentiation of erythroid cells exacerbates ineffective erythropoiesis in beta-thalassemia. *Blood, 112*(3), 875-885. doi:10.1182/blood-2007-12-126938

Lin, S. M., Chen, M., Ma, E. S., Lam, Y. H., Wong, K. Y., & Tang, M. H. (2014). Intrauterine therapy in a fetus with congenital dyserythropoietic

anaemia type I. *J Obstet Gynaecol, 34*(4), 352-353. doi:10.3109/01443615.2013.859238

Lind, L., Sandstrom, H., Wahlin, A., Eriksson, M., Nilsson-Sojka, B., Sikstrom, C., & Holmgren, G. (1995). Localization of the gene for congenital dyserythropoietic anemia type III, CDAN3, to chromosome 15q21-q25. *Hum Mol Genet, 4*(1), 109-112. doi:10.1093/hmg/4.1.109

Liu, G., Niu, S., Dong, A., Cai, H., Anderson, G. J., Han, B., & Nie, G. (2012). A Chinese family carrying novel mutations in SEC23B and HFE2, the genes responsible for congenital dyserythropoietic anaemia II (CDA II) and primary iron overload, respectively. *Br J Haematol, 158*(1), 143-145. doi:10.1111/j.1365-2141.2012.09102.x

Makyio, H., Ohgi, M., Takei, T., Takahashi, S., Takatsu, H., Katoh, Y., . . . Nakayama, K. (2012). Structural basis for Arf6-MKLP1 complex formation on the Flemming body responsible for cytokinesis. *EMBO J, 31*(11), 2590-2603. doi:10.1038/emboj.2012.89

Manco, L., Ribeiro, M. L., Maximo, V., Almeida, H., Costa, A., Freitas, O., . . . Tamagnini, G. (2000). A new PKLR gene mutation in the R-type promoter region affects the gene transcription causing pyruvate kinase deficiency. *Br J Haematol, 110*(4), 993-997. doi:10.1046/j.1365-2141.2000.02283.x

Mathias, L. A., Fisher, T. C., Zeng, L., Meiselman, H. J., Weinberg, K. I., Hiti, A. L., & Malik, P. (2000). Ineffective erythropoiesis in beta-thalassemia major is due to apoptosis at the polychromatophilic normoblast stage. *Exp Hematol, 28*(12), 1343-1353. doi:10.1016/s0301-472x(00)00555-5

Mayer, C., & Grummt, I. (2005). Cellular stress and nucleolar function. *Cell Cycle, 4*(8), 1036-1038. doi:10.4161/cc.4.8.1925

McGrath, K. E., Bushnell, T. P., & Palis, J. (2008). Multispectral imaging of hematopoietic cells: where flow meets morphology. *J Immunol Methods, 336*(2), 91-97. doi:10.1016/j.jim.2008.04.012

Merryweather-Clarke, A. T., Atzberger, A., Soneji, S., Gray, N., Clark, K., Waugh, C., . . . Robson, K. J. (2011). Global gene expression analysis of human erythroid progenitors. *Blood, 117*(13), e96-108. doi:10.1182/blood-2010-07-290825

Miano, M., Eikema, D. J., Aljurf, M., Van't Veer, P. J., Ozturk, G., Wolfl, M., . . . Dufour, C. (2019). Stem cell transplantation for congenital dyserythropoietic anemia: an analysis from the European Society for Blood and Marrow Transplantation. *Haematologica, 104*(8), e335-e339. doi:10.3324/haematol.2018.206623

Murphy, Z. C., Getman, M. R., Myers, J. A., Burgos Villar, K. N., Leshen, E., Kurita, R., . . . Steiner, L. A. (2020). Codanin-1 mutations engineered in human erythroid cells demonstrate role of CDAN1 in terminal erythroid maturation. *Exp Hematol, 91*, 32-38 e36. doi:10.1016/j.exphem.2020.09.201

Nair, V., Das, S., & Sharma, A. (2009). Hematopoietic stem cell transplantation in children with genetic defects. *Indian Pediatr, 46*(3), 241-243. Nandakumar, S. K., Ulirsch, J. C., & Sankaran, V. G. (2016). Advances in understanding erythropoiesis: evolving perspectives. *Br J Haematol, 173*(2), 206-218. doi:10.1111/bjh.13938

Nathan, D. G., Clarke, B. J., Hillman, D. G., Alter, B. P., & Housman, D. E. (1978). Erythroid precursors in congenital hypoplastic (Diamond-Blackfan) anemia. *J Clin Invest, 61*(2), 489-498. doi:10.1172/JCI108960

Nemeth, E., Tuttle, M. S., Powelson, J., Vaughn, M. B., Donovan, A., Ward, D. M., . . . Kaplan, J. (2004). Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science, 306*(5704), 2090-2093. doi:10.1126/science.1104742

Ng, B. G., Wolfe, L. A., Ichikawa, M., Markello, T., He, M., Tifft, C. J., . . . Freeze, H. H. (2015). Biallelic mutations in CAD, impair de novo pyrimidine biosynthesis and decrease glycosylation precursors. *Hum Mol Genet, 24*(11), 3050-3057. doi:10.1093/hmg/ddv057

Noy-Lotan, S., Dgany, O., Lahmi, R., Marcoux, N., Krasnov, T., Yissachar, N., . . . Tamary, H. (2009). Codanin-1, the protein encoded by the gene mutated in congenital dyserythropoietic anemia type I (CDAN1), is cell cycle-regulated. *Haematologica, 94*(5), 629-637. doi:10.3324/haematol.2008.003327

Ogawa, L. M., & Baserga, S. J. (2017). Crosstalk between the nucleolus and the DNA damage response. *Mol Biosyst, 13*(3), 443-455. doi:10.1039/c6mb00740f

Oikonomidou, P. R., & Rivella, S. (2018). What can we learn from ineffective erythropoiesis in thalassemia? *Blood Rev, 32*(2), 130-143. doi:10.1016/j.blre.2017.10.001

Palmblad, J., Sander, B., Bain, B., Klimkowska, M., & Bjorck, E. (2018). Congenital dyserythropoietic anemia type 1: a case with novel compound heterozygous mutations in the C15orf41 gene. *Am J Hematol*. doi:10.1002/ajh.25157

Parez, N., Dommergues, M., Zupan, V., Chambost, H., Fieschi, J. B., Delaunay, J., . . . Tchernia, G. (2000). Severe congenital dyserythropoietic anaemia type I: prenatal management, transfusion support and alpha-interferon therapy. *Br J Haematol, 110*(2), 420-423. doi:10.1046/j.1365-2141.2000.02168.x

Pellegrin, S., Haydn-Smith, K. L., Hampton-O'Neil, L. A., Hawley, B. R., Heesom, K. J., Fermo, E., . . . Toye, A. M. (2019). Transduction with BBF2H7/CREB3L2 upregulates SEC23A protein in erythroblasts and partially corrects the hypo-glycosylation phenotype associated with CDAII. *Br J Haematol, 184*(5), 876-881. doi:10.1111/bjh.15189

Pereira, J., Bento, C., Manco, L., Gonzalez, A., Vagace, J., & Ribeiro, M. L. (2016). Congenital dyserythropoietic anemia associated to a GATA1 mutation aggravated by pyruvate kinase deficiency. *Ann Hematol, 95*(9), 1551-1553. doi:10.1007/s00277-016-2720-0

Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S., & Goldstein, D. B. (2013). Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet, 9*(8), e1003709. doi:10.1371/journal.pgen.1003709

Pop, R., Shearstone, J. R., Shen, Q., Liu, Y., Hallstrom, K., Koulnis, M., . . . Socolovsky, M. (2010). A key commitment step in erythropoiesis is synchronized with the cell cycle clock through mutual inhibition between PU.1 and S-phase progression. *PLoS Biol, 8*(9). doi:10.1371/journal.pbio.1000484

Radosinska, J., & Vrbjar, N. (2016). The role of red blood cell deformability and Na,K-ATPase function in selected risk factors of cardiovascular diseases in humans: focus on hypertension, diabetes mellitus and hypercholesterolemia. *Physiol Res, 65 Suppl 1*, S43-54. doi:10.33549/physiolres.933402

Rao, A. P., Gopalakrishna, D. B., Bing, X., & Ferguson, P. J. (2016). Phenotypic Variability in Majeed Syndrome. *J Rheumatol, 43*(6), 1258-1259. doi:10.3899/jrheum.151193

Renella, R., Roberts, N. A., Brown, J. M., De Gobbi, M., Bird, L. E., Hassanali, T., . . . Wood, W. G. (2011). Codanin-1 mutations in congenital dyserythropoietic anemia type 1 affect HP1{alpha} localization in erythroblasts. *Blood, 117*(25), 6928-6938. doi:10.1182/blood-2010-09-308478

Ribeil, J. A., Arlet, J. B., Dussiot, M., Moura, I. C., Courtois, G., & Hermine, O. (2013). Ineffective erythropoiesis in beta -thalassemia. *ScientificWorldJournal, 2013*, 394295. doi:10.1155/2013/394295

Ricchi, P., Ammirabile, M., Costantini, S., Spasiano, A., Di Matola, T., Verna, R., . . . Pagano, L. (2016). Soluble form of transferrin receptor as a biomarker of overall morbidity in patients with non-transfusion-dependent thalassaemia: a cross-sectional study. *Blood Transfus, 14*(6), 538-540. doi:10.2450/2016.0279-15

Richard, C., & Verdier, F. (2020). Transferrin Receptors in Erythropoiesis. *Int J Mol Sci, 21*(24). doi:10.3390/ijms21249713

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (2015). 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, 17*(5), 405-424. doi:10.1038/gim.2015.30

Roy, N. B., Wilson, E. A., Henderson, S., Wray, K., Babbs, C., Okoli, S., . . . Schuh, A. (2016). A novel 33-Gene targeted resequencing panel provides accurate, clinical-grade diagnosis and improves patient management for rare inherited anaemias. *Br J Haematol, 175*(2), 318-330. doi:10.1111/bjh.14221

Roy, N. B. A., & Babbs, C. (2019). The pathogenesis, diagnosis and management of congenital dyserythropoietic anaemia type I. *Br J Haematol, 185*(3), 436-449. doi:10.1111/bjh.15817

Russo, R., Andolfo, I., Gambale, A., De Rosa, G., Manna, F., Arillo, A., . . . Iolascon, A. (2017). GATA1 erythroid-specific regulation of SEC23B expression and its implication in the pathogenesis of congenital dyserythropoietic anemia type II. *Haematologica, 102*(9), e371-e374. doi:10.3324/haematol.2016.162966

Russo, R., Andolfo, I., Manna, F., De Rosa, G., De Falco, L., Gambale, A., . . . Iolascon, A. (2016). Increased levels of ERFE-encoding FAM132B in patients with congenital dyserythropoietic anemia type II. *Blood, 128*(14), 1899-1902. doi:10.1182/blood-2016-06-724328

Russo, R., Andolfo, I., Manna, F., Gambale, A., Marra, R., Rosato, B. E., . . . Iolascon, A. (2018). Multi-gene panel testing improves diagnosis and management of patients with hereditary anemias. *Am J Hematol, 93*(5), 672-682. doi:10.1002/ajh.25058

Russo, R., Langella, C., Esposito, M. R., Gambale, A., Vitiello, F., Vallefuoco, F., . . . Iolascon, A. (2013). Hypomorphic mutations of SEC23B gene account for mild phenotypes of congenital dyserythropoietic anemia type II. *Blood Cells Mol Dis, 51*(1), 17-21. doi:10.1016/j.bcmd.2013.02.003

Russo, R., Marra, R., Rosato, B. E., Iolascon, A., & Andolfo, I. (2020). Genetics and Genomics Approaches for Diagnosis and Research Into Hereditary Anemias. *Front Physiol, 11*, 613559. doi:10.3389/fphys.2020.613559

Samkari, A., Borzutzky, A., Fermo, E., Treaba, D. O., Dedeoglu, F., & Altura, R. A. (2010). A novel missense mutation in MVK associated with MK deficiency and dyserythropoietic anemia. *Pediatrics, 125*(4), e964-968. doi:10.1542/peds.2009-1774

Schwarz, K., Iolascon, A., Verissimo, F., Trede, N. S., Horsley, W., Chen, W., . . . Heimpel, H. (2009). Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II. *Nat Genet, 41*(8), 936-940. doi:10.1038/ng.405

Secrest, M. H., Storm, M., Carrington, C., Casso, D., Gilroy, K., Pladson, L., & Boscoe, A. N. (2020). Prevalence of pyruvate kinase deficiency: A systematic literature review. *Eur J Haematol, 105*(2), 173-184. doi:10.1111/ejh.13424

Shalev, H., Al-Athamen, K., Levi, I., Levitas, A., & Tamary, H. (2017). Morbidity and mortality of adult patients with congenital dyserythropoietic anemia type I. *Eur J Haematol, 98*(1), 13-18. doi:10.1111/ejh.12778

Shefer Averbuch, N., Steinberg-Shemer, O., Dgany, O., Krasnov, T., Noy-Lotan, S., Yacobovich, J., . . . Tamary, H. (2018). Targeted next generation sequencing for the diagnosis of patients with rare congenital anemias. *Eur J Haematol, 101*(3), 297-304. doi:10.1111/ejh.13097

Shimizu, R., & Yamamoto, M. (2016). GATA-related hematologic disorders. *Exp Hematol, 44*(8), 696-705. doi:10.1016/j.exphem.2016.05.010

Shroff, M., Knebel, A., Toth, R., & Rouse, J. (2020). A complex comprising C15ORF41 and Codanin-1: the products of two genes mutated in congenital dyserythropoietic anaemia type I (CDA-I). *Biochem J, 477*(10), 1893-1905. doi:10.1042/BCJ20190944

Siatecka, M., & Bieker, J. J. (2011). The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood, 118*(8), 2044-2054. doi:10.1182/blood-2011-03-331371

Stewart, G. W., Amess, J. A., Eber, S. W., Kingswood, C., Lane, P. A., Smith, B. D., & Mentzer, W. C. (1996). Thrombo-embolic disease after splenectomy for hereditary stomatocytosis. *Br J Haematol, 93*(2), 303-310. doi:10.1046/j.1365-2141.1996.4881033.x

Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., . . . Hogenesch, J. B. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A, 101*(16), 6062-6067. doi:10.1073/pnas.0400782101

Swickley, G., Bloch, Y., Malka, L., Meiri, A., Noy-Lotan, S., Yanai, A., . . . Motro, B. (2020). Characterization of the interactions between Codanin-1 and C15Orf41, two proteins implicated in congenital dyserythropoietic anemia type I disease. *BMC Mol Cell Biol, 21*(1), 18. doi:10.1186/s12860-020-00258-1

Takegawa, S., Fujii, H., & Miwa, S. (1983). Change of pyruvate kinase isozymes from M2- to L-type during development of the red cell. *Br J Haematol, 54*(3), 467-474. doi:10.1111/j.1365-2141.1983.tb02121.x

Tamary, H., Shalev, H., Luria, D., Shaft, D., Zoldan, M., Shalmon, L., . . . Zaizov, R. (1996). Clinical features and studies of erythropoiesis in Israeli Bedouins with congenital dyserythropoietic anemia type I. *Blood, 87*(5), 1763-1770.

Tamary, H., Shalev, H., Perez-Avraham, G., Zoldan, M., Levi, I., Swinkels, D. W., . . . Miller, J. L. (2008). Elevated growth differentiation factor 15 expression in patients with congenital dyserythropoietic anemia type I. *Blood, 112*(13), 5241-5244. doi:10.1182/blood-2008-06-165738

Tanno, T., Bhanu, N. V., Oneal, P. A., Goh, S. H., Staker, P., Lee, Y. T., . . . Miller, J. L. (2007). High levels of GDF15 in thalassemia suppress expression of the iron regulatory protein hepcidin. *Nat Med, 13*(9), 1096-1101. doi:10.1038/nm1629

Unal, S., Russo, R., Gumruk, F., Kuskonmaz, B., Cetin, M., Sayli, T., . . . Uckan Cetinkaya, D. (2014). Successful hematopoietic stem cell transplantation in a patient with congenital dyserythropoietic anemia type II. *Pediatr Transplant, 18*(4), E130-133. doi:10.1111/petr.12254

UniProt, C. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res, 47*(D1), D506-D515. doi:10.1093/nar/gky1049

Voskou, S., Aslan, M., Fanis, P., Phylactides, M., & Kleanthous, M. (2015). Oxidative stress in beta-thalassaemia and sickle cell disease. *Redox Biol, 6*, 226-239. doi:10.1016/j.redox.2015.07.018

Weiss, G., & Goodnough, L. T. (2005). Anemia of chronic disease. *N Engl J Med, 352*(10), 1011-1023. doi:10.1056/NEJMra041809

Wickramasinghe, S. N. (1997). Dyserythropoiesis and congenital dyserythropoietic anaemias. *Br J Haematol, 98*(4), 785-797. doi:10.1046/j.1365-2141.1997.2513065.x

Xue, Y., Ankala, A., Wilcox, W. R., & Hegde, M. R. (2015). Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med, 17*(6), 444-451. doi:10.1038/gim.2014.122

Yang, H., Merica, E., Chen, Y., Cohen, M., Goldwater, R., Kosinski, P. A., . . . Barbier, A. J. (2019). Phase 1 Single- and Multiple-Ascending-Dose Randomized Studies of the Safety, Pharmacokinetics, and Pharmacodynamics of AG-348, a First-in-Class Allosteric Activator of Pyruvate Kinase R, in Healthy Volunteers. *Clin Pharmacol Drug Dev, 8*(2), 246-259. doi:10.1002/cpdd.604

Yuan, J., Angelucci, E., Lucarelli, G., Aljurf, M., Snyder, L. M., Kiefer, C. R., . . . Schrier, S. L. (1993). Accelerated programmed cell death (apoptosis) in erythroid precursors of patients with severe beta-thalassemia (Cooley's anemia). *Blood, 82*(2), 374-377.