Received: 31 August 2020 | Revised: 10 December 2020 | Accepted: 15 January 2021

DOI: 10.1002/jcla.23715

RESEARCH ARTICLE

WILEY

# Genetic analysis of 39 erythrocytosis and hereditary hemochromatosis-associated genes in the Slovenian family with idiopathic erythrocytosis

Aleša Kristan[1] | Jernej Gašperšič[1] | Tadeja Režen[2] | Tanja Kunej[3] | Rok Količ[4] | Andrej Vuga[4] | Martina Fink[5] | Špela Žula[5] | Saša Anžej Doma[5] | Irena Preložnik Zupan[5,6] | Tadej Pajič[5,7] | Helena Podgornik[5,8] | Nataša Debeljak[1]

[1]Medical Centre for Molecular Biology, Faculty of Medicine, Institute of Biochemistry and Molecular Genetics, University of Ljubljana, Ljubljana, Slovenia

[2]Centre for Functional Genomics and Bio-Chips, Faculty of Medicine, Institute of Biochemistry and Molecular Genetics, University of Ljubljana, Ljubljana, Slovenia

[3]Department of Animal Science, Biotechnical Faculty, University of Ljubljana, Ljubljana, Slovenia

[4]Kemomed Research and Development, Kemomed Ltd, Kranj, Slovenia

[5]Clinical Department of Haematology, University Medical Centre Ljubljana, Ljubljana, Slovenia

[6]Department of Internal Medicine, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

[7]Clinical Institute of Genomic Medicine, University Medical Centre Ljubljana, Ljubljana, Slovenia

[8]Chair of Clinical Biochemistry, Faculty of Pharmacy, University of Ljubljana, Ljubljana, Slovenia

**Correspondence**
Nataša Debeljak, Medical Centre for Molecular Biology, Institute of Biochemistry and Molecular Genetics, Faculty of Medicine, University of Ljubljana, Vrazov trg 2, SI-1000 Ljubljana, Slovenia.
Email: natasa.debeljak@mf.uni-lj.si

**Funding information**
Javna Agencija za Raziskovalno Dejavnost RS, Grant/Award Number: L3-9279 and P1-0390; University Medical Centre Ljubljana, Grant/Award Number: 20170073

## Abstract

**Background:** Erythrocytosis is a condition with an excessive number of erythrocytes, accompanied by an elevated haemoglobin and/or haematocrit value. Congenital erythrocytosis has a diverse genetic background with several genes involved in erythropoiesis. In clinical practice, nine genes are usually examined, but in approximately 70% of patients, no causative mutation can be identified. In this study, we screened 39 genes, aiming to identify potential disease-driving variants in the family with erythrocytosis of unknown cause.

**Patients and Methods:** Two affected family members with elevated haemoglobin and/or haematocrit and negative for acquired causes and one healthy relative from the same family were selected for molecular-genetic analysis of 24 erythrocytosis and 15 hereditary haemochromatosis-associated genes with targeted NGS. The identified variants were further analysed for pathogenicity using various bioinformatic tools and review of the literature.

**Results:** Of the 12 identified variants, two heterozygous variants, the missense variant c.471G>C (NM_022051.2) (p.(Gln157His)) in the *EGLN1* gene and the intron variant c.2572-13A>G (NM_004972.3) in the *JAK2* gene, were classified as low-frequency variants in European population. None of the two variants were present in a healthy

family member. Variant c.2572-13A>G has potential impact on splicing by one prediction tool.

**Conclusion:** For the first time, we included 39 genes in the erythrocytosis clinical panel and identified two potential disease-driving variants in the Slovene family studied. Based on the reported functional in vitro studies combined with our bioinformatics analysis, we suggest further functional analysis of variant in the *JAK2* gene and evaluation of a cumulative effect of both variants.

**KEYWORDS**
DNA, erythrocytosis, genetic variation, haemochromatosis, sequence analysis

## 1 | INTRODUCTION

Erythrocytosis is a heterogeneous group of disorders that occur when there is an increase in the red blood cell (RBC) mass by more than 125% of the predicted value for specific body height and weight, which defines increased haemoglobin levels and/or haematocrit.[1,2] Erythrocytosis is classified according to its cause, as primary or secondary, and as congenital or acquired. Primary erythrocytosis occurs when there is an intrinsic defect in the erythroid progenitor cells that is associated with expanded proliferation of RBCs, with erythropoietin (EPO) levels remaining below normal. On the contrary, secondary erythrocytosis occurs when the defect is extrinsic to the erythroid compartment and is instead associated with normal or inappropriately high EPO levels.[3] Most cases of erythrocytosis are acquired, which defines erythrocytosis secondary to various cardiac, pulmonary or renal diseases, or to external hypoxia, or where erythrocytosis is the consequence of somatic gene variants. Polycythaemia vera is the most common primary acquired erythrocytosis, which is characterized by the somatic variant p.Val617Phe or variants in exon 12 of the Janus kinase 2 (*JAK2*) gene.[4] The JAK2 protein is a tyrosine kinase that has important roles in the signal transduction for proliferation and differentiation of myeloid cells. JAK2 binds to the intracellular domains of the cytokine receptors that lack catalytic activity, including the EPO receptor (EPOR). When ligands such as the cytokine EPO bind to their receptors, they induce conformational changes and phosphorylation events that activate JAK2, with the consequent signal transduction via the downstream JAK2/STAT5, MAPK/ERK and PI3K/AKT pathways. In haematopoiesis, JAK2 is a critical mediator of an effective erythropoiesis.[5]

Congenital erythrocytosis, which is also known as familial erythrocytosis, is a rare clinical condition that is present from birth due to a germline defect in one of the various genes involved in the regulation of oxygen homeostasis.[4,6] For males, the haemoglobin levels are usually >185 g/L (with haematocrit >0.52), and a little lower for females, at >165 g/L (for haematocrit >0.48).[1,2] Eight types of congenital erythrocytosis have been classified according to Online Mendelian Inheritance in Man (ECYT1-8).[7] Primary congenital erythrocytosis is defined as type 1 (ECYT1), and this is caused by variants in the *EPOR* gene. Secondary congenital erythrocytosis is classified into remaining seven types (ECYT2-8), and these are caused by variants in genes involved in the oxygen-sensing pathway (*VHL*, *EGLN1*, *EPAS1*, *EPO*) or variants that affect the haemoglobin oxygen affinity (*HBB*, *HBA1*, *HBA2*, *BPGM*) (Table 1).[6-9] Oxygen homeostasis is

**TABLE 1** Classification of congenital erythrocytosis.

| Type | OMIM # | Inheritance[a] | Gene | Gene location | Protein (synonym) |
|---|---|---|---|---|---|
| ECYT1 | 1333100 | AD | *EPOR* | 19p13.2 | Erythropoietin receptor |
| ECYT2 | 263400 | AR | *VHL* | 3p25.3 | von Hippel-Lindau tumor suppressor |
| ECYT3 | 609820 | AD | *EGLN1* | 1q42.2 | Egl−9 family hypoxia-inducible factor 1 (HIF-prolyl hydroxylase 2, PHD2) |
| ECYT4 | 611783 | AD | *EPAS1* | 2p21 | Endothelial PAS domain protein 1 (hypoxia-inducible factor 2 subunit alpha, HIF2α) |
| ECYT5 | 617907 | AD | *EPO* | 7q22.1 | Erythropoietin |
| ECYT6 | 617980 | AD | *HBB* | 11p15.4 | Haemoglobin subunit beta |
| ECYT7 | 617981 | AD | *HBA1*, *HBA2* | 16p13.3 | Haemoglobin subunit alpha |
| ECYT8 | 222800 | AR | *BPGM* | 7q33 | Bisphosphoglycerate mutase |

Abbreviations: AD, autosomal-dominant; AR, autosomal-recessive

[a]Inheritance of the specific CE type was attained from OMIM database.[7]

regulated through a complex pathway that is mediated by the essential transcription factors known as hypoxia-inducible factors (HIFs). HIFs can up-regulate a number of target genes involved in maintenance of sufficient tissue oxygenation, including the hormone EPO, which regulates proliferation and differentiation of erythroid progenitors. Transcription complex HIF is a dimer, regulated through oxygen-dependent regulation of alpha subunit. In humans, three isoforms of alpha subunit are known, with HIF2α [officially termed endothelial PAS domain protein 1, EPAS1] as major isoform involved in erythropoiesis. At normal oxygen concentrations (ie, normoxia), the HIF-α subunit is hydroxylated by HIF-prolyl hydroxylase 2 (PHD2) [officially termed EGLN1] at prolyl residues in its oxygen-dependent degradation domain. Post-translational hydroxylation regulates the oxygen-dependent stability of EPAS1, as it allows binding of the von Hippel-Lindau tumor suppressor (VHL) , which results in ubiquitination, followed by degradation by proteasomes. Under hypoxic conditions (eg 1% $O_2$), oxygen availability is limited, and thus, this hydroxylation is diminished, which prevent interactions with the VHL protein. This results in stabilization of HIF2α and formation of the stable HIF dimer complex that is responsible for the subsequent transcription activation of the targeted genes.[10]

The majority of the variants associated with ECYT1-8 have been collected in the Global Variome shared Leiden Open Variation Database (LOVD), which is supported by the LOVD 3.x software.[11] Along with the nine genes responsible for ECYT1-8, other key genes are involved in the erythropoiesis pathway and are hence plausible target genes for the development of erythrocytosis.[12] Because of heterogenic genetic background, the cause for erythrocytosis remains unknown in 70% of patients with indication of congenital erythrocytosis that have been screened for known pathogenic variants in genes causative for ECYT1-8. Those patients are therefore diagnosed with erythrocytosis of unknown cause, so-called idiopathic erythrocytosis. No proper diagnosis and prognosis can be made by clinicians, and further genetic studies are needed for patients with idiopathic erythrocytosis.[12,13]

Enhanced erythropoiesis can be also induced by iron overload, as iron is necessary for the synthesis of haemoglobin.[14] An excess iron in the blood could be a sign of hereditary haemochromatosis, a well-known disorder caused by defects in genes that are involved in iron metabolism, transport and balance. The most common hereditary haemochromatosis is type 1, associated with homozygous or compound heterozygous variants in homeostatic iron regulator (*HFE*) gene.[15,16] Some authors have observed high frequency of heterozygous *HFE* variants among patients with idiopathic erythrocytosis, indicating the involvement of haemochromatosis genes in the development of erythrocytosis.[15]

The aim of the present study was to investigate the hereditary gene variants of multiple genes involved in erythropoiesis, in a family with indication of congenital erythrocytosis. An expanded selection of 24 erythrocytosis and 15 hereditary haemochromatosis-associated genes was included in the targeted molecular-genetic analysis with next-generation sequencing (NGS). Here, we report on these NGS screening results where we describe variants with plausible causative effects on increased erythropoiesis.

## 2 | MATERIALS AND METHODS

### 2.1 | Patients

The index patient (Figure 1, individual I:2) was selected from the individuals followed at the University Medical Centre Ljubljana (UMC) over an 8-year period according to the diagnostic algorithm for

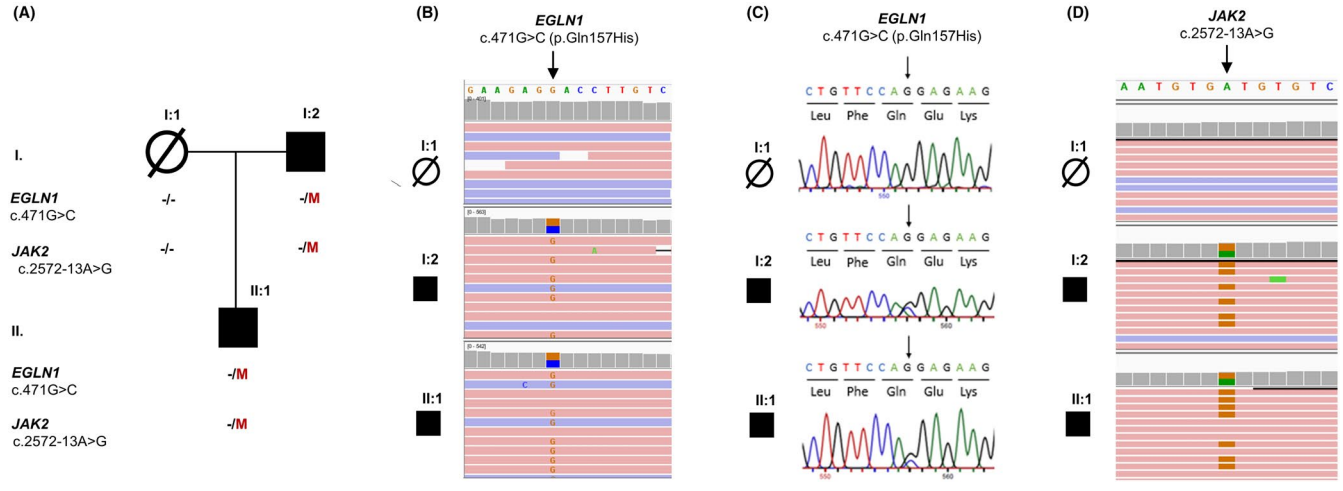


**FIGURE 1** Identification of a heterozygous nucleotide variants in the *EGLN1* and *JAK2* genes (A) Family pedigree and segregation of the variants identified within the family. (B) Next-generation sequencing (NGS) and (C) Sanger sequencing, showing heterozygous *EGLN1* variant in the two affected patients , but not in the unaffected family member. (D) Results of NGS showing intron variant in the *JAK2* gene in both of the affected family members, but not in the unaffected family member. Squares, circles, represent males, females, respectively. Filled symbols, patients with clinical diagnosis of erythrocytosis; open symbols, unaffected family members; slashed symbol, subject is deceased. Carriers of heterozygous *EGLN1* and *JAK2* variants are indicated as -/M. Arrow indicates mutation point

erythrocytosis.[17] The inclusion criteria were as follows: (a) haemoglobin and/or haematocrit above reference values at least twice over 2 months; (b) absence of variants *JAK2* p.Val617Phe and *JAK2* exon 12; (c) absence of any defined cause of secondary acquired erythrocytosis; and (d) absence of variants in genes for the thrombopoietin receptor (*MPL*), calreticulin (*CALR*) and receptor tyrosine kinase (*KIT*).[17] The index patient was a 68-year-old male who was referred to the Clinical Department of Haematology at UMC Ljubljana in 2013, due to his high RBC count of 6.36 x $10^{12}$ cells/L, high haemoglobin of 221 g/L and high haematocrit of 0.650. His EPO level was in the normal range (11.4 IU/L). The patient also had elevated blood pressure and heart rate. The patient was prescribed for therapeutic phlebotomy, and at the last follow-up in 2019, he had a slightly increased haemoglobin of 171 g/L. Upon medical history revision, he reported that his son also had clinical signs of erythrocytosis.

The son (Figure 1, individual II:1) of the index patient was 41 years old, and at his first appointment in 2015, he presented with a RBC count of 6.29 × $10^{12}$ cells/L, high haemoglobin of 192 g/L and high haematocrit of 0.552, while his EPO level was normal (8.3 IU/L). The son was a regular blood donor, and so he was not referred for a phlebotomy. These two patients were tested for polycythaemia vera and hereditary haemochromatosis and were negative for variants p.Val617Phe and exon 12 in the *JAK2* gene, and also for variants p.Cys282Tyr, p.His63Asp and p.Ser65Asp in the gene for homostatic iron regulator (HFE). Neither of these two patients showed any pulmonary, cardiac or renal abnormalities.

The wife and mother of these elder and younger patients, respectively (Figure 1, individual I:1), is deceased, but had had normal RBC counts, haemoglobin levels and haematocrit, and no signs of erythrocytosis.

All three affected and healthy family members were, together with a reference DNA control NA12878, included in the NGS genetic testing for congenital erythrocytosis. The study was approved by the Slovenian Ethical Committee, No. KME 115/07/15.

## 2.2 | Genetic analysis

Patient's peripheral blood was collected for genetic analysis, together with written informed consent signed by all patients. Granulocytes were isolated from collected peripheral blood, and the genomic DNA was extracted from 1 to 2 × $10^7$ cells using QIAamp DNA mini kits (Qiagen). The patients underwent targeted NGS, with a custom gene panel that covered the target regions of 39 selected genes: 21 genes previously associated with congenital erythrocytosis and polycythaemia vera,[12] three additional erythrocytosis-associated genes (*PKLR*, *TET2* and *GATA*); and 15 genes previously associated with hereditary haemochromatosis.[18] All of the selected genes were targeted for exon regions, while *EPO*, *VHL* and some of the other genes had the first intron, promoter and enhancer regions also included[19,20] (Table 2).

Libraries were prepared using Nextera DNA library preparation kits (Illumina), with enrichment performed by probe hybridization approach using custom gene panel (Integrated DNA Technologies) followed by sequencing (MiniSeq, Illumina). The disease-risk variants identified were validated by Sanger sequencing (GATC Biotech). Sanger sequencing and prior PCR amplification were performed with custom-designed primers (Integrated DNA Technologies; available upon request).

**TABLE 2** List of sequenced genes and regions

| Association | Exon | Intron 1 | Promotor and enhancer |
|---|---|---|---|
| Erythrocytosis-associated genes[a] | *BHLHE41*, *BPGM*, *EGLN1*, *EGLN2*, *EGLN3*, *EPAS1*, *EPO*, *EPOR*, *GFI1B*, *HBA1*, *HBA2*, *HBB*, *HIF1A*, *HIF1AN*, *HIF3A*, *JAK2*, *KDM6A*, *OS9*, *PKLR*, *SH2B3*, *VHL*, *ZNF197*, *GATA1* and *TET2* | *VHL*, *EPO*, *EPOR*, *HBB*, *HBA1*, *HBA2* | *EPO* |
| Hereditary haemochromatosis-associated genes[b] | *HFE*, *HJV*, *HAMP*, *TFR2*, *SLC40A1*, *FTH1*, *TF*, *B2 M*, *CP*, *FTL*, *CDAN1*, *SEC23B*, *SLC25A38*, *STEAP3* and *ALAS2* | - | - |

*BHLHE41* gene for basic helix-loop-helix e41, *BPGM* gene for bisphosphoglycerate mutase, *EGLN1* gene for egl 1, *EGLN2* gene for egl, *EGLN3* gene for egl 3, *EPAS1* gene for endothelial PAS domain protein 1, *EPO* gene for erythropoietin, *EPOR* gene for erythropoietin receptor, *GFI1B* gene for growth factor-independent 1B transcriptional repressor, *HBA1* gene for haemoglobin subunit alpha 1, *HBA2* gene for haemoglobin subunit alpha 2, *HBB* gene for haemoglobin subunit beta, *HIF1A* gene for hypoxia-inducible factor 1 subunit alpha, *HIF1AN* gene for hypoxia-inducible factor 1 subunit alpha inhibitor, *HIF3A* gene for hypoxia-inducible factor 3 subunit alpha, *JAK2* gene for Janus kinase 2, *KDM6A* gene for lysine demethylase 6A, *OS9* gene for OS9 *PKLR* gene for pyruvate kinase L/R, *SH2B3* gene for SH2B adaptor protein 3, *VHL* gene for von Hippel-Lindau tumor suppressor, *ZNF197* gene for zinc finger protein 197, *GATA1* gene for GATA binding protein 1, *TET2* gene for tet methylcytosine dioxygenase 2, *HFE* gene for homeostatic iron regulator, *HJV* gene for hemojuvelin BMP co-receptor, *HAMP* gene for hepcidin antimicrobial peptide, *TFR2* gene for transferrin receptor 2, *SLC40A1* gene for solute carrier family 40 member 1, *FTH1* gene for ferritin heavy chain 1, *TF* gene for transferrin, *B2M* gene for beta-2-microglobulin, *CP* gene for ceruloplasmin, *FTL* gene for ferritin light chain, *CDAN1* gene for codanin 1, *SEC23B* SEC23 homolog B, *SLC25A38* gene for solute carrier family 25 member 38, *STEAP3* STEAP3 metalloreductase, *ALAS2* gene for 5'-aminolevulinate synthase 2.

[a]Adopted from Camps et al. (2016).[12]

[b]Adopted form Lanktree et al. (2017).[18]

## 2.3 | Bioinformatics analysis

The sequencing analysis was performed with built-in bioinformatics tools (Illumina) and variant annotation with an online tool (Variant Interpreter Illumina). The sequences were aligned to reference genome hg19 (GRCh37). To remove variants with low sequencing quality, the following filters were used within the variant caller: genotype quality (ie GQX value) <30 or not present; quality by depth <2; root mean square mapping quality <20; strand bias > −10; and read depth <1. The variants were first selected based on the relationship between genotypes and phenotypes; this selection was for variants identified as heterozygous in the affected family members and not in the healthy family members, or identified as homozygous in the affected family members and as heterozygous in the healthy family members. For the final selection, the variants with minor allele frequencies (MAFs) <0.05 in the European population were filtered, as MAF <0.05 is distinct for low-frequency variants.[21] We selected data from a European Non-Finnish population managed by the GnomAD genome and GnomAD exome databases, and data from a European population managed by the 1000 Genomes database. Before filtering, the presence of high-frequency variants in the *JAK2* and *HFE* genes that cause polycythaemia vera and hereditary haemochromatosis was also assessed. The focus of our analysis was on small variants that involved one or a few nucleotides, such as single nucleotide variants (SNVs) and small insertions/deletions (INDELs).

With the aim to determine the degree of variant position conservation during evolution, conservation analysis was performed with the ConSurf server (https://consurf.tau.ac.il/).[22] The ConSurf server predicts the evolutionary conservation of amino acids or nucleotides based on multiple alignment and phylogenetic relations of homologous sequences that result in position-specific conservation scores. Continuous conservation scores are divided into scale of nine colour grades for visualization, from the most variable position (grade 1), through intermediately conserved position (grade 5), to the most conserved position (grade 9). To each colour grade, a confidence interval is assigned to each conservation score. If the interval spans four or more colour grades, less than six homologous sequences are aligned and the conservation score is unreliable.[22] The parameters for the conservation analysis of protein sequences were as follows: CSI-BLAST search algorithm; 3 iterations; E-value cut-off 0.0001; protein sequence database UNIREF90; number of analysed homologues 150; minimal 35% identity between homologues; maximal 95% identity between homologues; multiple sequence alignment algorithm ClustalW; and method for calculation of the evolution rate Bayesian paradigm. For conservation analysis of nucleotide sequences, we manually searched for homologous sequences using Basic Local Alignment Search Tool (BLAST) program BLASTN (https://blast.ncbi.nlm.nih.gov/Blast.cgi)[23] and perform multiple sequence alignment with ClustalW (https://www.ebi.ac.uk/Tools/msa/clustalo/).[24] For calculation of evolution rate in ConSurf server, we selected Bayesian paradigm.

The pathogenicity of variants in coding regions was assessed using *in silico* prediction tools CADD score (https://cadd.gs.washington.edu/snv),[25] PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/),[26] SIFT (https://sift.bii.a-star.edu.sg/),[27] MutPred2 (http://mutpred.mutdb.org/index.html),[28] SNPs&GO (https://snps.biofold.org/snps-and-go/snps-and-go.html),[29] PANTHER (http://www.pantherdb.org/tools/csnpScoreForm.jsp),[30] PhD-SNP (https://snps.biofold.org/phd-snp/phd-snp.html),[31] PROVEAN (http://provean.jcvi.org/index.php)[32] and Mutation Taster 2 (http://www.mutationtaster.org/)[33] as described by Schiemann and Stowell, 2016,[34] Bris et al. 2018,[35] Wang et al. 2020.[36] The pathogenicity prediction of intron variants and impact on splicing features were analysed using tools CADD score,[25] the RegSNP-intron (https://regsnps-intron.ccbb.iupui.edu/), Human Splicing Finder (http://www.umd.be/HSF/),[37] IntSplice (https://www.med.nagoya-u.ac.jp/neurogenetics/IntSplice/index.html)[38] as reviewed in Ohno et al. 2018[39] and Lin et. al. 2019.[40]

With the CADD (Combined Annotation Dependent Depletion) tool, the prediction of the deleteriousness of variants results in 'raw' score and 'PHRED-scaled' score. A PHRED-scaled score expresses the rank of variant pathogenicity; for example, a score of 10 indicates that the variant is predicted to be in the 10% of the most deleterious substitutions and a score of 20 between the 1% most deleterious. The suggested threshold is between 10 and 20, and we set the cut-off score at >15.[25] The PolyPhen-2 (Polymorphism Phenotyping vs.-2.0) predicts the effect of an amino acid change with classifiers 'benign', 'possibly damaging' and 'probably damaging' and scores from 0.0 to 1.0: variants with values closer to 0.0 are classified as benign and variants with values closer to 1.0 are more confidently predicted as probably damaging. Variants with score over 0.50 were predicted to be pathogenic. In bioinformatics analysis with PolyPhen-2, we present values obtained with HumVar prediction model, which is preferred model for diagnostics of Mendelian diseases.[26] The prediction tool SIFT (Sorting Intolerant From Tolerant) categorizes impact of amino acid change on protein function based on SIFT score ranging from 0 to 1: <0.05 is classified as damaging and >0.05 as tolerated.[27] The output of MutPred2 (Mutation prediction vs.-2.0) consists of a general score that ranges between 0.0 and 1.0, with a higher score indicating greater probability to be pathogenic. The cut-off general score >0.5 was considered to be pathogenic.[28] The prediction tool SNPs&GO uses the reliability index (RI) to evaluate how reliable is the prediction, with 0 being the most unreliable and 10 being the most reliable. Variants are predicted as neutral polymorphism (neutral) or disease related (disease) when the probability score is >0.5.[29] PANTHER (Protein Analysis Through Evolutionary Relationships) estimates the probability of a variant to impact protein function based on evolutionary preservation of a position in protein. The likelihood of deleterious effect increases with the longer preservation time. The thresholds values are >450 million years (my) for 'probably damaging' prediction, between 200 my and 450 my for 'possibly damaging' and <200 my for 'probably benign' prediction.[30] The prediction tool PhD-SNP (Predictor of Human Deleterious SNP) classifies variants into 'neutral polymorphism' or 'disease related' with values from 0 to 1 and the decision threshold for disease causing is >0.5. It also uses reliability index (RI) ranges from 0 to 10, from

the most unreliable to the most reliable prediction. Prediction was used with 20-fold cross-validation.[31] PROVEAN (Protein Variant Effect Analyzer) uses PROVEAN score for binary classification of variants into either deleterious or neutral. The threshold is set to > −2.5 for neutral predictions and < −2.5 for deleterious predictions.[32] The prediction tool Mutation Taster 2 employs a Bayes classifier to predict a variant as one of the four possible types: 'disease causing', that is probably deleterious, 'disease-causing automatic', that is known to be deleterious, 'polymorphism', that is probably harmless and 'polymorphism automatic', that is known to be harmless. It also uses a probability values of the predictions ranging between 0 and 1, with values closer to 1 as more reliable prediction.[33] The RegSNP-intron is a tool that predicts the pathogenicity of intron variants with probability scores 0.00 −0.36 for benign variants; 0.36–0.45 for possibly damaging variants and 0.45–1.00 for damaging variants.[40] The HSF (Human Splicing Finder) allows the identification of all splicing features through multiple algorithms and also the prediction of the impact of intron variants on these features.[37] IntSplice is a tool that predicts a splicing consequence of intron variant close to the 3' end of an intron, and the result is either an abnormal or normal splicing.[38]

The variants were named in agreement with the standard international nomenclature guidelines of the Human Genome Variation Society.[41]

## 3 | RESULTS

### 3.1 | Genetic analysis

Next-generation sequencing revealed 74, 76 and 79 small variants in individuals II:1, I:2 and I:1, respectively. Twelve variants identified in genes associated with erythrocytosis and hereditary haemochromatosis were consistent with autosomal-dominant or autosomal-recessive disease inheritance patterns in the family. Three variants in the *EPAS1* and *VHL* genes were also identified in a reference DNA control NA12878, while the remaining nine variants in the *CDAN1*, *EGLN1* and *JAK2* genes were present only in family members (Table 3). After filtering, three variants in the *VHL*, *JAK2* and *EGLN1* genes were recognized as low-frequency variants (MAF < 0.05) with an autosomal-dominant inheritance pattern.

**TABLE 3** List of all small variants in the analysed family trio consistent with autosomal-dominant or autosomal-recessive inheritance patterns

| Genomic location (hg19) | Gene | Position | HGVS coding DNA/ HGVS protein | SNV ID | MAF[a] | Average DP | Inheritance in family[b] |
|---|---|---|---|---|---|---|---|
| chr. 15:43017426 | *CDAN1* | Exon 27 | NM_138477.2:c.3474A>C / p.(Leu1158=) | rs16957091 | 0.23 | 288 | AD |
| chr. 15:43018486 | *CDAN1* | Intron 24 | NM_138477.2:c.3204+22C>T | rs2305085 | 0.19 | 101 | AD |
| chr. 15:43020983 | *CDAN1* | Exon 20 | NM_138477.2:c.2671C>T / p.(Arg891Cys) | rs8023524 | 0.19 | 421 | AD |
| chr. 15:43021563 | *CDAN1* | Intron 17/ splice region | NM_138477.2:c.2408-3C>T | rs12905385 | 0.19 | 348 | AD |
| chr. 15:43021986 | *CDAN1* | Intron 16/ splice region | NM_138477.2:c.2352+8C>T | rs12594483 | 0.12 | 231 | AD |
| **chr. 1:231557164** | ***EGLN1*** | **Exon 1** | **NM_022051.2:c.471G>C / p.(Gln157His)** | **rs61750991** | **0.03** | **93** | **AD** |
| chr. 2:46603671[c] | *EPAS1* | Intron 8/ splice region | NM_001430.4:c.1035-7C>G | rs7557402 | 0.47 | 184 | AR |
| chr. 2:46605954[c] | *EPAS1* | Intron 11 | NM_001430.4:c.1554+48G>C | rs7598371 | 0.46 | 91 | AR |
| chr. 9:5050706 | *JAK2* | Exon 6 | NM_004972.3:c.489C>T / p.(His163=) | rs2230722 | 0.31 | 297 | AD |
| **chr. 9:5089661** | ***JAK2*** | **Intron 19** | **NM_004972.3:c.2572-13A>G** | **rs780797578** | **0.0001** | **140** | **AD** |
| chr. 9:5090934 | *JAK2* | Intron 22 | NM_004972.3:c.3059+23A>T | rs2274649 | 0.29 | 104 | AD |
| **chr. 3:10187858[c]** | ***VHL*** | **Intron 1** | **NM_000551.3:c.341-325_341-324delTT** | **-** | **0.0004** | **131** | **AD** |

[a]MAF in European population was selected from a European Non-Finnish population managed by GnomAD genome database.

[b]Inheritance in family was determined based on the relationship between genotypes and phenotypes: variants identified as heterozygous in the affected family members and not in the healthy ones are defined with AD; variants identified as homozygous in the affected family members and as heterozygous in the healthy family member are defined with AR.

[c]Variants also identified in a reference DNA control NA12878.

Variants with MAF <0.05 are indicated in bold. HGVS, Human Genome Variation Society; SNV, single nucleotide variant; MAF, minor allele frequency; DP, read depth; *CDAN1*, gene for codanin 1; *EGLN1* gene for egl-9 family hypoxia inducible factor 1, *EPAS1* gene for endothelial PAS domain protein 1, *JAK2*, gene for Janus kinase 2; *VHL*, gene for von Hippel-Lindau tumour suppressor; AD, autosomal dominant; AR, autosomal recessive.

**TABLE 4** Two single nucleotide variants consistent with inheritance pattern and with minor allele frequency below 0.05 in the Slovene family studied

| Genomic location (hg19) | Gene | HGVS coding DNA / HGVS protein | SNV ID | Location | MAF (database) | Functional predictions | Patient | Genotype | VRF |
|---|---|---|---|---|---|---|---|---|---|
| chr. 1:231557164 | *EGLN1* | c.471G>C /p.(Gln157His) | rs61750991 | Exon 1 | 0.03 (GnomAD)<br>0.03 (GnomAD Exome)<br>0.03 (1000 Genomes) | CADD: low pathogenicity<br>PolyPhen−2: benign<br>SIFT: tolerated<br>MutPred2: low pathogenicity<br>SNPs&GO: neutral<br>PANTHER: probably benign<br>PhD-SNP: neutral<br>PROVEAN: neutral<br>Mutation Taster 2: polymorphism<br>ClinVar: benign | I:2<br>II:1<br>I:1 | Heterozygous<br>Heterozygous<br>WT | 0.5<br>0.5<br>- |
| chr. 9:5089661 | *JAK2* | c.2572-13A>G | rs780797578 | Intron 19 | 0.0001 (GnomAD)<br>0.00003 (GnomAD Exome) | CADD: low pathogenicity<br>RegSNP-intron: benign<br>HSF: potential alteration on splicing<br>IntSplice: normal splicing | I:2<br>II:1<br>I:1 | Heterozygous<br>Heterozygous<br>WT | 0.5<br>0.5<br>- |

Abbreviations: HGVS, Human Genome Variation Society; SNV, single nucleotide variant; MAF, minor allele frequency; VRF, variant read frequency; *EGLN1*, gene for egl-9 family hypoxia-inducible factor 1; *JAK2*, gene for Janus kinase 2; WT, wild type.

Variant c.341-325_341-324delTT (NM_000551.3) in the *VHL* gene will not be discussed further here, as it is located in a part of the sequence that is rich in thymine (ie the T nucleotide), and therefore, the deletion of two T nucleotides might be due to sequencer error. Next-generation sequencing confirmed the absence of polycythaemia vera-causing variant p.Val617Phe and exon 12 variants in the *JAK2* gene and hereditary haemochromatosis-causing variants p.Cys282Tyr, p.His63Asp and p.Ser65Asp variants in the *HFE* gene in both of these patients and in healthy family member.

Both affected family members (father I:2, son II:1) were heterozygous for two SNVs in the *JAK2* and *EGLN1* genes, while unaffected family member (individual I:1) did not show either SNV (Table 4; Figure 1A). A missense variant in exon 1 of the *EGLN1* gene was a heterozygous G > C substitution at nt. c.471, which caused an amino acid change at residue 157 from glutamine to histidine (Figure 1B,C). This is a known variant, designated with SNV ID number rs61750991. According to the ClinVar online database, this variant is classified in association with congenital erythrocytosis type 3 (ECYT3) as benign.[42] A non-coding variant in intron 19 of the *JAK2* gene (rs780797578) is located 13 base pairs from the exon 20 (c.2572-13A>G), in proximity to the splicing region (Figure 1D).

### 3.2 | Bioinformatics analysis

A conservation analysis of amino acids in EGLN1 protein using ConSurf server showed that Gln157 is a low conserved amino acid, indicated as a variable position (grade 3), which rapidly evolved. The confidence interval was (4, 2) (Supplementary Figure S1). An amino acid change in p.Gln157His was predicted to have mild effects on protein function according to all nine prediction algorithms (Table 4; Supplementary Table S1).

The ConSurf conservation analysis of 132 nucleotides at the 3′end of *JAK2* intron 19 showed that nucleotide position c.2572-13A is intermediately conserved (grade 6). The confidence interval was (8, 5) (Supplementary Figure S2). The pathogenicity predictions of *JAK2* variant c.2572-13A>G using CADD score and RegSNP-intron tools showed that variant has low deleterious effect on protein function. Similarly, the prediction with IntSplice indicated that variant has no potential impact on splicing. However, the analysis with splice prediction tool HSF revealed that variant c.2572-13A>G could have an effect on splicing, since potential new donor site was identified because of nucleotide change.

## 4 | DISCUSSION

The aim of this study was to identify causative or disease-associated variants in patients with erythrocytosis of unknown cause, in a family with obvious blood parameters and other signs for congenital erythrocytosis. The total number of total SNVs and small insertion/deletions found in the target regions of 39 genes in two affected and one healthy family member varied from 74 to 79. The called

genotypes were consistent with autosomal-dominant or autosomal-recessive inheritance patterns of the disease at only 12 variant sites (approximately 15%) (Table 3). As expected, only a few were missense variants located in coding regions, while the majority of variants were in introns (Table 3). Through evolution, introns have been recognized as a much less conserved parts of sequences and their potential variants have mainly had indirect impacts on structural changes to proteins. Our sequencing data show that a large part of the variants identified were common variants, with MAF >0.05 in European populations, while only three variants identified within the family had frequencies <0.05. As congenital erythrocytosis is a rare clinical disorder, we focused only on the low-frequency variants. Overall, one SNV in the coding region of the *EGLN1* gene and one SNV in the intron region of the *JAK2* gene were examined in detail.

The *EGLN1* missense variant c.471G>C/ p.(Gln157His) was already identified in previous studies.[43,44] Albiero et al.[43] identified this variant in two family members (father and son) with increased haemoglobin, where the father was also positive for *JAK2* p.Val617Phe.[43] Also Ladroue et al.[44] reported this variant in a patient with erythrocytosis.[44] Potential germline mutations in *EGLN1* lead to stabilization of EPAS1 under normoxic conditions, which results in abnormal EPO levels and secondary erythrocytosis.[45] To date, a total of 39 different *EGLN1* variants have been identified in patients with erythrocytosis, which comprise heterozygous missense, frameshifts and nonsense variants.[11] The patients carrying these variants had predominantly normal to elevated EPO levels, similar to the phenotypes of the affected father and son from the present study, who had EPO levels in the normal range. The variant c.471G>C segregated in the family with an autosomal-dominant inheritance pattern (Figure 1A–C), which is consistent with an autosomal-dominant inheritance mode that was recorded for the variants in the *EGLN1* gene, that are causative for the ECYT3 (Table 1). The majority of variants reported in previous studies lie between or near the C-terminal catalytic domain, which is responsible for the HIFα/2-oxoglutarate or ferrous iron binding. Nonsense and frameshift variants can have substantial effects on the EGLN1 protein in terms of impaired function, as these can produce a protein truncated in its C-terminal region; in contrast, missense substitutions have less clear effects.[45] Variant p.(Gln157His) is positioned in the first exon of the *EGLN1* gene, approximately 100 amino acids from the MYND zing-finger domain at the N-terminus (amino acids 21–58) and almost equally distant from the β substrate recognition loop at the C-terminus (amino acids 241–251).[46] The conservation analysis showed low evolutionary conservation of residue p.Gln157 (Supplementary Figure S1), although not all causative variants are positioned at fully conserved residues (eg p.Asn203Lys, p.Lys204Glu, p.Gly285Arg, p.Lys291Ile).[45] According to *in silico* pathogenicity predictions, substitution of glutamine with histidine at residue 157 will have a mild effect on the protein structure and function (Table 4, Supplementary Table S1), which corresponds to a benign interpretation in the ClinVar database.[42] Ladroue et al.[44] also performed *in vitro* hydroxylation assays and showed that this variant has no functional effect on the protein.[44]

Variant c.2572-13A>G is located in intron 19 of the *JAK2* gene (Table 4; Figure 1D). Somatic *JAK2* variants are disease-causing events in patients with polycythaemia vera, as types of myeloproliferative neoplasms. On the other hand, *JAK2* germline variants rarely correlate with haematological disorders. They have been described, however, in patients with polycythaemia vera and other myeloproliferative neoplasms, such as essential thrombocythemia and primary myelofibrosis.[47-49] Kapralova et al.[50] reported on two heterozygous missense germline variants p.Glu846Asp and p.Arg1063His that they identified in a patient with erythrocytosis and abnormal megakaryopoiesis of the bone marrow.[50] They showed that both of these variants lead to erythrocytosis with megakaryocyte abnormalities through hyper-activation of the JAK2/STAT5 signalling pathway via EPOR.[50] Similarly, they suggested that combined impact of two variants, p.Glu846Asp in the *JAK2* gene and p.Gln157His in the *EGLN1* gene, resulted in erythrocytosis phenotype in one patient.[51] Interestingly, we found the same *EGLN1* variant in our studied family, which could imply on the cumulative effect of the two identified variants in *JAK2* and *EGLN1* gene.

The JAK2 protein has four important domains: the FERM domain (amino acids 37–380) and the SH2 domain (amino acids 401–482) have roles in receptor and ligand binding; the JAK homology 2 (JH2) pseudo-kinase domain (amino acids 545–809) acts as a negative regulator of the adjacent JAK2 homology 1 (JH1) kinase domain (amino acids 849–1124).[5,46] The variants associated with haematopoietic disorders lie predominantly in the pseudo-kinase and kinase domains; for instance, the most prevalent somatic variant p.Val617Phe lies in the JH2 domain and disrupts the JH2-JH1 auto-inhibitory interaction, which leads to increased JAK2 kinase activity and activation of its downstream effectors, independent of cytokine binding.[5] Variant c.2572-13A>G identified in the present study is located in intron 19, which is positioned between exons 19 and 20, and these code for the JH1 domain. Previous in silico models have shown that variants in the JH1 domain can have effects on prolonged activation of STAT5 in a low-cytokine environment.[50] The bioinformatic analysis with three prediction tools showed that variant has no impact on splicing or protein function, while the algorithm of Human Splicing Finder pointed to a potential alteration of splicing event (Table 4, Supplementary Table S1). The effect of variant c.2572-13A>G should be further tested in vitro. In accordance with possible impact of variant on protein function were also results from the ConSurf server, as the conservation analysis showed that nucleotide position c.2572–13 was intermediately conserved through evolution (Supplementary Figure S2). Intron variant c.2572-13A>G is a known variant, as it has already been assigned to the SNV ID number rs780797578, although this variant has not been cited in any publication, and also no clinical significance is reported in the ClinVar database.[42]

To study the development of erythrocytosis in the present family, the cumulative effects of both identified variants need to be assessed. In addition, the cumulative effects of common variants (ie MAF >0.05) should not be neglected, as small contributions of common variants can explain large proportions of those at disease risk in the population described by an infinitesimal model.[21] It has been reported that inherited common genetic variants can establish the background for a rare disease, or even rare disorders that are typically considered to be monogenic.[52,53]

## 5 | CONCLUSIONS

In the 39 sequenced erythrocytosis and haemochromatosis-associated genes, we identified two variants in one Slovene family with suspicion of congenital erythrocytosis: c.471G>C/p.(Gln157His) in the *EGLN1* gene and c.2572-13A>G in the *JAK2* gene. Hitherto, neither of these variants can be interpreted as solely causative for the development of erythrocytosis and the cumulative effects of both variants and also between other variants should be assessed. The impact of variant p.(Gln157His) on EGLN1 function was well described by Ladroue et al.[44] in which the functional analysis showed no impairment of hydroxylation of EPAS1.[44] On the contrary, further functional assays for variant c.2572-13A>G in the *JAK2* gene are necessary to better understand the impact of this intron variant on the regulation of protein function.

In the current study, we targeted exon regions and some regulatory regions of 39 genes involved in erythropoiesis, which are not all part of commercially available gene panels for congenital erythrocytosis. As far as we know, this is the first time that this extended selection of both erythrocytosis and hereditary haemochromatosis-associated genes was included in the molecular-genetic analysis of erythrocytosis. We believe, this selection of genes represents a good initial step toward better diagnosis of patients with idiopathic erythrocytosis. For further investigation of the aetiology of congenital erythrocytosis cases, we propose full screening of the non-coding regions of erythrocytosis-associated genes for all statistically significant variants that are potentially associated with increased expansion of RBCs, followed by whole-exome sequencing or whole-genome sequencing, if necessary.

### ACKNOWLEDGEMENTS

This study was supported by the Slovenian Research Agency, grant no. L3-9279, grant no. P1-0390 and Young Researcher funding to AK and the University Medical Centre, Ljubljana, grant no. 20170073.

### CONFLICT OF INTEREST

RK and AV are employees of Kemomed Ltd., Kemomed Research and Development. The other authors declare that they have no conflict of interest.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### ORCID

*Aleša Kristan* https://orcid.org/0000-0001-8644-4347
*Tadeja Režen* https://orcid.org/0000-0001-6210-7370
*Tanja Kunej* https://orcid.org/0000-0002-0465-1762
*Helena Podgornik* https://orcid.org/0000-0003-3752-2067
*Nataša Debeljak* https://orcid.org/0000-0003-1294-292X

### REFERENCES

1. McMullin MF. The classification and diagnosis of erythrocytosis. *Int J Lab Hematol*. 2008;30(6):447-459.
2. Patnaik MM, Tefferi A. The complete evaluation of erythrocytosis: congenital and acquired. *Leukemia*. 2009;23(5):834-844.
3. Lee FS, Percy MJ. The HIF pathway and erythrocytosis. *Annu Rev Pathol*. 2011;6:165-192.
4. Bento C, Percy MJ, Gardie B, et al. Genetic basis of congenital erythrocytosis: mutation update and online databases. *Hum Mutat*. 2014;35(1):15-26.
5. Nangalia J, Grinfeld J, Green AR. Pathogenesis of myeloproliferative disorders. *Annu Rev Pathol*. 2016;11:101-126.
6. McMullin MF. Investigation and management of erythrocytosis. *Curr Hematol Malig Rep*. 2016;11(5):342-347.
7. Amberger JS, Hamosh A. Searching online Mendelian inheritance in man (OMIM): a knowledgebase of human genes and genetic phenotypes. *Curr Protoc Bioinformatics*. 2017;58:1 2 1-2 12.
8. Kristan A, Debeljak N, Kunej T. Genetic variability of hypoxia-inducible factor alpha (HIFA) genes in familial erythrocytosis: analysis of the literature and genome databases. *Eur J Haematol*. 2019;103(4):287-299.
9. Vocanec D, Prijatelj T, Debeljak N, Kunej T. Genetic variants of erythropoietin (EPO) and EPO receptor genes in familial erythrocytosis. *Int J Lab Hematol*. 2019;41(2):162-167.
10. Semenza GL. Oxygen sensing, hypoxia-inducible factors, and disease pathophysiology. *Annu Rev Pathol*. 2014;9:47-71.
11. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD vol 2.0: the next generation in gene variant databases. *Hum Mutat*. 2011;32(5):557-563.
12. Camps C, Petousi N, Bento C, et al. Gene panel sequencing improves the diagnostic work-up of patients with idiopathic erythrocytosis and identifies new mutations. *Haematologica*. 2016;101(11):1306-1318.
13. Bento C. Genetic basis of congenital erythrocytosis. *Int J Lab Hematol*. 2018;40:62-67.
14. Ginzburg YZ, Feola M, Zimran E, Varkonyi J, Ganz T, Hoffman R. Dysregulated iron metabolism in polycythemia vera: etiology and consequences. *Leukemia*. 2018;32(10):2105-2116.
15. Biagetti G, Catherwood M, Robson N, et al. HFE mutations in idiopathic erythrocytosis. *Br J Haematol*. 2018;181(2):270-272.
16. Brissot P, Pietrangelo A, Adams PC, de Graaff B, McLaren CE, Loreal O. Haemochromatosis. *Nat Rev Dis Primers*. 2018;4.18016
17. Debeljak N, Lazarevič J, Miskič D, et al. Characterization of erythrocytosis and a proposed diagnostic algorithm in Slovenia. *Zdrav Vestn*. 2019;88:5-6.
18. Lanktree MB, Sadikovic B, Waye JS, et al. Clinical evaluation of a hemochromatosis next-generation sequencing gene panel. *Eur J Haematol*. 2017;98(3):228-234.
19. Lenglet M, Robriquet F, Schwarz K, et al. Identification of a new VHL exon and complex splicing alterations in familial erythrocytosis or von Hippel-Lindau disease. *Blood*. 2018;132(5):469-483.
20. Zmajkovic J, Lundberg P, Nienhold R, et al. A gain-of-function mutation in EPO in familial erythrocytosis. *N Engl J Med*. 2018;378(10):924-930.
21. Gibson G. Rare and common variants: twenty arguments. *Nat Rev Genet*. 2012;13(2):135-145.
22. Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016;44(W1):W344-W350.
23. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389-3402.
24. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019;47(W1):W636-W641.
25. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D894.

26. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249.
27. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc*. 2016;11(1):1-9.
28. Pejaver V, Urresti J, Lugo-Martinez J, et al. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *bioRxiv*. 2017;134981.
29. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genom*. 2013;14(Suppl 3):S6.
30. Tang H, Thomas PD. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics*. 2016;32(14):2230-2232.
31. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*. 2006;22(22):2729-2734.
32. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015;31(16):2745-2747.
33. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014;11(4):361-362.
34. Schiemann AH, Stowell KM. Comparison of pathogenicity prediction tools on missense variants in RYR1 and CACNA1S associated with malignant hyperthermia. *Br J Anaesth*. 2016;117(1):124-128.
35. Bris C, Goudenege D, Desquiret-Dumas V, et al. Bioinformatics tools and databases to assess the pathogenicity of mitochondrial DNA variants in the field of next generation sequencing. *Front Genet*. 2018;9:632.
36. Wang Z, Huang C, Lv H, Zhang M, Li X. In silico analysis and high-risk pathogenic phenotype predictions of non-synonymous single nucleotide polymorphisms in human Crystallin beta A4 gene associated with congenital cataract. *PLoS One*. 2020;15(1):e0227859.
37. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human splicing finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37(9):e67.
38. Shibata A, Okuno T, Rahman MA, et al. IntSplice: prediction of the splicing consequences of intronic single-nucleotide variations in the human genome. *J Hum Genet*. 2016;61(7):633-640.
39. Ohno K, Takeda JI, Masuda A. Rules and tools to predict the splicing effects of exonic and intronic mutations. *Wiley Interdiscip Rev RNA*. 2018;9(1).e1451
40. Lin H, Hargreaves KA, Li R, et al. RegSNPs-intron: a computational framework for predicting pathogenic impact of intronic single nucleotide variants. *Genome Biol*. 2019;20(1):254.
41. den Dunnen JT. Sequence variant descriptions: HGVS nomenclature and mutalyzer. *Curr Protoc Hum Genet*. 2016;90:7.13.1-7.13.19.
42. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062-D1067.
43. Albiero E, Ruggeri M, Fortuna S, et al. Analysis of the oxygen sensing pathway genes in familial chronic myeloproliferative neoplasms and identification of a novel EGLN1 germ-line mutation. *Br J Haematol*. 2011;153(3):405-408.
44. Ladroue C, Hoogewijs D, Gad S, et al. Distinct deregulation of the hypoxia inducible factor by PHD2 mutants identified in germline DNA of patients with polycythemia. *Haematologica*. 2012;97(1):9-14.
45. Gardie B, Percy M, Hoogewijs D, et al. The role of PHD2 mutations in the pathogenesis of erythrocytosis. *Hypoxia (Auckl)*. 2014;2:71-90.
46. UniProt C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*. 2019;47(D1):D506-D515.
47. Milosevic Feenstra JD, Nivarthi H, Gisslinger H, et al. Whole-exome sequencing identifies novel MPL and JAK2 mutations in triple-negative myeloproliferative neoplasms. *Blood*. 2016;127(3):325-332.
48. Maie K, Yokoyama Y, Yano Y, et al. Progression to polythythemia vera from familial thrombocytosis with germline JAK2 R867Q mutation. *Ann Hematol*. 2018;97(4):737-739.
49. Chiang YH, Chang YC, Lin HC, et al. Germline variations at JAK2, TERT, HBS1L-MYB and MECOM and the risk of myeloproliferative neoplasms in Taiwanese population. *Oncotarget*. 2017;8(44):76204-76213.
50. Kapralova K, Horvathova M, Pecquet C, et al. Cooperation of germ line JAK2 mutations E846D and R1063H in hereditary erythrocytosis with megakaryocytic atypia. *Blood*. 2016;128(10):1418-1423.
51. Kapralova K, Horvathova M, Kucerova J, et al. JAK2 E846D germline mutation associated with erythrocytosis causes increased and prolonged Epo-induced activation of STAT5. *Blood*. 2014;124(21):4008-4008.
52. Bezzina CR, Barc J, Mizusawa Y, et al. Common variants at SCN5A-SCN10A and HEY2 are associated with Brugada syndrome, a rare disease with high risk of sudden cardiac death. *Nat Genet*. 2013;45(9):1044-1049.
53. Niemi MEK, Martin HC, Rice DL, et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature*. 2018;562(7726):268-271.


## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Kristan A, Gašperšič J, Režen T, et al. Genetic analysis of 39 erythrocytosis and hereditary hemochromatosis-associated genes in the Slovenian family with idiopathic erythrocytosis. *J Clin Lab Anal*. 2021;35:e23715. https://doi.org/10.1002/jcla.23715