Characterization of genetic variants in the *EGLN1/PHD2* gene identified in a European collection of patients with erythrocytosis

Marine Delamare,[1,2*] Amandine Le Roy,[1,2*] Mathilde Pacault,[2,3*] Loïc Schmitt,[1,2*] Céline Garrec,[3] Nada Maaziz,[4] Matti Myllykoski,[5] Antoine Rimbert,[2] Valéna Karaghiannis,[1,2] Bernard Aral,[4] Mark Catherwood,[6] Fabrice Airaud,[3] Lamisse Mansour-Hendili,[7,8] David Hoogewijs,[9,10] Edoardo Peroni,[11,12] Salam Idriss,[2] Valentine Lesieur,[2] Amandine Caillaud,[2] Karim Si-Tayeb,[2] Caroline Chariau,[13] Anne Gaignerie,[13] Minke Rab,[14,15] Torsten Haferlach,[16] Manja Meggendorfer,[16] Stéphane Bézieau,[2,3] Andrea Benetti,[17] Nicole Casadevall,[18] Pierre Hirsch,[18] Christian Rose,[19†] Mathieu Wemeau,[20] Frédéric Galacteros,[7,21] Bruno Cassinat,[22] Beatriz Bellosillo,[23] Celeste Bento,[24] Richard van Wijk,[14,15] Petro E. Petrides,[25] Maria Luigia Randi,[17] Mary Frances McMullin,[6,26] Peppi Koivunen,[5] ECYT-3 consortium,[‡] François Girodon,[4,27,28#] and Betty Gardie[1,2,28#]

[1]Ecole Pratique des Hautes Etudes, EPHE, PSL University, Paris, France; [2]Université de Nantes, CNRS, INSERM, l'Institut du Thorax, Nantes, France; [3]Service de Génétique Médicale, CHU de Nantes, Nantes, France; [4]Service d'Hématologie Biologique, Pôle Biologie, CHU de Dijon, Dijon, France; [5]Biocenter Oulu and Faculty of Biochemistry and Molecular Medicine, Oulu Center for Cell-Matrix Research, University of Oulu, Oulu, Finland; [6]Belfast Health and Social Care Trust, Belfast, North Ireland; [7]Département de Biochimie-Biologie Moléculaire, Pharmacologie, Génétique Médicale, AP-HP, Hôpitaux Universitaires Henri Mondor, Créteil, France; [8]Université Paris-Est Créteil, IMRB Equipe Pirenne, Laboratoire d'Excellence LABEX GRex, Créteil, France; [9]Section of Medicine, Department of Endocrinology, Metabolism and Cardiovascular System, University of Fribourg, Fribourg, Switzerland; [10]National Center of Competence in Research "Kidney.CH", Zurich, Switzerland; [11]Immunology and Molecular Oncology Unit, Veneto Institute of Oncology, IOV-IRCCS, Padova, Italy; [12]Medical Genetics Unit, Mater Domini University Hospital, Catanzaro, Italy; [13]Nantes Université, CHU Nantes, CNRS, Inserm, BioCore, Nantes, France; [14]Central Diagnostic Laboratory - Research, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; [15]Department of Hematology, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands; [16]Munich Leukemia Laboratory, Munich, Germany; [17]Department of Medicine-DIMED, University of Padua, Padua, Italy; [18]Sorbonne Université, INSERM, Centre de Recherche Saint-Antoine, CRSA, AP-HP, SIRIC CURAMUS, Hôpital Saint-Antoine, Service d'Hématologie Biologique, Paris, France; [19]Service d'Onco-Hématologie, Saint-Vincent de Paul Hospital, Université Catholique de Lille, Université Nord de France, Lille, France; [20]Hematology Department, Claude Huriez Hospital, Lille Hospital, Lille, France; [21]Red Cell Disease Referral Center-UMGGR, AP-HP, Hôpitaux Universitaires Henri Mondor, Créteil, France; [22]Université Paris Cité, APHP, Hôpital Saint-Louis, Laboratoire de Biologie Cellulaire, Paris, France; [23]Pathology Department, Hospital del Mar-IMIM, Barcelona, Spain; [24]Hematology Department, Centro Hospitalar e Universitário de Coimbra; CIAS, University of Coimbra, Coimbra, Portugal; [25]Hematology Oncology Center and Ludwig-Maximilians-University Munich Medical School, Munich, Germany; [26]Queen's University, Belfast, North Ireland; [27]Inserm U1231, Université de Bourgogne, Dijon, France and [28]Laboratoire d'Excellence GR-Ex, Paris, France.

**MD, ALR, MP and LS contributed equally as first authors.*

#FGi and BG contributed equally as senior authors.

†deceased

‡ An appendix with all ECYT-3 consortium members can be found at the end of the manuscript.

Correspondence: B. Gardie
betty.gardie@inserm.fr

Received: February 10, 2023.
Accepted: June 6, 2023.
Early view: June 15, 2023.

https://doi.org/10.3324/haematol.2023.282913

Abstract

Hereditary erythrocytosis is a rare hematologic disorder characterized by an excess of red blood cell production. Here we describe a European collaborative study involving a collection of 2,160 patients with erythrocytosis sequenced in ten different laboratories. We focused our study on the *EGLN1* gene and identified 39 germline missense variants including one gene deletion in 47 probands. *EGLN1* encodes the PHD2 prolyl 4-hydroxylase, a major inhibitor of hypoxia-inducible factor. We performed a comprehensive study to evaluate the causal role of the identified *PHD2* variants: (i) *in silico* studies of localization, conservation, and deleterious effects; (ii) analysis of hematologic parameters of carriers identified in the UK Biobank; (iii) functional studies of the protein activity and stability; and (iv) a comprehensive study of *PHD2* splicing. Altogether, these studies allowed the classification of 16 pathogenic or likely pathogenic mutants in a total of 48 patients and relatives. The *in silico* studies extended to the variants described in the literature showed that a minority of *PHD2* variants can be classified as pathogenic (36/96), without any differences from the variants of unknown significance regarding the severity of the developed disease (hematologic parameters and complications). Here, we demonstrated the great value of federating laboratories working on such rare disorders in order to implement the criteria required for genetic classification, a strategy that should be extended to all hereditary hematologic diseases.

Introduction

Red blood cell production is tightly regulated by the oxygen-sensing pathway, which controls the expression of erythropoietin (EPO), a glycoprotein hormone that stimulates the survival, proliferation, and differentiation of erythroid progenitors. The main enzymes that are capable of sensing oxygen are the dioxygenases whose enzymatic activity controls the hydroxylation of target proteins, utilizing oxygen and 2-oxoglutarate as co-substrates.

This study focuses on the *EGLN1* (Egl nine homolog 1) gene that encodes the dioxygenase prolyl hydroxylase domain-containing protein 2 (PHD2) (also called HIF prolyl 4-hydroxylase-2, HIF-P4H-2).[1-3] In the presence of oxygen, PHD2 hydroxylates its main substrate, the α subunit of the hypoxia-inducible factor (HIF), which is a heterodimeric transcription factor (α/β) that plays a central role in oxygen homeostasis. The hydroxylated HIF-α (at proline residues 402 and 564 for HIF-1α and 405 and 531 for HIF-2α) then binds to von Hippel-Lindau (VHL) tumor suppressor protein, which directs HIF-α for proteasomal degradation through VHL E3 ubiquitin ligase activity. Thus, the orchestrated action of PHD and VHL proteins drives HIF degradation in the presence of oxygen. When oxygen concentration decreases, PHD enzymatic activity is diminished, which causes the stabilization of HIF-α. As a result, HIF-α accumulates in the nucleus, forms an active heterodimer complex with HIF-1β that binds hypoxia-responsive elements (HRE) and induces target gene expression. HIF regulates the transcription of more than 200 genes involved in many pathways,[4] such as erythropoiesis (via the synthesis of EPO), iron regulation, angiogenesis, metabolism, cell proliferation and survival. *EGLN1* is a HIF target gene and its expression contributes to a negative feedback mechanism that limits HIF-1 responses during reoxygenation.[5-7] There are three isoforms of PHD (PHD1-3), with PHD2 showing the highest oxygen-dependent activity.[8]

Germline loss-of-function mutations in the *EGLN1/PHD2* gene were first described in 2006 in a patient with hereditary erythrocytosis.[9] Erythrocytoses are characterized by an elevated red cell mass reflected by increased hemoglobin and hematocrit[10] levels. Primary hereditary erythrocytosis occurs when the mutations target the intrinsic mechanism of erythroid progenitors with overproliferation (i.e., *EPOR*), and a compensatory lowered EPO serum level. Secondary hereditary erythrocytosis occurs when a high level of EPO is produced, thus indirectly driving red blood cell overproduction. In this case, the circulating EPO levels are inappropriately normal or elevated. Patients with *PHD2* mutations develop secondary hereditary erythrocytosis with high to normal EPO serum levels, a normal concentration being inappropriate in view of the high hematocrit. In contrast to patients with the Chuvash *VHL*-R200W mutation, erythroid progenitors do not exhibit hypersensitivity to EPO (a feature of primary erythrocytosis). However, mice targeted for *Phd2* inactivation in hematopoietic precursors showed hypersensitivity to EPO.[11] As discussed below, the impact of *PHD2* mutations on erythroid progenitors is still under debate. Interestingly, genetic variants in the *PHD2* and *HIF2A* genes have been associated with the adaptation of Tibetans to high altitude.[12,13] The mechanisms involved appear to be complex,[14,15] but studies have shown that particular variants of PHD2 (c.12C>G, p.Asp4Glu and c.380G>C, Cys127Ser *in cis*) result in a gain of function of the protein.[15]

Since the first described case, 64 different *PHD2* genetic variants (in 72 families) have been described, including missense, frameshift and nonsense mutations (*Online Supplementary Table S1*, p.Asp4Glu and p.Cys127Ser not included).[16-18] Patients with erythrocytosis due to a mutation in the *EGLN1* gene are all heterozygous, with the exception of two siblings described as homozygous for p.Cys42Arg.[19] The phenotypes of patients carrying these mutations are usually limited to erythrocytosis, but complications such as

thrombosis, hypertension, renal cysts, angiomas, and rarely paraganglioma[20,21] or pheochromocytoma[22] may arise.[16,17] *In vitro* functional studies performed on *PHD2* variants have not been able to explain the differences in clinical presentation.[9,19-21,23-27] In particular, studies demonstrated that some PHD2 mutations induce a subtle loss of function, sometimes very close to the wild-type protein.[26]

This article describes a collaborative study of 34 novel genetic variants (in addition to 5 previously described[16,19,28-33]) identified in the *EGLN1* gene using next-generation sequencing panels. The study includes data from sequencing of a total of 2,160 patients with hereditary or idiopathic erythrocytosis recruited in seven European countries. Forty-seven cases carrying genetic variants in *PHD2* were identified. We performed comprehensive *in silico* and functional studies to decipher their potential causal role in the disease pathogenesis.

Methods

Sequencing

All study participants signed written informed consent. Blood samples were collected for research purposes after receiving approval from the different local ethics committees. DNA was extracted and molecular screening was performed by next-generation sequencing with different technologies, depending on the sequencing center.

In silico analyses

The MobiDetails annotation platform[34] was used for the interpretation of DNA variations (frequencies in the control population, prediction of the impact of missense variants and analysis of splicing with the Splicing Prediction Pipeline (SPiP)) and for the localization of the mutated amino acids on the three-dimensional protein structure (AlphaFold Protein Structure Database). The localization of the affected amino acids on the two-dimensional structure was analyzed using the MetaDome website.[35] The UK Biobank resource was used to analyze the hematologic parameters of the carriers of *EGLN1* genetic variants.

Luciferase reporter assay

End-point Luciferase assays were performed as previously described.[20,26] Real-time Luciferase assays were performed on HEK 293T cells transfected with jetPRIME® (Ozyme Polyplus). Expression vectors pcDNA3-HA-PHD2 were co-transfected with pcDNA3-HA-HIF-2α, pGL3 3xHRE-luciferase reporter plasmid and pCMV-HA-empty vector. Luciferase activity was measured over 24 h using the bioluminometer WSL-1565 Kronos HT® (ATTO). Cells were harvested at different timepoints and lysed in passive lysis buffer (Invitrogen) for immunoblot detection with a mouse anti-HA antibody (clone 16B12, BioLegend).

In vitro enzyme activity assays

Flag-tagged PHD2 was expressed in insect cells and affinity-purified using anti-Flag, and the His-tagged HIF-2α oxygen-dependent degradation domain (ODDD) protein was expressed in *E. coli* and affinity-purified using NiNTA as described previously.[36,33] A PHD2 enzymatic activity assay was performed to measure the radioactive CO_2 produced during the decarboxylation of 2-oxo[1-^{14}C]glutarate (Perkin-Elmer), which co-occurs with the substrate proline hydroxylation.[37]

Cycloheximide chase assay

HEK cells were transfected with plasmids encoding pcDNA3-HA-PHD2 (100-800 ng) in addition to pCMV-HA-empty vector for a total amount of 800 ng of transfected DNA. Twenty-four hours after transfection, cells were treated with cycloheximide (Sigma-Aldrich) at a final concentration of 100 μg/mL. Cells were harvested at different timepoints. An equal volume of protein lysates was analyzed by western blot assay using a mouse anti-HA antibody, anti-actin (Sigma) and a goat anti-mouse secondary antibody (Jackson Immunoresearch).

Splicing reporter assay

Minigene constructs were prepared as described by Cooper and Gaildrat.[38] Cells were transfected with 2 μg pCas2 plasmid containing the exon of interest and flanking intronic sequences surrounded by artificial exons (named A and B). Total RNA was extracted 24 h after transfection. Reverse transcriptase polymerase chain reaction (RT-PCR) was performed with primers located in exons A and B. PCR products were resolved on a 2% agarose gel or quantified by using the Agilent TapeStation® system.

Generation of human induced pluripotent stem cells and differentiation into hepatocyte-like cells

Human induced pluripotent stem cells (hiPSC) were generated from peripheral blood mononuclear cells and characterized in the hiPSC Core Facility of Nantes University. hiPSC from passage numbers 21 to 25 were differentiated into hepatocyte-like cells, as described previously.[39,40] After 22 days of differentiation, cells were cultured for 24 h in normoxia or at 1% O_2 in an Invivo 400 Hypoxia Workstation (Baker Ruskinn).

Results

Sequencing of patients with erythrocytosis

Patients with erythrocytosis were recruited according to the following criteria: red cell mass >125% and/or elevated hematocrit (>52% in men, >47% in women) and hemoglobin (>18 g/dL in men and >16 g/dL in women) levels.

Cases of polycythemia vera and secondary erythrocytosis, particularly related to cardiac or pulmonary insufficien-

cy, had been previously excluded. A flow chart detailing the steps that led to the molecular screening performed by high-throughput sequencing is presented in Figure 1. Next-generation sequencing enabled the investigation of large rearrangements (deletions and duplications) together with point mutations. Samples from a total of 2,160 patients were sequenced in seven countries (10 diagnostic centers listed in Table 1). We selected variants with a gnomAD frequency ≤5x10^{-4}. In total, we identified 39 genetic variants including one complete deletion in 47 families (Table 1). We identified, for the first time, a deletion of one copy of the entire *EGLN1* gene by next-generation sequencing. The

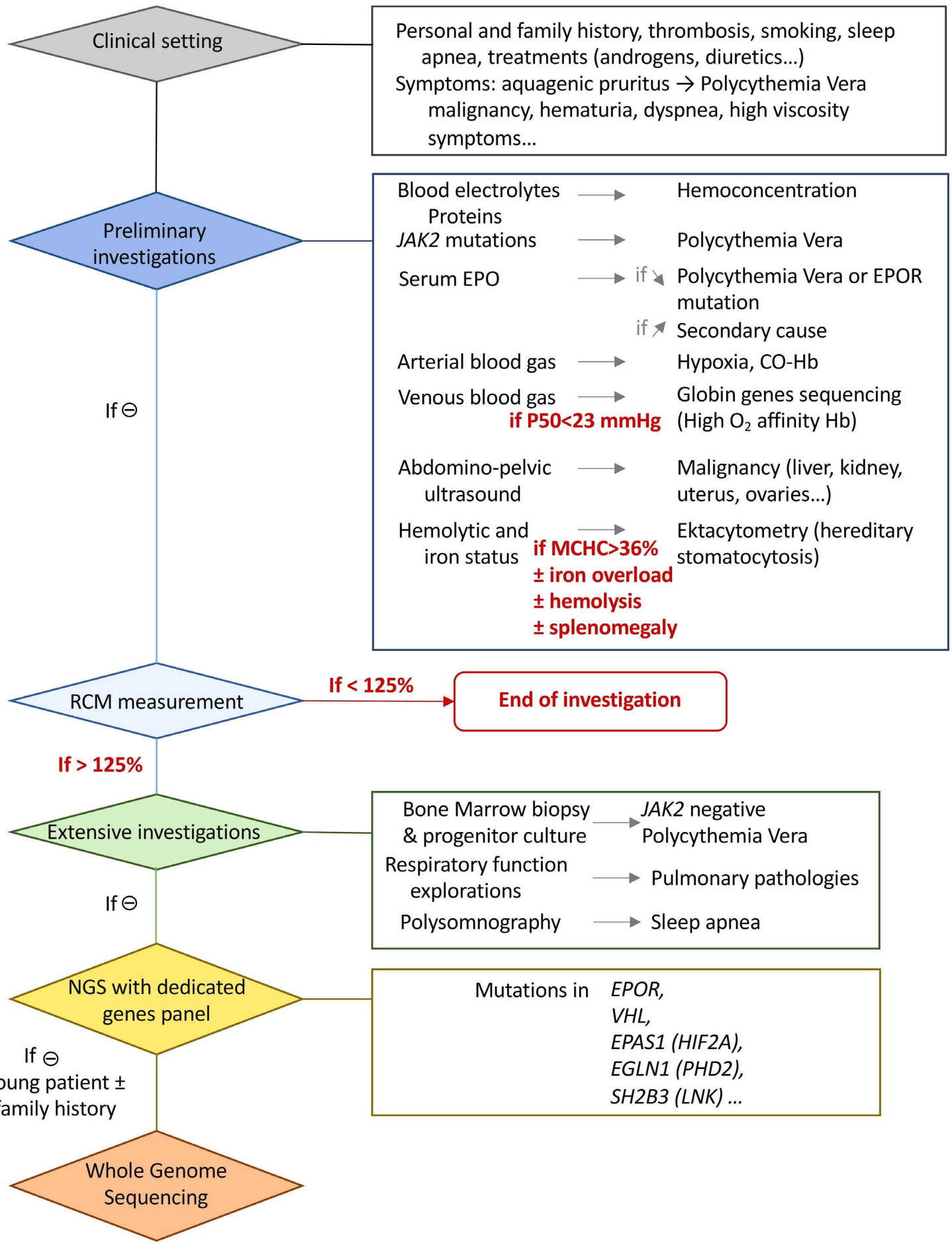


Figure 1. Diagnostic flow chart for patients presenting with erythrocytosis. EPO: erythropoietin; EPOR: erythropoietin receptor, CO: carbon monoxide; Hb: hemoglobin; P50: partial pressure; O_2: oxygen; MCHC: mean corpuscular hemoglobin concentration; RCM: red cell mass; NGS: next-generation sequencing.

Table 1. List of genetic variants identified in the *EGLN1* gene in patients with erythrocytosis.

ID	Exon	Position cDNA	Position protein	Letter code	Age/ age at Dx in yrs	Sex	Hb g/dL	Hct %	RBC 10^6/mm^3	EPO mIU/mL	Family history	Other symptoms	Treatment	Classification	Dx center	Ref
P #1	5'UTR	c.-410G>T	NA	NA	73	M	17	51	6.05	17	-	Hypereosinophilia	IFN-α	VUS	1	-
P #2	1	c.104G>A	p.Arg35His	R35H	75	M	17.1	51.2	5.71	-	-	-	-	VUS	7	-
P #3	1	c.122A>G	p.Tyr41Cys	Y41C	18	M	19.2	57	-	9	No	Headaches	Asp	LP	1	2, 4
P #4	1	c.148G>C	p.Asp50His	D50H	50/39	M	19.9	54.8	6.17	17	No	Portal vein thrombosis	-	VUS	1	-
P #5	1	c.230C>T	p.Pro77Leu	P77L	81	M	-	-	-	-	-	-	-	VUS	3	-
P #6	1	c.287C>T	p.Ala96Val	A96V	25	F	22.6	65.6	6.97	6.3	-	-	Asp	LB	1	-
P #7	1	c.400C>T	p.Gln134*	Q134*	65	M	18.4	54	6.1	6.03	+ 2	-	-	P	1	10ψ
P #8	1	c.470A>G	p.Gln157Arg	Q157R	30	M	20.2	57.9	5.07	3.2	-	-	-	LB	1	-
P #9	1	c.489C>A	p.Tyr163*	Y163*	-	M	20.5	58	6.83	9.2	+ 2	Ear pain, ulcerative colitis	Asp	P	6	-
P #10	1	c.568_569delinsTT	p.Ala190Leu	A190L	22	M	15	51.8	-	30	+ 2	-	-	VUS	3	-
P #11	1	c.629_631delTGG	p.Val210del	V210Del	28	F	16.7	49	-	-	+	-	-	VUS	4	-
P #12	1	c.661C>T	p.Gln221*	Q221*	-	F	16.7	53	5.09	-	No	None	Phleb	LP	6	-
P #13	1	c.665T>G	p.Ile222Ser	I222S	50	M	-	-	-	-	-	-	-	VUS	7	-
P #14	1	c.698G>C	p.Gly233Ala	G233A	31	M	17.4	49	-	10.4	+ 2	-	-	VUS	4	-
P #15	1	c.715C>T	p.Gln239*	Q239*	51	F	18	55.4	5.7	10	+ 4	MI, heavy smoker	Phleb/ Asp	P	5	18, 20
P#16	1	c.763A>G	p.Lys255Glu	K255N	58	F	16.9	52	-	-	No	PTE	-	VUS	4	-
P #17	1	c.794_795delinsTT	p.Gly265Val	G265V	14	M	15.4	46.6	-	-	+ 5	Psychomotor retardation	No	VUS	1	-
P #18	1	c.803C>T	p.Thr268Ile	T268I	31	M	-	-	-	-	-	-	-	VUS	4	-
P #19	1	c.806T>C	p.Ile269Thr	I269T	50	M	20.4	61.5	7	6.6	-	Exertional dyspnea	Asp	P	1	-
P #20	1	c.806T>C	p.Ile269Thr	I269T	18	M	17.1	52.6	6.23	3.5	+ 2	None	Asp	P	1, 5	-
P #21	1	c.806T>C	p.Ile269Thr	I269T	43	M	18.3	55.6	6.11	6	No	None	Phleb	P	9	-
P #22	1	c.806T>C	p.Ile269Thr	I269T	22	M	18.6	52	6.11	17.9	+ 3	-	Phleb/Asp	P	2	-
P #23	1	c.806T>C	p.Ile269Thr	I269T	-	M	17.1	51.5	-	5.8	-	-	-	P	10	-
P #24	1	c.808_811dup	p.Leu271Argfs*15	L271fs	28	M	16.8	48.5	5.87	-	-	-	-	P	3	-
P #25	1	c.887C>T	p.Thr296Met	T296M	22	M	18.6	53.8	6.48	11.9	-	Asthenia, HFE	-	VUS	1	-
P #26	/	c.891+1G>A	NA	NA	35	M	18.8	54	5.58	15	No	None	Phleb/Asp	P		-
P #27	2	c.908A>G	p.Tyr303Cys	Y303C	-	M	19.8	56	7.1	-	No	-	-	VUS	2	-

Continued on following page.

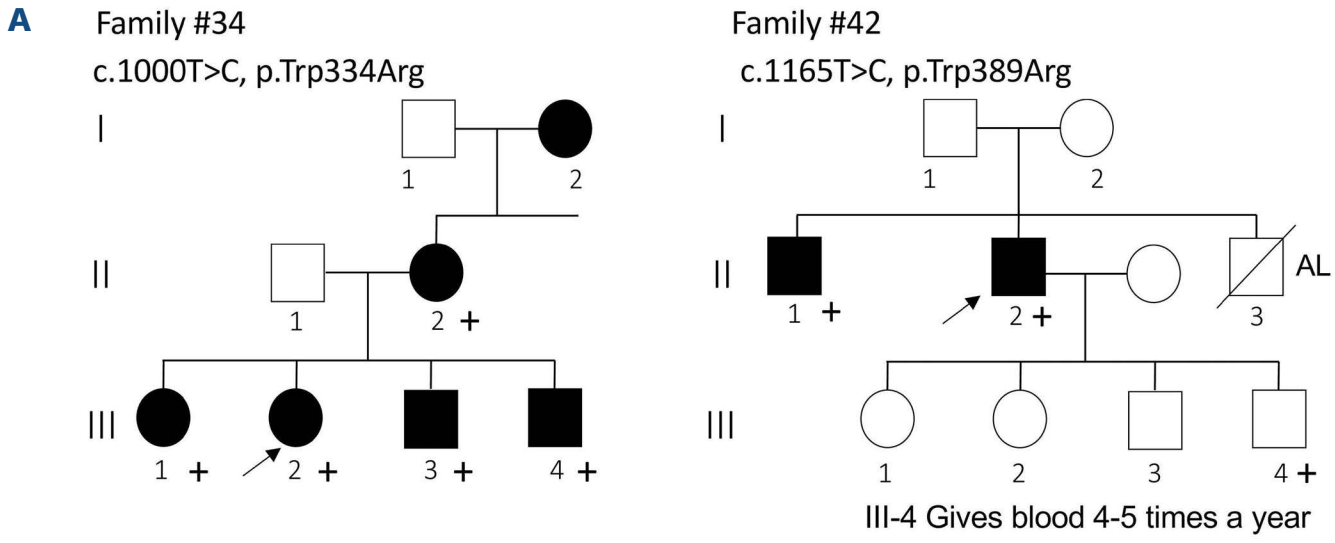
ID	Exon	Position cDNA	Position protein	Letter code	Age/ age at Dx in yrs	Sex	Hb g/dL	Hct %	RBC 10^6/mm^3	EPO mIU/mL	Family history	Other symptoms	Treatment	Classification	Dx center	Ref
P #28	2	c.935G>A	p.Arg312His	R312H	-	M	18.5	52	5.95	5	-	-	-	LP	6	-
P #29	2	c.935G>A	p.Arg312His	R312H	-	F	15.6	47	5.85	45.5	+ 3	None	Phleb	LP	6	-
P #30	2	c.935G>A	p.Arg312His	R312H	49	M	20.1	61.5	6.54	15	No	CVA, OSAS	Phleb	LP	2	-
P #31	2	c.935G>A	p.Arg312His	R312H	62	M	-	-	-	-	-	-	-	LP	3	-
P #32	2	c.985T>C	p.Tyr329His	Y329H	1	F	16.8	49.7	5.75	6.2	+ 5	-	-	VUS	9	-
P #33	2	c.990dup	p. Asn331*	N331*	63	M	20.7	62	-	Normal	+ 4	-	-	P	9	-
P #34	2	c.1000T>C	p.Trp334Arg	W334R	35	M	17.6	49.9	-	-	+ 6	None	Asp	P	2, 1	9ψ
P #35	3	c.1045G>T	p.Gly349Cys	G349C	75	F	17.6	48-58	6.22	16.5	+ 3	Phlebitis, RVO	Phleb/ Asp	VUS	1	-
P #36	3	c.1045G>A	p.Gly349Ser	G349S	21	M	16.9	50.1	-	-	No	-	-	VUS	9	-
P #37	3	c.1072C>A	p.Pro358Thr	P358T	88	F	18.9	53.4	5.87	8.07	-	-	-	VUS	7	-
P #38	3	c.1096T>C	p.Phe366Leu	F366L	46	M	18.8	55.8	6.66	15	-	-	-	VUS	2, 5	26
P #39	3	c.1108C>G	p.Arg370Gly	R370G	58	M	19.6	53.8	-	5	+ 2	None	Phleb	VUS	9	-
P #40	3	c.1129C>T	p.Gln377*	Q377*	50/34	F	18	53.3	5.87	69	-	-	-	P	10	14
P #41	4	c.1152C>T	p.Tyr384Tyr	Y384Y	26	M	13.4	42.8	-	91	Ad	TIA, myalgia, arthralgia	Phleb/Asp	LP	1	-
P #42	4	c.1165T>C	p.Trp389Arg	W3389R	41	M	20	56	6.08	2.4	+ 3	-	Phleb/Asp	P	1	-
P #43	4	c.1165T>C	p.Trp389Arg	W3389R	26	F	18.1	54	-	50	+ 3	MI	Hydroxyurea	P	1	-
P #44	/	c.1216+1G>T	NA	NA	62	F	12.9	44	5.91	24.1	+ 3	-	-	LP	3	-
P #45	5	c.1257T>G	p.Asp419Glu	D419E	23	M	-	49.6	5.33	9.7	No	Thrombocytopenia	Phleb	VUS	1	-
P #46	5	c.1259C>T	p.Ser420Leu	S420L	-	M	18	53	-	-	No	-	-	VUS	5	-
P #47	1-5	Deletion of the entire gene	NA	NA	56	F	15.6	47.4	-	-	No	-	Asp	P	1	-

ID: identification; Dx: diagnosis; Hb: hemoglobin (normal values are 13-18 g/dL for men and 12-16 g/dL for women); Hct: hematocrit (normal values are 40-52% for men and 37-47% for women); RBC: red blood cells (normal values are 4.2-5.7x10^6/mm^3 for men and 4.2-5.2x10^6/mm^3 for women); EPO: erythropoietin (normal values are 5-25 mIU/mL). It is not excluded that the hematologic values were measured after phlebotomies. Family history (+: erythrocytosis diagnosed in the family with the number of additional members indicated; No: no family history; Ad: adopted); Ref: references (detailed references are listed in the *Online Supplementary Data*); P #: patient number; UTR: untranslated region; NA: not applicable, -: data not available; M: male; F: female; *: stop codon, Fs: frameshift; MI: myocardial infarction; PTE: post-renal transplant erythrocytosis; HFE: hemochromatosis; CVA: cerebral vascular accident; OSAS: obstructive sleep apnea syndrome; RVO: retinal vein occlusion; TIA: transient ischemic attack; IFN-α: interferon-alpha; Asp: aspirin; Phleb: phlebotomies; VUS: variant of unknown significance; LB: likely benign; LP: likely pathogenic; P: pathogenic; Diagnosis Center: 1: Nantes (France), 2: Dijon (France), 3: Munich (Germany), 4: Belfast (UK), 5: Créteil (France), 6: Utrecht (Netherland), 7: Coimbra (Portugal), 8: Padova (Italy), 9: Barcelona (Spain), 10: Paris (France). ψ: same proband as already published but with updated information in the present paper.

result was confirmed by quantitative PCR on DNA from additional biological samples. Remarkably, the patient with this deletion (P #47) did not present a more severe phenotype than that of patients carrying missense mutations. Of note, as usually described in hereditary erythrocytoses, the majority of probands were men (72.3%; 34 men *vs*. 13 women). Segregation studies of the identified variants were possible in a limited number of cases and allowed the identification of 34 additional carriers. Examples of pedigrees are shown in Figure 2A. One pedigree shows a large family with a history of erythrocytosis over three generations. The variant c.1000T>C, p.Trp334Arg was identified in proband III.2[16] and detected in four additional relatives. The other pedigree shows the family of patient P#42 carrying the variant c.1165T>C, p.Trp389Arg, with segregation of the variant in family members with erythrocytosis, except for patient III.4, who was described as an asymptomatic carrier. After investigation, we found that this patient was a very assiduous blood donor, suggesting an incidental regulation of his hematologic parameters.

In silico analyses of the variants

First, we performed an *in silico* analysis of the variants identified to localize the missense variants in the protein structure. For this purpose, we first used the MetaDome server,[35] which allows visualization of gene‐wide profiles of genetic tolerance on the two-dimensional protein structure. We localized the PHD2 variants described in the literature



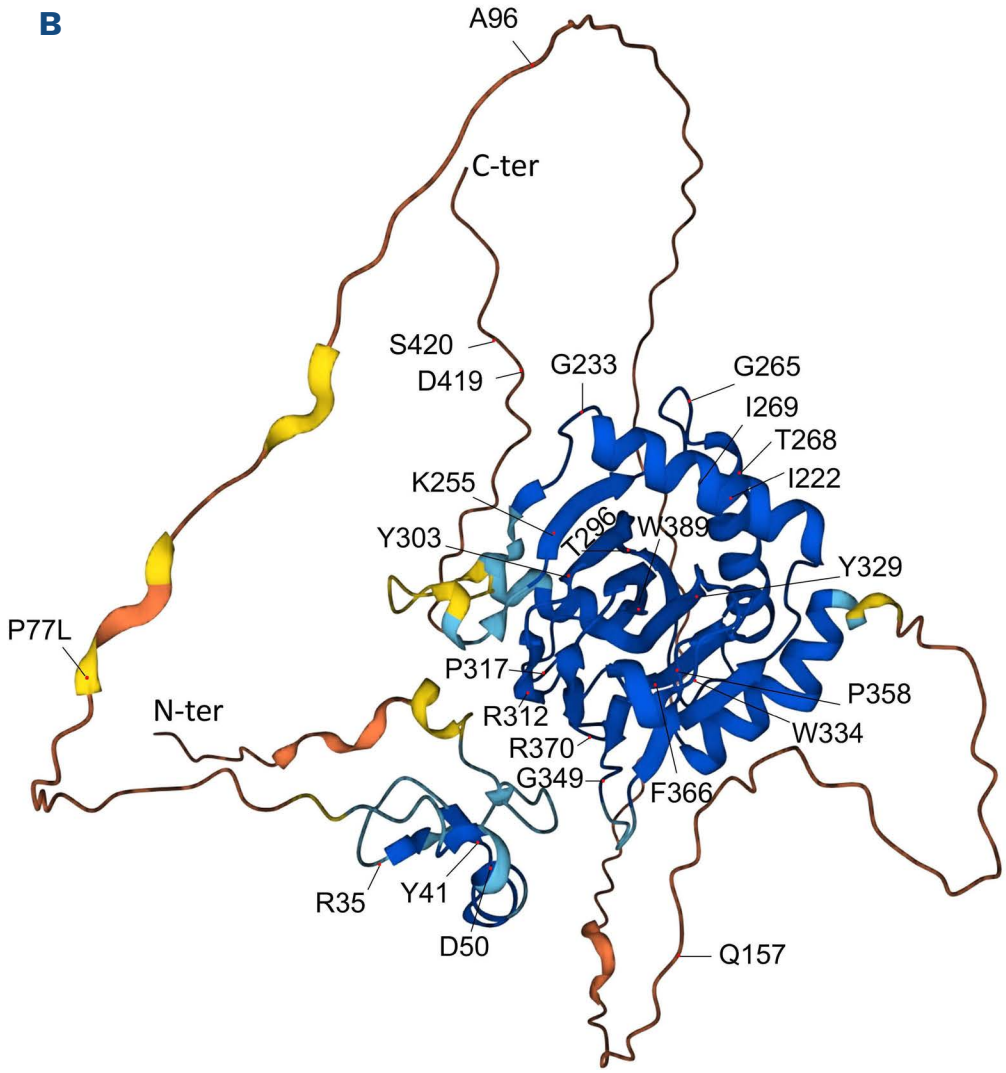


Figure 2. Pedigrees of families carrying genetic variants in *EGLN1* and localization of the mutated amino acids on the PHD2 protein. (A) Pedigrees of two families carrying genetic variants in *EGLN1*. (Left) The pedigree of patient #34 carrying the variant c.1000T>C, p.Trp334Arg. (Right) The pedigree of patient #42 carrying the variant c.1165T>C, p.Trp389Arg. +: carrier of the genetic variant. The arrow indicates the proband. (B) Localization of the targeted amino acids on the three-dimensional structural prediction of PHD2 obtained from the AlphaFold Protein Structure Database via the MobiDetails website. The regions of the structure modeled with high, medium, low, or very low confidence are colored blue, light blue, yellow, and orange, respectively. The prediction indicates that the relative positions of the two folded domains are not reliably modeled.

(*Online Supplementary Figure S1A*, left panel) and the missense variants described in this study (right panel) on the Protein's Tolerance Landscape. The variants were distributed throughout the protein, with no particular hotspot of mutations. The majority of the amino acids targeted by the genetic variations in the present study are expected to have a moderate impact (orange on the scale). Only three are expected to be highly intolerant (red on the scale): D50H, V210del, and Y329H (scores are detailed in *Online Supplementary Table S3*). The targeted amino acids on the three-dimensional structure were localized using the AlphaFold Protein Structure Database through the MobiDetails website (Figure 2B). PHD2 has two main domains: a N-terminal MYND-type zinc finger domain and a C-terminal prolyl hydroxylase domain with the double-stranded β-helix core fold supported by surrounding α-helices (in blue in Figure 2B). Most of the amino acids mutated in patients are located within the "core" domain, whereas R35, Y41 and D50 are located within the N-terminal zinc finger domain and P77L, A96, Q157, D419 and S420 are located outside the main domains. The following *in silico* analyses showed that the four variants located outside the structured domains are not pathogenic. Careful analysis of the N-terminal sequence of PHD2 revealed that the amino acids affected by the variations are located near to cysteines residues that play a major role in the function of the MYND-type zinc finger domain (*Online Supplementary Figure S1B*).

In silico analysis was performed for the variant c.-410G>T, which is located upstream of the coding sequence. This variant is positioned precisely in the HRE-HIF binding consensus sequence responsible for the previously described regulation of *PHD2* expression in hypoxia.[6] Indeed, the variant targets the first G of the core HIF binding consensus sequence (ACGTG) and consequently may disrupt the regulation of PHD2 in hypoxia (*Online Supplementary Figure S1C*).

The conservation of the PHD2 targeted amino acids was studied throughout a broad range of taxonomic classes covering primates, mammals, lower vertebrates as well as invertebrates (*Online Supplementary Figure S2*). A majority of them (63%, 17/27) are highly conserved through species and PHD isoforms (highlighted in black in the figure).

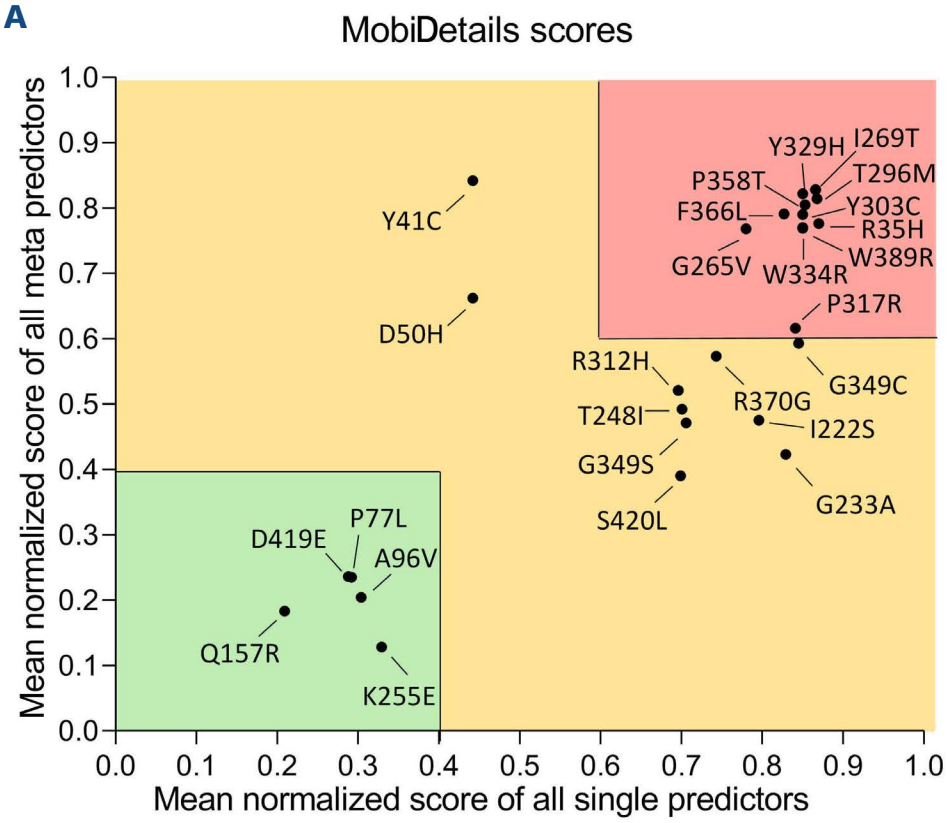
We examined the likely functional impact of *PHD2* missense variations using MobiDetails,[34] a dedicated annotation platform for interpreting DNA variations. We used Radar graphs to represent the *in silico* prediction by the main software tools (*Online Supplementary Figure S3*). For each variant, the scores of individual and meta-predictors were plotted on a graph to allow their classification into three categories: probably damaging, possibly damaging/tolerated, or benign (Figure 3A). These analyses showed that all identified variants had elevated scores, with the exception of variants P77L, A96V, Q157R, K255E and D419H (which are mainly located outside the main domains in the three-dimensional structure).

We analyzed the UK Biobank database, a large-scale biomedical database containing in-depth genetic and health information from half a million participants from the United Kingdom. We selected the *PHD2* variants identified in the literature and this study (Figure 3B) and/or other variants associated with elevated hematocrit and hemoglobin levels (*Online Supplementary Figure S4*). We focused on men because, for this pathology, it may be difficult to interpret hematocrit and hemoglobin levels in women because of their menstrual cycles. For some variants (A96V, C127S, Q157H and Q157R, in green, Figure 3B), a significant number of carriers had hematocrit and hemoglobin levels distributed equally and comparably to those of wild-type individuals, strengthening the arguments in favor of a benign effect of the variants. Of note, we found two men carrying the Q221* and I269T variants (in red, Figure 3B) with hematocrit and hemoglobin values above the 99th percentile, a finding in favor of a deleterious effect of these variants. Some variants (R371H and T296M, light red) are associated with hematocrit and hemoglobin levels >90th to 99th percentile but their interpretation remains reserved. Additional variants have been associated with these elevated hematocrit and hemoglobin levels (*Online Supplementary Figure S4*), a list that could be very informative when identifying future variants.

Functional studies of PHD2 variants using luciferase reporter tests

The effect of the genetic variants on PHD2 activity was assessed using a reporter assay for HIF transcriptional activity. The HIF-2α subunit was co-expressed with a luciferase gene driven by HRE (Figure 4A). The activity of accumulated luciferase was quantified 24 h after transfection. The addition of wild-type PHD2 causes a dose-dependent suppression of HIF2-mediated induction of the reporter gene. When immunoblotted to monitor expression of the PHD2 protein, W334R and W389R proteins accumulated at levels below that of the wild-type protein for equivalent amounts of transfected plasmid (*data not shown*). To account for this effect, the amounts of transfected plasmid were adjusted to yield equivalent amounts of protein. We observed an inhibitory activity comparable to the wild-type for W334R, G349S and G349C variants, as well as for the previously published P200Q and R371H[26] variants. Indeed, only the activity of the W389R mutant was impaired (Figure 4A). To test the hypothesis of a potential effect on the kinetics of HIF-2 inhibition by the different PHD2 variants, we performed a real-time luciferase reporter assay using the Kronos system, which measures the activity of expressed luciferase instantaneously. We detected a severe impairment for the control mutants D254H and P317R, as described elsewhere,[9,26] and intermediate activity for the Y41C variant. All other PHD2 variants had a similar inhibitory effect to that of the wild-type protein (Figure 4B). Immunoblotting quantification showed that protein expression was reduced for two variants (I269T

A



B

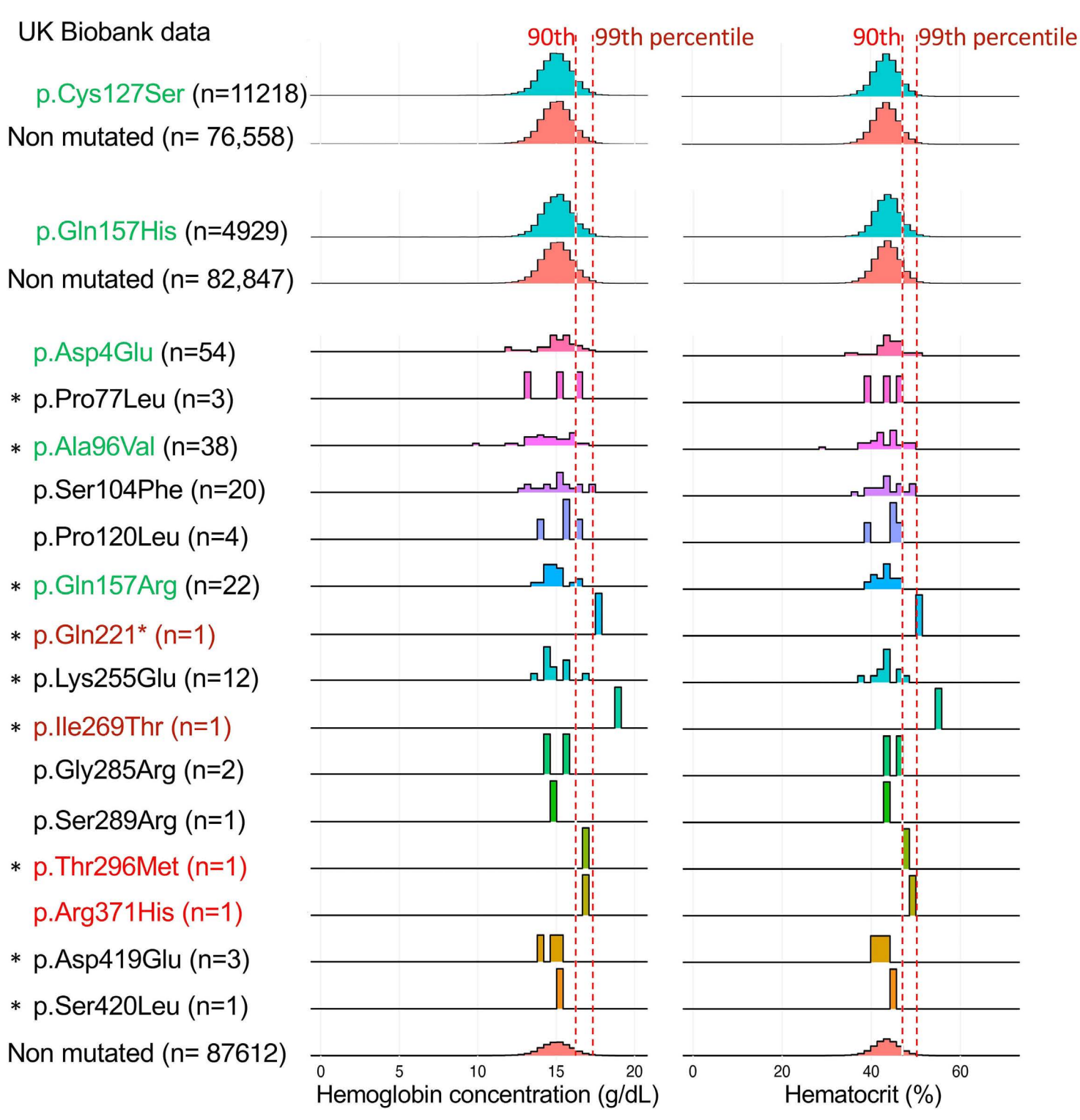
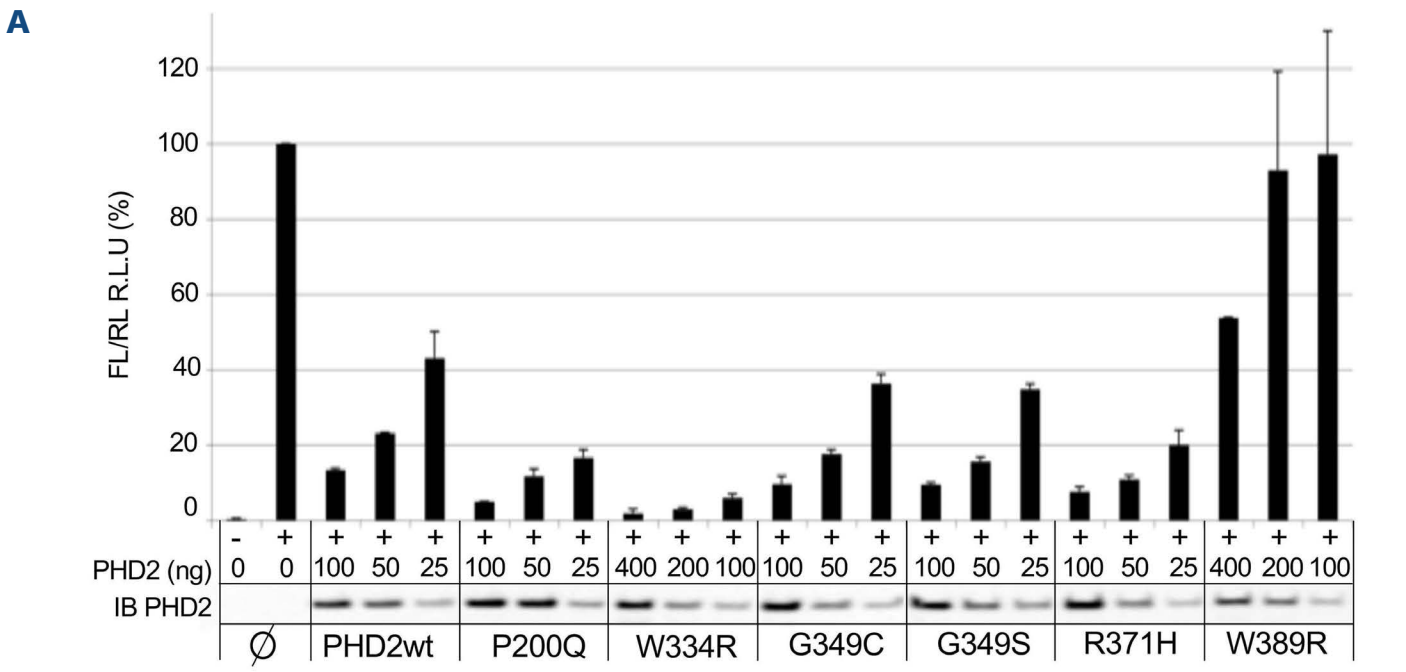


Figure 3. Scores from single and meta-predictors of studied variants and analysis of the UK Biobank data. (A) Representation of scores obtained by the *in silico* predictors analyzed by the MobiDetails annotation platform. Values are normalized (0-1), 0 being the least damaging and 1 the most for each predictor. The graph shows mean normalized scores obtained by single predictors (SIFT, Polyphen 2 HumDiv and HumVar) and meta-predictors (Fathmm, REVEL, ClinPred, Meta SVM, Meta LR, Mistic). For each variant, we analyzed the scores obtained with single and meta-predictors and classified the variants as benign when both scores were <0.4, and as deleterious when both scores were >0.6. (B) Analysis of hemoglobin and hematocrit levels in male carriers of *PHD2* variants identified in the UK Biobank. The colored bars represent cases. The red dotted lines represent the 90th and 99th percentile values of the control population. *Variants present in the UK Biobank also identified in the present study. n: number of cases.

and F366L), so the quantity of transfected plasmid was adjusted (Figure 4B and *data not shown*).

Study of enzymatic activity

The catalytic activity and kinetic properties of W334R, G349S, G349C, and R371H variants, expressed in insect cells and affinity-purified (*Online Supplementary Figure S5*), were determined by an enzymatic assay to detect the substrate hydroxylation-coupled decarboxylation of 2-oxglutarate compared with that of the wild-type PHD2. The results



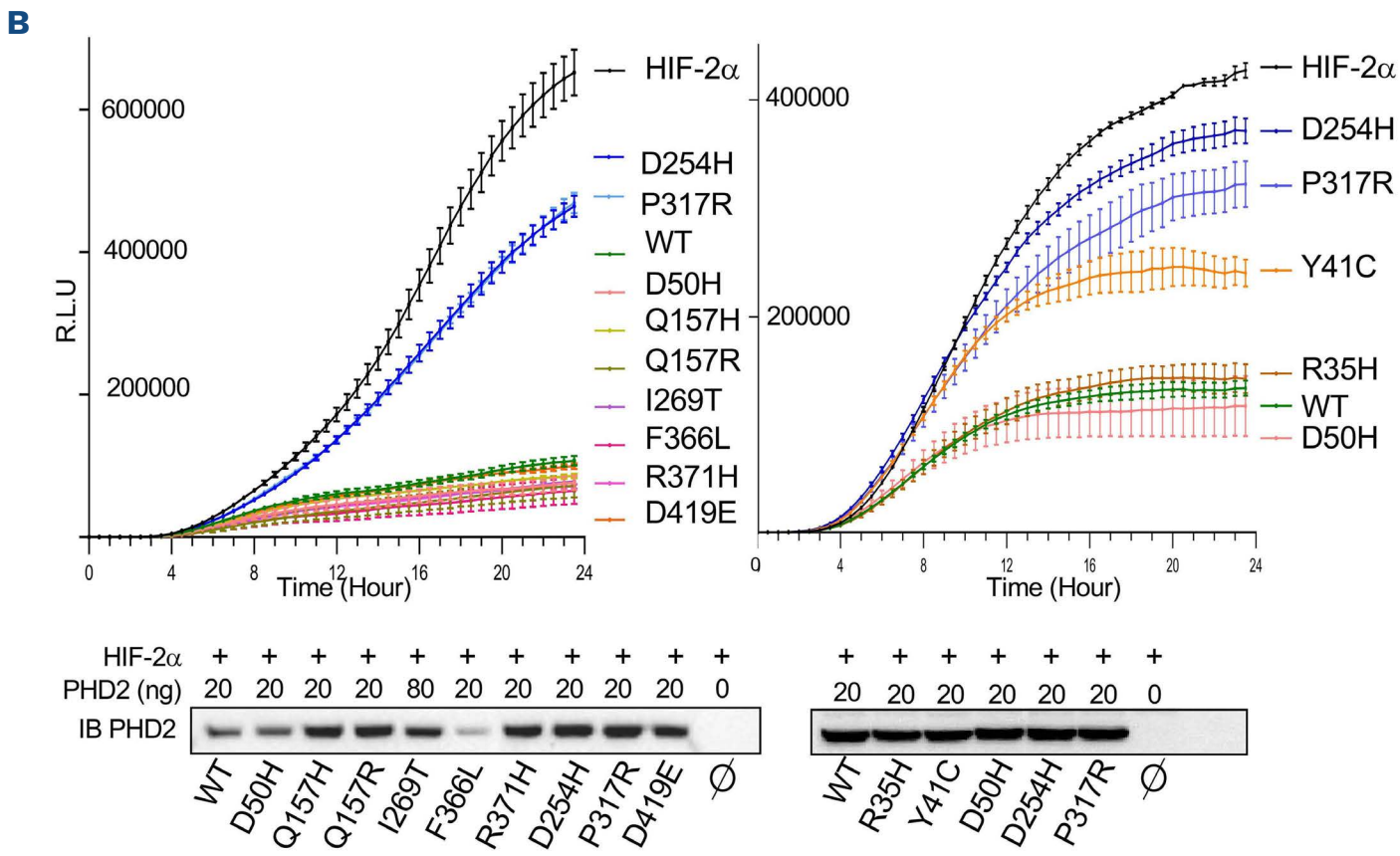


Figure 4. Functional study of *PHD2* mutants using luciferase reporter assays. (A) Functional study of *PHD2* mutants using end-point luciferase reporter assays. Cells were co-transfected with various amounts of PHD2 expression vectors (to enable the expression of the same amount of PHD2 proteins) in addition to HIF-2α expression vector, firefly luciferase reporter plasmid driven by hypoxia responsive elements and *Renilla* luciferase plasmid as a control of transfection efficiency. Luciferase activity was measured 24 h after transfection. Results are given as percentage of firefly luciferase activity normalized to *Renilla* luciferase activity. The amount of HA-PHD2 transfected (PHD2) was quantified by immunoblotting using anti-HA antibody. Results are mean values of experiments performed in triplicate. (B) Functional study of *PHD2* mutants using real time luciferase reporter assays. Cells were co-transfected with various amounts of PHD2 expression vectors in addition to HIF-2α expression vector and firefly luciferase reporter plasmids driven by hypoxia-responsive elements. Cells were incubated for 24 h in the bioluminometer Kronos HT® (ATTO) and luciferase activity was measured during 10 sec every 30 min. Results are given in relative light units (counts/10 sec). The amount of HA-PHD2 transfected was quantified by immunoblotting using an anti-HA antibody. Results are mean values of experiments performed in triplicate. FL: firefly luciferase; RL: *Renilla* luciferase; R.L.U.: relative light unit; IB PHD2: immunoblot of transfected HA-PHD2 by using anti-HA antibody; WT: wild-type; ∅: cells transfected with an empty pcDNA3 vector.

showed little difference between the variants and wild-type PHD2 through the K_m values for 2-oxoglutarate or two peptide substrates; a short synthetic HIF-1α peptide or the recombinant HIF-2α ODDD (Table 2). The G349S and G349C variants had higher K_m values for 2-oxo-glutarate compared with the wild-type protein, but these were probably compensated by their higher V_{max} values (Table 2). The V_{max} values of the R371H and W334R variants using either substrate were lower compared with those of the wild-type protein, but this could be, at least partially, compensated by their higher affinity (lower K_m values). However, a higher V_{max} value of the G349S variant was observed, especially with the HIF-1α substrate (Table 2).

Study of protein stability by a cycloheximide chase assay

The stability of PHD2 variants was assessed by measuring the kinetics of PHD2 expression after treatment with cycloheximide, an inhibitor of protein biosynthesis due to its suppressive effect on translational elongation. Plasmids expressing the HA-PHD2 mutants were transfected 24 h before cycloheximide chase and expression was measured at different timepoints (*Online Supplementary Figure S6A*). We noticed decreased expression 10 h after cycloheximide treatment for the variants I269T, W334R, R371H, and W389R. Quantification of replicates confirmed the significant loss of stability of these variants (Figure 5; *Online Supplementary Figure S6B*).

Study of the impact of variants on splicing using a minigene assay

Since some of the genetic variants studied do not seem to have a major impact on the function or the stability of PHD2 proteins, we attempted to test their impact on splicing. The potential effects of some *PHD2* variants on splicing were analyzed using Alamut Visual®, an exploration software application for genomic variation (*Online Supplementary Figure S7A*). These analyses showed a potential impact on splicing protein binding in all cases, except for the c.1165T>C, p.W389R, which was used as a control to study exon 4 splicing. *In silico* analysis using the SPiP site via MobiDetails, was more restrictive and showed only three deleterious splicing variants (c.891+1G>A, c.1152C>T, p.Y384Y, and c.1216+1G>T) (*Online Supplementary Figure S7B*). We first focused our study on variants located in exon 3 which are most likely to

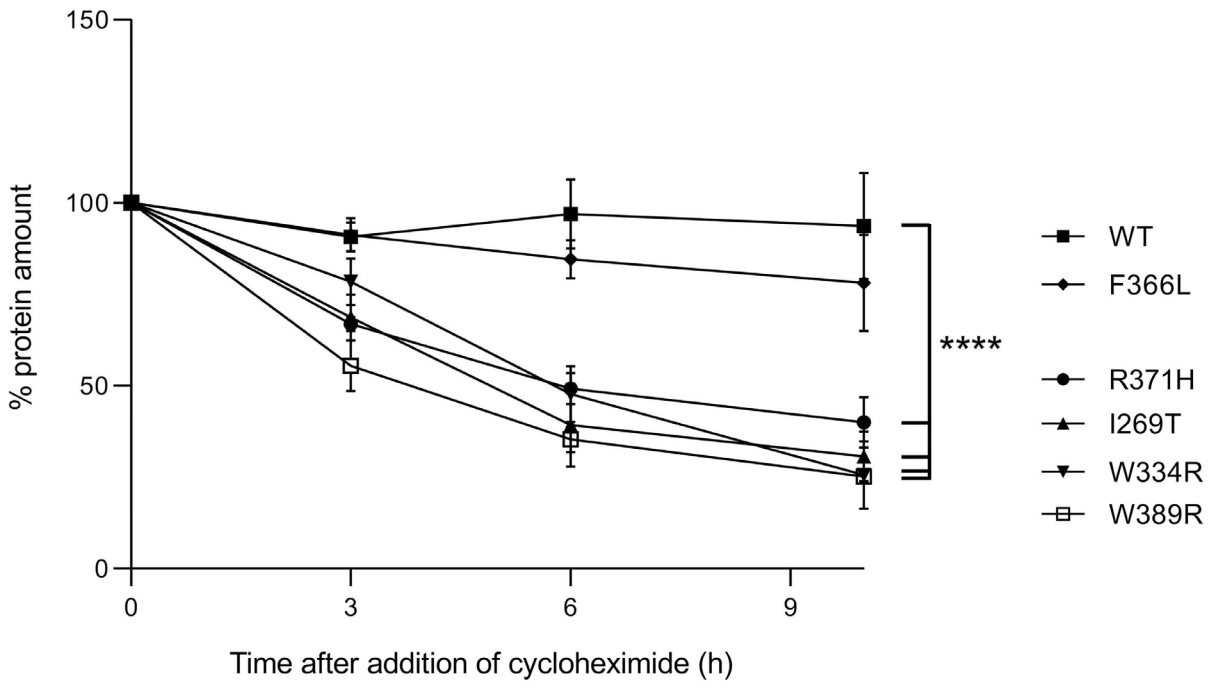


Figure 5. Study of PHD2 protein stability by the cycloheximide chase assay. The graph shows the amount of transfected HA-PHD2 protein after treatment with cycloheximide, reported as a percentage of the initial HA-PHD2 protein level (100% at 0 h of cycloheximide treatment) normalized to the intensity of actin. Data are shown as mean ± standard error of the mean of three independent experiments. Two-way analysis of variance was used for statistics (****$P \leq 0.0001$). WT: wild-type.

Table 2. K_m and V_{max} values of erythrocytosis-associated PHD2 mutants for 2-oxoglutarate and HIF-1α and HIF-2α substrates.

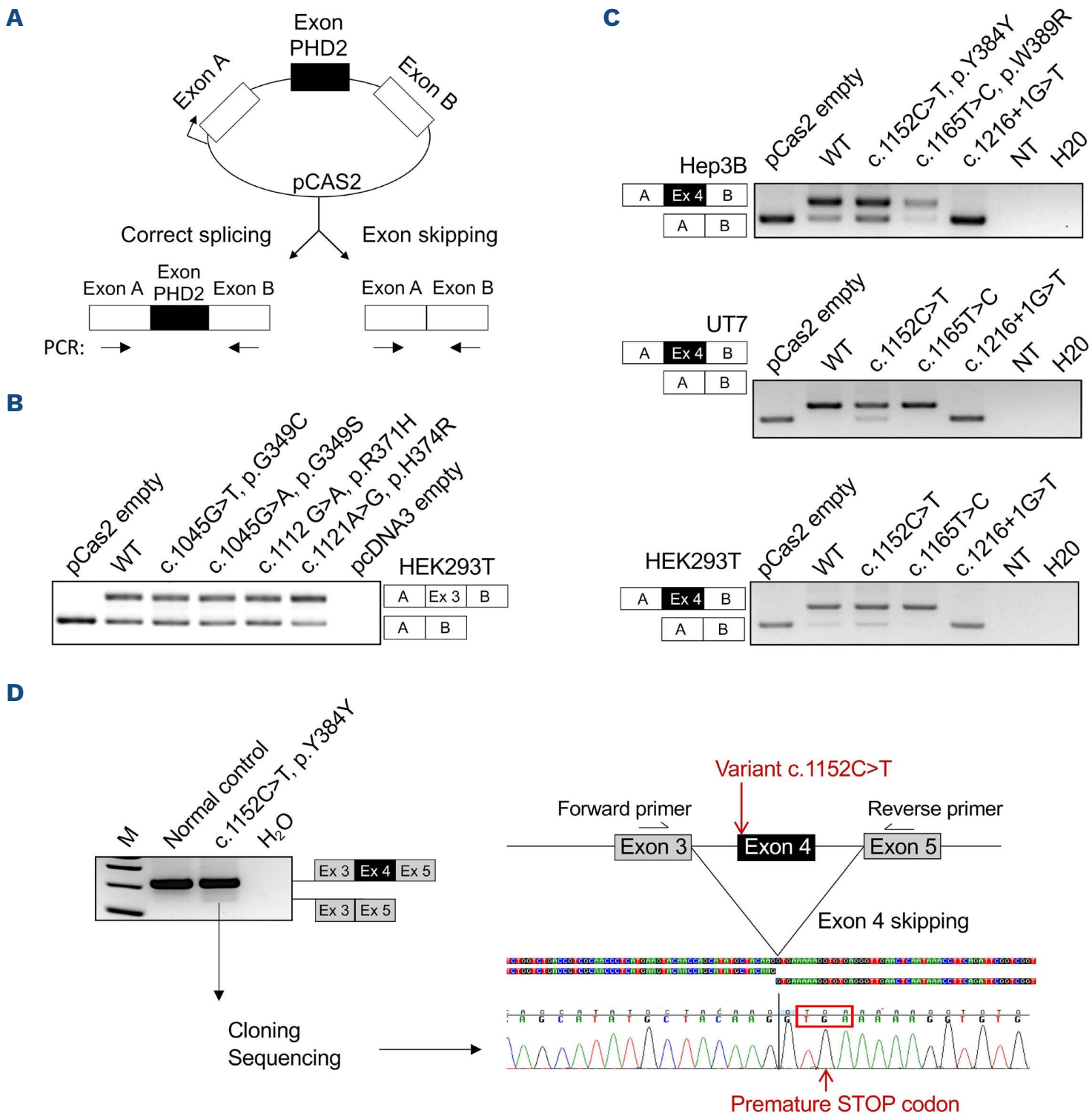
Parameter	Unit	Enzyme				
		WT PHD2	W334R	G349S	G349C	R371H
K_m of 2-oxoglutarate V_{max}	μM % of WT PHD2	5.0 ± 1.5 100	10.5 ± 4.0 100	9.5 ± 2.5 170	8.8 ± 2.9 160	4.8 ± 1.0 60
K_m of HIF-1α C-terminal peptide V_{max}	μM % of WT PHD2	11.5 ± 7.7 100	5.3 ± 1.8 50	14.5 ± 13.1 200	10.7 ± 6.4 110	3.9 ± 1.1 35
K_m of HIF-2α ODDD V_{max}	μM % of WT PHD2	0.27 ± 0.18 100	0.14 ± 0.11 70	0.24 ± 0.07 120	0.29 ± 0.16 90	0.15 ± 0.11 60

K_m and V_{max} values were determined for 2-oxoglutarate, HIF-1α C-terminal peptide (19 residues with Pro564) and for HIF-2α-ODDD (250 residues, contains both hydroxylatable prolines). WT: wild-type; PHD2: prolyl hydroxylase domain-containing protein 2; ODD: oxygen-dependent degradation domain.

be involved in the disease i.e., the same mutation described in two different families (c.1112 G>A, p.R371H)[26,27] or the same nucleotide and amino acid targeted in two different families (c.1045G>A, p.G349C and c.1045G>T, p.G349S). The mutant c.1121A>G, p.H374R, which completely inhibits the activity of PHD2, was used as a control.[20] The *PHD2* exon 3 and intronic flanking sequences were between the *SERPING1* exons (noted as A and B) of the splicing reporter plasmid pCAS2.[38] Mutagenesis was then performed to obtain mutant constructs (Figure 6A). HEK293T cells were transfected with the different minigene constructs. Spliced transcripts were detected by RT-PCR using primers located in exons A and B and analyzed on agarose gel. Two bands were present (Figure 6B) corresponding to transcripts containing exons A and B spliced with or without *PHD2* exon 3, demonstrating that wild-type *PHD2* exon 3 is not fully spliced in this cell type. No significant difference was observed for the different mutated minigene constructs, suggesting that the variants studied do not have an impact on splicing. We pursued the study with variants located in exon 4: c.1152C>T, p.Y384Y, and c.1216+1G>T in addition to c.1165T>C, p.W389R as a control. The results showed a deleterious effect of the c.1216+1G>T variant associated with a complete absence of *PHD2*-exon 4 inclusion in the three cell lines tested (Figure 6C). No impact on splicing was observed with c.1165T>C, p.W389R compared to the wild-type control, whereas the band corresponding to exon 4 skipping was slightly more highly expressed in the construct containing the synonymous variant c.1152C>T, p.Y384Y, in all the cell types tested (Figure 6C).

Study of the impact of variants on splicing using patients' cells

To confirm the observed *PHD2*-exon 4 splicing defect, we next focused our study on a patient's cells. RT-PCR using



Continued on following page.

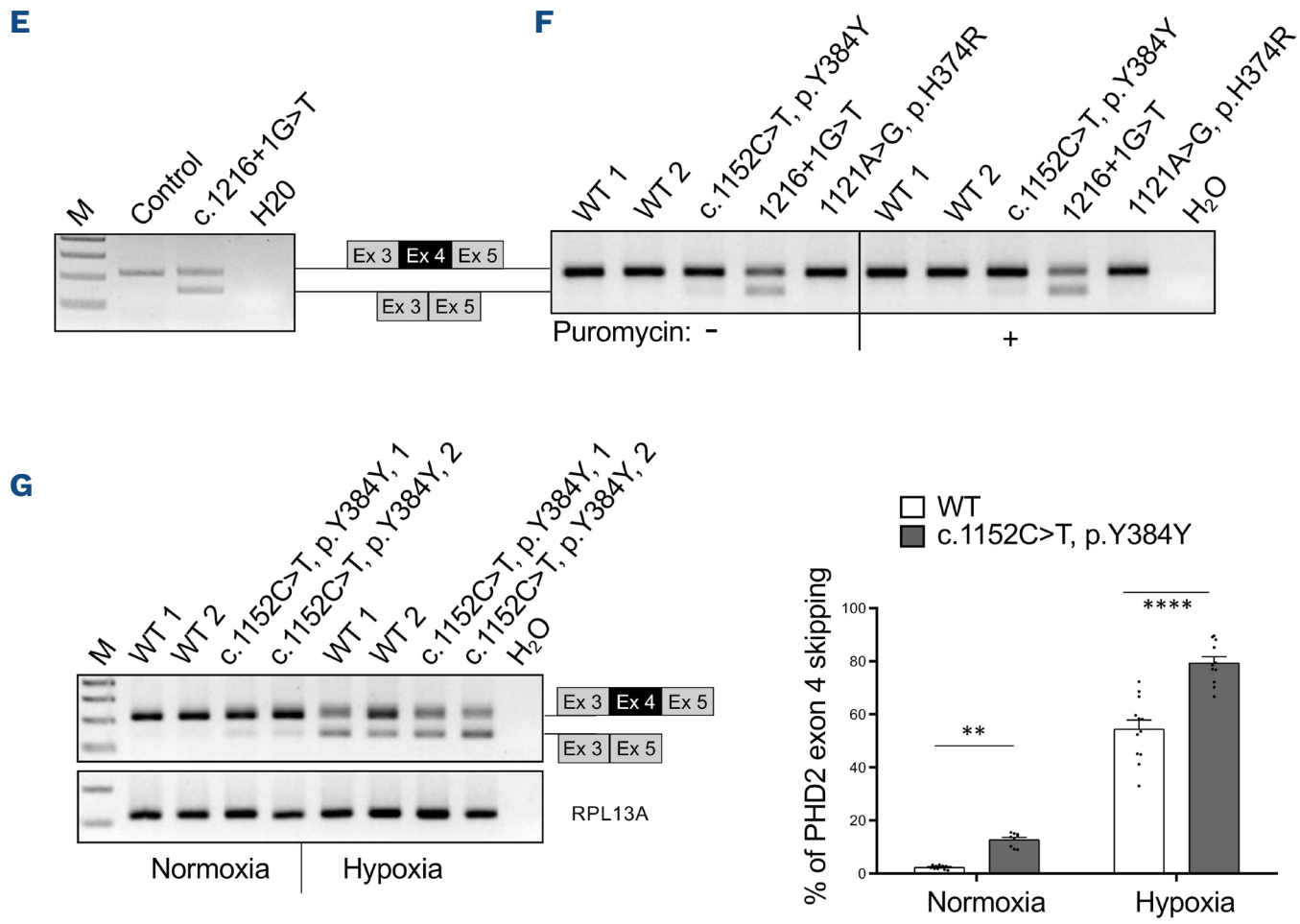


Figure 6. Functional studies of splicing variants. (A) Schematic representation of the splicing reporter assay (minigene experiment). The *PHD2* exon of interest and flanking intronic sequences were cloned in a minigene pCAS2 plasmid between the *SERPING1* exons (named A and B) flanked by short intronic sequences and conserved consensus splicing sequences. Mature and spliced mRNA were studied by reverse transcriptase polymerase chain reaction (RT-PCR) using primers located in exons A and B (schematized by arrows). (B) Characterization of *PHD2*-exon 3 splicing by a minigene experiment. RT-PCR was performed on mRNA obtained from HEK293T cells transfected with a minigene construct containing *PHD2*-exon 3 (wild-type or mutated) cloned in pCAS2. The pCAS2 plasmids were transfected, and the expression of the spliced chimeric transcripts was analyzed. Bands corresponding to exon A [A] and exon B [B] spliced together or with *PHD2* exon 3 (Ex3) are indicated on the right (representative picture of agarose gel; N=3). (C) Characterization of *PHD2*-exon 4 splicing by the minigene experiments. RT-PCR was performed on mRNA obtained from cell lines (HEK293T, Hep3B or UT-7) transfected with a minigene construct containing *PHD2*-exon 4 (wild-type or mutated) flanked by intronic sequences cloned into pCas2 plasmids. The plasmids were transfected, and the expression of the spliced chimeric transcripts was analyzed. Bands corresponding to exon A and exon B spliced together or with *PHD2*-exon 4 (Ex4) are indicated on the left (representative picture of agarose gel, N=3). (D) Study of endogenous *PHD2* splicing in a patient's cells. RT-PCR using primers located in exons 3 and 5 of the *PHD2* gene was performed on mRNA extracted from whole blood cells of the patient carrying the c.1152C>T, p.Y384Y variant collected into Paxgene® tubes. The lower band was purified and sequenced. The sequencing chromatogram is presented and shows the sequence of a transcript containing the exon 3 spliced with exon 5. (E) RT-PCR was performed on *PHD2* mRNA (exons 3-5) extracted from peripheral blood mononuclear cells of the patient carrying the c.1216+1G>T variant. (F) RT-PCR was performed on *PHD2* mRNA (exons 3-5) extracted from lymphoblastoid cell lines established from different patients. The cells were cultured in the absence (-) or presence (+) of puromycin, an inhibitor of nonsense-mediated mRNA decay mechanisms (representative picture of agarose gel, N=3). (G) Study of human induced pluripotent stem cells established from the patient carrying the c.1152C>T, p.Y384Y variant and differentiated into the hepatocyte-like cells. A representative gel of RT-PCR performed on *PHD2* mRNA (exons 3-5) is shown on the left. The quantification of the percentage of *PHD2*-exon 4 skipping in all replicates was performed using the TapeStation® migration system and the results are shown on the right. Each column represents the mean ± standard error of the mean of independent experiments (see details in *Online Supplementary Figure S8B*). Two-way analysis of variance was used for statistics (**$P \leq 0.01$, ****$P \leq 0.0001$). M: molecular-weight size marker; WT: wild-type.

primers in exons 3 and 5 of the endogenous *PHD2* mRNA was performed on whole blood cells obtained from the patient carrying the c.1152C>T, p.Y384Y variant; a barely detectable lower band was visible in the patient's sample (Figure 6D). Purification, cloning, and sequencing identified the band as a *PHD2* transcript containing exon 3 spliced with exon 5, thus confirming exon 4 skipping. We confirmed severe exon 4 skipping on peripheral blood mononuclear cells obtained from the patient carrying the heterozygous c.1216+1G>T mutation (equal intensity of the two bands corresponding to the spliced and skipped exon 4) (Figure 6E). Theoretically, translation of these mis-spliced transcripts introduces, in frame with the *PHD2* coding sequence, a premature stop codon at the beginning of exon 5 (Figure 6D, right panel). Although the premature stop codon is located in the last exon, we opted to assess whether these transcripts could be the target of nonsense-mediated mRNA decay. To this purpose, we established lymphoblastoid cell lines from

blood cells of the two patients and cultured them in the absence or presence of puromycin, an inhibitor of nonsense-mediated mRNA decay. The results obtained with these cells confirmed a severe effect of the c.1216+1G>T mutant and a very weak effect of the c.1152C>T, p.Y384Y variant on splicing (Figure 6F). No effect of puromycin was observed, indicating that the mis-spliced transcripts were not degraded by nonsense-mediated mRNA decay.

Because splicing is cell type-dependent and the effect of the c.1152C>T, p.Y384Y variant appeared to be more important, in minigene experiments, in the hepatocarcinoma cell line Hep3B (Figure 6C), we used a cellular model that mimics hepatocytes. We generated hiPSC from the patient's peripheral blood mononuclear cells which were differentiated into hepatocyte-like cells (*Online Supplementary Figure S8A*). At the end of the differentiation (day 22), hepatocyte-like cells were cultured in 1% oxygen for 24 h. RT-PCR was performed to investigate *PHD2*-exon 4 skipping (Figure 6G, left panel; *Online Supplementary Figure S8B*). PCR products were quantified after their migration on TapeStation® (Figure 6G, right panel). Our results demonstrated a significant effect of the c.1152C>T, p.Y384Y variant on *PHD2*-exon 4 skipping in normoxia and hypoxia (at 1% oxygen). Remarkably, in control and mutant cells, we noted that hypoxia has a major impact on *PHD2* splicing, with upregulation of the transcripts that do not retain exon 4.

Compiled analysis of data for variant classification

We interpreted and classified the identified variants according to American College of Medical Genetics (ACMG) criteria[41] and the recommendations of the French NGS group, based on genetic data (segregation in family, number of unrelated families, etc.), and the *in silico* and functional studies described here.

In this study, 16 *EGLN1* variants (in 24 patients) were classified as likely pathogenic or pathogenic, and 23 variants (in 23 patients) were classified as being of unknown significance (VUS) or likely benign. When comparing these two groups, there was no difference in the occurrence of thrombosis, either in the past history or after the diagnosis (4/24 [16.6%] *vs.* 2/24 [8.3%], respectively). Patients with a pathogenic variant were more frequently treated with phlebotomy (9/24 [37.5%] *vs.* 3/23 [13%]) and low-dose aspirin (10/24 [41.6%] *vs.* 2/23 8.7%]), whereas no significant difference was noted between these two groups of patients with respect to complete blood count values, including hemoglobin and hematocrit values (*Online Supplementary Table S2*). Surprisingly, two patients with a pathogenic *EGLN1* variant were treated with cytoreductive drugs (interferon and hydroxyurea). A family history of erythrocytosis was found in a higher proportion of patients with a pathogenic variant (41.6%) compared to patients with a VUS (26%).

We extended the *in silico* analysis to the genetic variants identified in the literature and compiled all these results with those of our patients (*Online Supplementary Table S3*). A total of 96 *PHD2* variants were identified, of which 36 can be classified as pathogenic (or probably pathogenic), five as benign, and 55 are still of unknown significance. As in our study, no difference was observed between the pathogenic/not pathogenic categories regarding hematologic parameters (*Online Supplementary Table S4*).

Discussion

The development of next-generation sequencing panels for use in diagnostic laboratories allowed molecular screening of a larger number of patients and increased the number of VUS identified. Mutations in the *EGLN1* gene are distributed throughout the protein with no hotspot, and no more than two families, of small size, per mutation have been described.[31] In addition, mutations in hypoxia pathway genes associated with erythrocytosis may be hypomorphic.[42] All these limitations make the classification of the identified variants difficult. The aim of this study was to bring together specialized genetic laboratories dedicated to genetic screening of mutations in *EGLN1* and to propose a comprehensive approach to *in silico* and functional analyses in order to collect information and facilitate genetic diagnosis.

Among the 2,160 patients sequenced by ten laboratories, we identified 39 genetic variants, including a complete heterozygous deletion of the gene, in 47 families. Our compiled *in silico* and functional studies allowed the classification of 16 variants as likely pathogenic/pathogenic mutations in 48 patients (24 probands and 24 relatives). The analysis extended to all variants in the literature showed that of the 96 PHD2 variants identified so far (including those in our study), 36 can be classified as pathogenic, five as benign and 55 are, in fact, of unknown significance (VUS). Although there are still many variants to be classified, the pooling of results from all our laboratories has been very beneficial. It has allowed the identification of families carrying similar mutations, which enabled the addition of ACMG classification criteria (such as PS4) and their classification into pathogenic mutations (I269T, R312H). In addition, the study of databases such as the UK Biobank has been very useful for the classification of benign variants (A96V, C127S, Q157H, Q157R).

Examination of the clinical data showed that complications are rare in patients carrying *EGLN1* mutations (in the present study, only 1 patient with portal vein thrombosis). Notably, clinical data from all the families described in the literature and in the present study (N=72 and N=47, respectively) showed that tumor development can be considered a very rare event (only 2 cases of paraganglioma[20,21] and 1 case of pheochromocytoma[42]) and may, therefore, be associated with additional genetic events. However, given the demonstrated role of PHD2 in the pathogenesis of pseudo-hypoxic

pheochromocytoma,[43] it is essential to medically monitor the occurrence of a possible tumor.

In this study there was no clear distinction in clinical or biological presentations between patients carrying *PHD2* variants classified as pathogenic compared to those carrying VUS. This means that clinical parameters cannot help in the classification. Nevertheless, our data are retrospective, and in a certain number of cases, we were unable to obtain precise dates of the biological results, i.e., before any treatment, phlebotomies or during evolution, which are parameters that can modify the hematologic data. In addition, the number of variants classified as pathogenic according to ACMG criteria may be underestimated because a family history was found for a significant number of VUS (26%) but it was not possible to explore their segregation in the family. Further exploration in the families would allow the variants to be shifted to the pathogenic class or to open research into other causal genetic events.

This study confirmed that erythrocytosis is more frequent in men than in women. In women, the disease could go undetected, mainly due to blood loss associated with menstruation. The example of the man carrying the W389R mutant (Family #42, P #III-4) who has no symptoms but is an assiduous blood donor supports this hypothesis (Figure 2A).

The functional studies using endpoint and real-time luciferase assays showed loss of function for only two mutants, Y41C and W389R. Development of the luciferase assays under even more sensitive conditions (reducing the amount of vectors transfected, using *PHD2* knock-out cells, etc.) did not improve the results (*data not shown*). We therefore used more sensitive *in vitro* enzymatic assays to evaluate the impact of some variants on PHD2 catalytic function. Unfortunately, these assays were not more informative. Measurement of K_m and V_{max} values did not demonstrate any significant loss of function of the tested variants. Indeed, the higher K_m value (showing a lower affinity for the substrate or co-substrate) measured for the W334R, G349S, and G349C variants was compensated by a higher V_{max} (speed of reaction) value. The same difficulty in demonstrating a loss of function was encountered with the R371H variant. The experiments showed a lower V_{max} value compensated by a substrate with increased binding affinity (lower K_m), confirming the results of previous studies.[23,25]

The examination of a loss of protein stability was more conclusive. The cycloheximide chase assay showed a decrease in the half-life of four variants: I269T, W334R, R371H, and W389R. Alteration of protein stability may theoretically have some impact on HIF activity, but was detected in our luciferase assays only for the W389R variant (for an equivalent amount of protein expressed).

The absence of significant detectable PHD2 loss of function or protein stability suggests that other PHD2 partners need to be tested. Indeed, PHD2 binds a number of proteins, and the mutations could conceivably affect these interactions.[44-46] For example, PHD2 binds FKPB38, which plays a major role in PHD2 stability;[47] LIMD1 is known to form a complex with PHD2 and VHL, creating an enzymatic niche that enables efficient degradation of HIF-1;[48] and PHD2 binds p23, which allows recruitment of PHD2 to the heat shock protein 90 (HSP90) machinery to facilitate hydroxylation of HIF-1α.[44] Of note, three variants in our series (R35H, Y41C and D50H), which are are located in the zinc finger domain, are involved in the stability of the PHD2/p23/HSP90/HIF-α complex. Nonetheless, none of these variants targets a conserved cysteine that plays a major role in zinc chelating cysteine residues, and only the Y41C variant showed a decreased ability to downregulate HIF in our luciferase reporter assay. Interestingly, recent work has identified a new partner of the PHD2 zinc finger domain that binds to the ribosomal chaperone NACA, allowing PHD2 to co-translationally modify the nascent HIF-α polypeptide.[49] More specific assays need to be developed in order to be able to investigate all these complex interactions.

To test the hypothesis of a potential impact of variants on splicing, which has been already demonstrated for other hypoxia pathway genes involved in erythrocytosis,[50] we performed splicing reporter assays. No impact was detected for the variants located in exon 3 (G349C, G349S, R371H), confirming the *in silico* prediction of the SPiP tool rather than that of the ALAMUT site which suggested a possible impact on the binding of the spliceosome proteins. Interesting results were obtained for exon 4 splicing. For the c.1216+1G>T variant targeting the splicing donor site, a deleterious effect with a high level of exon 4 skipping was detected, in all cells tested. Exon 4 skipping results in a transcript that contains an in-frame translation termination codon introduced by exon 5. We demonstrated that this transcript is not targeted by the nonsense-mediated decay machinery, which can, in consequence, lead to the expression of a truncated PHD2 protein (starting at amino acid 382). We have previously shown that a PHD2 protein truncated from amino acid 398 completely loses its function.[26] Thus, we may conclude that exon 4 skipping is equivalent to loss of *PHD2* function. A more subtle effect was detected for the synonymous c.1152C>T (Y384Y) located near the 5' end of exon 4 of *PHD2*. Because splicing is cell type-specific, we developed a cellular model using hiPSC derived from the patient and differentiated into hepatocyte-like cells. All experiments showed a slight but readily reproducible impact of the variant on exon 4 skipping. Interestingly, our experiments performed in hiPSC cells confirmed that hypoxia upregulates the expression of *PHD2*, a described HIF target gene, but also showed for the first time that hypoxia increases exon 4 skipping. Since we know that the exon 4 skipping results in a non-functional PHD2, the described negative feedback loop of hypoxia-induced *PHD2* might be attenuated because the expressed PHD2 is not fully effective. This result paves a new way to fine-tune the hypoxia pathway which regulates expression and splicing of its different players.

In conclusion, our study provides an example of a comprehensive approach to classification of genetic variants that could be applied to many rare hematologic diseases. Here, we demonstrated the importance of collaborating and combining results from different diagnostic centers dedicated to rare diseases. *In silico* studies with analysis of databases such as the UK Biobank may also facilitate classification, especially of non-pathogenic variants. We have shown the importance of performing a wide range of functional analyses and of studying finer regulatory mechanisms of the gene, such as the splicing studied here, whose complexity remains to be explored.

Finally, accurate classification is essential in order to make the most appropriate diagnosis in patients and thus ensure proper follow-up and treatment. In the absence of targeted or specific treatment, by analogy with polycythemia vera, phlebotomies and the use of low-dose aspirin, have been suggested,[51] albeit still under debate.[52] The recent encouraging results obtained with an HIF-2α inhibitor in *VHL*- and HIF-2α-associated diseases[53,54] opens up new promising therapeutic perspectives in all disorders associated with the hypoxia pathway.

Disclosures

No conflicts of interest to disclose.

Contributions

MD, ALR, MP, LS, MM, VK, DH, EP, VL, PK, and BG designed and/or performed in vitro and cellular functional experiments. MD, AC, KS-T, CC, and AG worked on hiPSC reprogramming and differentiation. AR performed the bioinformatics analyses. CG, NM, BA, MC, FA, LM, MR, TH, MM, SB, AB, NC, PH, CR, MW, FGa, BC, BB, CB, RvW, PP, MLR, MFMcM, FGi, and the ECYT3 consortium conducted the medical and genetic diagnostic studies. BG and FGi wrote the manuscript and designed the study. SI corrected the manuscript. BG directed the study. All authors contributed to the research and approved the final version of the manuscript.

Acknowledgments

The present research has been conducted using the UK Biobank resource under application number 49823. We are most grateful to the Bioinformatics Core Facility of Nantes BiRD, member of Biogenouest, Institut Français de Bioinformatique (IFB) (ANR-11-INBS-0013) for the use of its resources and for its technical support. We thank Thomas Besnard for minigene expertise, Morgane Taligot for technical assistance and Dr Emmanuelle Verger for expertise in next-generation sequencing analysis.

Funding

This study was supported by grants from the Région des Pays de la Loire (project "EryCan"); the ANR (PRTS 2015 "GenRED" and AAPG 2020 "SplicHypoxia"); the labex GR-Ex, reference ANR-11-LABX-0051; Fonds Européen de Développement Régional (FEDER) Bourgogne Franche Comté; the VHL Alliance USA, the VHL France; the Génavie association and the Fondation Maladies Rares (FMR).

Data-sharing statement

Data and detailed information related to the study are available from the corresponding author upon request.

Appendix: ECYT-3 members

Annalisa Andreoli, Emmanuel Bachy, Sarah Bonnet, Françoise Boyer, Serge Carillo, Brieuc Cherel, Florian Chevillon, Nataša Debeljak, Justine Decroocq, Roxana Dragan, Martine Escoffre, Jonathan Farhi, Arnaud Hot, Ludovic Karkowski, Catherine Humbrecht-Kraut, Philippe Joly, Jean-Jacques Kiladjian, Adrienne de Labarthe, Franck Lellouche, Guy Leverger, Emmanuel Raffoux, Dana Ranta, Benoit de Ranzis and Aline Schmidt.

References

1. Ivan M, Kondo K, Yang H, et al. HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. Science. 2001;292(5516):464-468.
2. Jaakkola P, Mole DR, Tian YM, et al. Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. Science. 2001;292(5516):468-472.
3. Epstein AC, Gleadle JM, McNeill LA, et al. C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. Cell. 2001;107(1):43-54.
4. Wenger RH, Stiehl DP, Camenisch G. Integration of oxygen signaling at the consensus HRE. Sci STKE. 2005;2005(306):re12.
5. D'Angelo G, Duplan E, Boyer N, Vigne P, Frelin C. Hypoxia up-regulates prolyl hydroxylase activity: a feedback mechanism that limits HIF-1 responses during reoxygenation. J Biol Chem. 2003;278(40):38183-38187.
6. Metzen E, Stiehl DP, Doege K, Marxsen JH, Hellwig-Bürgel T, Jelkmann W. Regulation of the prolyl hydroxylase domain protein 2 (phd2/egln-1) gene: identification of a functional hypoxia-responsive element. Biochem J. 2005;387(Pt 3):711-717.
7. Stiehl DP, Wirthner R, Köditz J, Spielmann P, Camenisch G, Wenger RH. Increased prolyl 4-hydroxylase domain proteins compensate for decreased oxygen levels. Evidence for an autoregulatory oxygen-sensing system. J Biol Chem. 2006;281(33):23482-23491.
8. Berra E, Benizri E, Ginouves A, Volmat V, Roux D, Pouyssegur J. HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1alpha in normoxia. Embo J. 2003;22(16):4082-4090.
9. Percy MJ, Zhao Q, Flores A, et al. A family with erythrocytosis establishes a role for prolyl hydroxylase domain protein 2 in oxygen homeostasis. Proc Natl Acad Sci U S A. 2006;103(3):654-659.
10. Pearson TC, Guthrie DL, Simpson J, et al. Interpretation of

measured red cell mass and plasma volume in adults: Expert Panel on radionuclides of the International Council for Standardization in Haematology. Br J Haematol. 1995;89(4):748-756.
11. Arsenault PR, Pei F, Lee R, et al. A knock-in mouse model of human PHD2 gene-associated erythrocytosis establishes a haploinsufficiency mechanism. J Biol Chem. 2013;288(47):33571-33584.
12. Xiang K, Ouzhuluobu, Peng Y, et al. Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. Mol Biol Evol. 2013;30(8):1889-1898.
13. Huerta-Sánchez E, Jin X, Asan, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 2014;512(7513):194-197.
14. Song D, Li LS, Arsenault PR, et al. Defective Tibetan PHD2 binding to p23 links high altitude adaptation to altered oxygen sensing. J Biol Chem. 2014;289(21):14656-14665.
15. Lorenzo FR, Huff C, Myllymäki M, et al. A genetic mechanism for Tibetan high-altitude adaptation. Nat Genet. 2014;46(9):951-956.
16. Gardie B, Percy MJ, Hoogewijs D, et al. The role of PHD2 mutations in the pathogenesis of erythrocytosis. Hypoxia (Auckl). 2014;2:71-90.
17. 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.
18. Lappin TR, Lee FS. Update on mutations in the HIF: EPO pathway and their role in erythrocytosis. Blood Rev. 2019;37:100590.
19. Sinnema M, Song D, Guan W, et al. Loss-of-function zinc finger mutation in the EGLN1 gene associated with erythrocytosis. Blood. 2018;132(13):1455-1458.
20. Ladroue C, Carcenac R, Leporrier M, et al. PHD2 mutation and congenital erythrocytosis with paraganglioma. N Engl J Med. 2008;359(25):2685-2692.
21. Yang C, Zhuang Z, Fliedner SM, et al. Germ-line PHD1 and PHD2 mutations detected in patients with pheochromocytoma/paraganglioma-polycythemia. J Mol Med (Berl). 2015;93(1):93-104.
22. Provenzano A, Chetta M, De Filpo G, et al. Novel germline PHD2 Variant in a metastatic pheochromocytoma and chronic myeloid leukemia, but in the absence of polycythemia. Medicina (Kaunas). 2022;58(8):1113.
23. Chowdhury R, Leung IKH, Tian Y-M, et al. Structural basis for oxygen degradation domain selectivity of the HIF prolyl hydroxylases. Nat Commun. 2016;7(1):12673.
24. Moore JA, Hubbi ME, Wang C, et al. Isolated erythrocytosis associated with 3 novel missense mutations in the EGLN1 gene. J Investig Med High Impact Case Rep. 2020;8:2324709620947256.
25. Pappalardi MB, Martin JD, Jiang Y, et al. Biochemical characterization of human prolyl hydroxylase domain protein 2 variants associated with erythrocytosis. Biochemistry. 2008;47(43):11165-11167.
26. 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.
27. Percy MJ, Furlow PW, Beer PA, Lappin TR, McMullin MF, Lee FS. A novel erythrocytosis-associated PHD2 mutation suggests the location of a HIF binding groove. Blood. 2007;110(6):2193-2196.
28. Oliveira JL, Coon LM, Frederick LA, et al. Genotype–phenotype correlation of hereditary erythrocytosis mutations, a single center experience. Am J Hematol. 2018;93(8):1029-1041.
29. Bonnin A, Gardie B, Girodon F, et al. [A new case of rare erythrocytosis due to EGLN1 mutation with review of the literature]. Rev Med Interne. 2020;41(3):196-199.
30. Bento C, Almeida H, Maia TM, et al. Molecular study of congenital erythrocytosis in 70 unrelated patients revealed a potential causal mutation in less than half of the cases (where is/are the missing gene(s)?). Eur J Haematol. 2013;91(4):361-368.
31. Barradas J, Rodrigues CD, Ferreira G, et al. Congenital erythrocytosis – discover of a new mutation in the EGLN1 gene. Clin Case Rep. 2018;6(6):1109-1111.
32. Al-Sheikh M, Moradkhani K, Lopez M, Wajcman H, Prehu C. Disturbance in the HIF-1alpha pathway associated with erythrocytosis: further evidences brought by frameshift and nonsense mutations in the prolyl hydroxylase domain protein 2 (PHD2) gene. Blood Cells Mol Dis. 2008;40(2):160-165.
33. Ansar S, Malcolmson J, Farncombe KM, Yee K, Kim RH, Sibai H. Clinical implementation of genetic testing in adults for hereditary hematologic malignancy syndromes. Genet Med. 2022;24(11):2367-2379.
34. Baux D, Van Goethem C, Ardouin O, et al. MobiDetails: online DNA variants interpretation. Eur J Hum Genet. 2021;29(2):356-360.
35. MetaDome: pathogenicity analysis of genetic variants through aggregation of homologous human protein domains - Wiel - 2019 - Human Mutation - Wiley Online Library. https://onlinelibrary-wiley-com.proxy.insermbiblio.inist.fr/doi/10.1002/humu.23798 Accessed October 18, 2021.
36. Hirsilä M, Koivunen P, Xu L, Seeley T, Kivirikko KI, Myllyharju J. Effect of desferrioxamine and metals on the hydroxylases in the oxygen sensing pathway. FASEB J. 2005;19(10):1308-1310.
37. Koivunen P, Myllyharju J. Kinetic analysis of HIF prolyl hydroxylases. Methods Mol Biol. 2018;1742:15-25.
38. Gaildrat P, Killian A, Martins A, Tournier I, Frebourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. Methods Mol Biol. 2010;653:249-257.
39. DeLaForest A, Nagaoka M, Si-Tayeb K, et al. HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. Development. 2011;138(19):4143-4153.
40. Si-Tayeb K, Noto FK, Nagaoka M, et al. Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. Hepatology. 2010;51(1):297-305.
41. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-424.
42. Couvé S, Ladroue C, Laine E, et al. Genetic evidence of a precisely tuned dysregulation in the hypoxia signaling pathway during oncogenesis. Cancer Res. 2014;74(22):6554-6564.
43. Eckardt L, Prange-Barczynska M, Hodson EJ, et al. Developmental role of PHD2 in the pathogenesis of pseudohypoxic pheochromocytoma. Endocr Relat Cancer. 2021;28(12):757-772.
44. Song D, Li LS, Heaton-Johnson KJ, Arsenault PR, Master SR, Lee FS. Prolyl hydroxylase domain protein 2 (PHD2) binds a Pro-Xaa-Leu-Glu motif, linking it to the heat shock protein 90 pathway. J Biol Chem. 2013;288(14):9662-9674.
45. Vogel KS, Brannan CI, Jenkins NA, Copeland NG, Parada LF. Loss of neurofibromin results in neurotrophin-independent survival of embryonic sensory and sympathetic neurons. Cell. 1995;82(5):733-742.
46. Huo Z, Ye JC, Chen J, et al. Prolyl hydroxylase domain protein 2 regulates the intracellular cyclic AMP level in cardiomyocytes through its interaction with phosphodiesterase 4D. Biochem

Biophys Res Commun. 2012;427(1):73-79.
47. Barth S, Nesper J, Hasgall PA, et al. The peptidyl prolyl cis/trans isomerase FKBP38 determines hypoxia-inducible transcription factor prolyl-4-hydroxylase PHD2 protein stability. Mol Cell Biol. 2007;27(10):3758-3768.
48. Foxler DE, Bridge KS, James V, et al. The LIMD1 protein bridges an association between the prolyl hydroxylases and VHL to repress HIF-1 activity. Nat Cell Biol. 2012;14(2):201-208.
49. Song D, Peng K, Palmer BE, Lee FS. The ribosomal chaperone NACA recruits PHD2 to cotranslationally modify HIF-α. EMBO J. 2022;41(22):e112059.
50. 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.
51. McMullin MFF, Mead AJ, Ali S, et al. A guideline for the management of specific situations in polycythaemia vera and secondary erythrocytosis: a British Society for Haematology guideline. Br J Haematol. 2019;184(2):161-175.
52. Gordeuk VR, Miasnikova GY, Sergueeva AI, et al. Thrombotic risk in congenital erythrocytosis due to up-regulated hypoxia sensing is not associated with elevated hematocrit. Haematologica. 2020;105(3):e87-e90.
53. Kamihara J, Hamilton KV, Pollard JA, et al. Belzutifan, a potent HIF2α inhibitor, in the Pacak-Zhuang syndrome. N Engl J Med. 2021;385(22):2059-2065.
54. Yu Y, Yu Q, Zhang X. Allosteric inhibition of HIF-2α as a novel therapy for clear cell renal cell carcinoma. Drug Discov Today. 2019;24(12):2332-2340.