Comprehensive *in silico* and functional studies for classification of *EPAS1/HIF2A* genetic variants identified in patients with erythrocytosis

Valéna Karaghiannis,[1,2*] Darko Maric,[3,4*] Céline Garrec,[5] Nada Maaziz,[6] Alexandre Buffet,[7,8] Loïc Schmitt,[2] Vincent Antunes,[3,4] Fabrice Airaud,[5] Bernard Aral,[6] Amandine Le Roy,[2] Sébastien Corbineau,[2] Lamisse Mansour-Hendili,[9,10] Valentine Lesieur,[2] Antoine Rimbert,[2] Fabien Laporte,[2] Marine Delamare,[2] Minke Rab,[11,12] Stéphane Bézieau,[2,5] Bruno Cassinat,[13] Frédéric Galacteros,[10,14] Anne-Paule Gimenez-Roqueplo,[7,8] Nelly Burnichon,[7,8] Holger Cario,[15] Richard van Wijk,[11] Celeste Bento,[16] ECYT-4 consortium,° François Girodon,[6,17,18#] David Hoogewijs,[3,4#] and Betty Gardie[1,2,18#]

[1]Ecole Pratique des Hautes Etudes, EPHE, Université Paris Sciences et Lettres, Paris, France; [2]Nantes Université, CNRS, INSERM, l'Institut du Thorax, Nantes, France; [3]Section of Medicine, Department of Endocrinology, Metabolism and Cardiovascular System, University of Fribourg, Fribourg, Switzerland; [4]National Center of Competence in Research "Kidney.CH", Switzerland; [5]Service de Génétique Médicale, CHU de Nantes, Nantes, France; [6]Service d'Hématologie Biologique, Pôle Biologie, CHU de Dijon, Dijon, France; [7]Université Paris Cité, INSERM, PARCC, Paris, France; [8]Département de Médecine Génomique des Tumeurs et des Cancers, AP-HP, Hôpital Européen Georges Pompidou, Paris, France; [9]Département de Biochimie-Biologie Moléculaire, Pharmacologie, Génétique Médicale, AP-HP, Hôpitaux Universitaires Henri Mondor, Créteil, France; [10]Université Paris-Est Créteil, IMRB Equipe Pirenne, Laboratoire d'Excellence LABEX GRex, Créteil, France; [11]Central Diagnostic Laboratory - Research, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands; [12]Department of Hematology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands; [13]Université Paris Cité, APHP, Hôpital Saint-Louis, Laboratoire de Biologie Cellulaire, Paris, France; [14]Red Cell Disease Referral Center-UMGGR, AP-HP, Hôpitaux Universitaires Henri Mondor, Créteil, France; [15]Department of Pediatrics and Adolescent Medicine, University Medical Center, Ulm, Germany; [16]Hematology Department, Centro Hospitalar e Universitário de Coimbra, CIAS, University of Coimbra, Coimbra, Portugal; [17]Université de Bourgogne, INSERM U1231, Dijon, France and [18]Laboratoire d'Excellence GR-Ex, Paris, France

VK and DM contributed equally as co-first authors.
#FG, DH and BG contributed equally as co-senior authors.
°An appendix with all ECYT-4 consortium members can be found at the end of the manuscript.

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

Received: September 1, 2022.
Accepted: January 17, 2023.
Early view: January 26, 2023.

https://doi.org/10.3324/haematol.2022.281698

Abstract

Gain-of-function mutations in the *EPAS1/HIF2A* gene have been identified in patients with hereditary erythrocytosis that can be associated with the development of paraganglioma, pheochromocytoma and somatostatinoma. In the present study, we describe a unique European collection of 41 patients and 28 relatives diagnosed with an erythrocytosis associated with a germline genetic variant in *EPAS1*. In addition we identified two infants with severe erythrocytosis associated with a mosaic mutation present in less than 2% of the blood, one of whom later developed a paraganglioma. The aim of this study was to determine the causal role of these genetic variants, to establish pathogenicity, and to identify potential candidates eligible for the new hypoxia-inducible factor-2 α (HIF-2α) inhibitor treatment. Pathogenicity was predicted with *in silico* tools and the impact of 13 HIF-2β variants has been studied by using canonical and real-time reporter luciferase assays. These functional assays consisted of a novel edited vector containing an expanded region of the erythropoietin promoter combined with distal regulatory elements which substantially enhanced the HIF-2α-dependent induction. Altogether, our studies allowed the classification of 11 mutations as pathogenic in 17 patients and 23 relatives. We described four new mutations (D525G, L526F, G527K, A530S) close to the key proline P531, which broadens the spectrum of mutations involved in erythrocytosis. Notably, we identified patients with only erythrocytosis associated with germline mutations A530S and Y532C previously identified at somatic state in tumors, thereby raising the complexity of the genotype/phenotype correlations. Altogether, this study allows accurate clinical follow-up of patients and opens the possibility of benefiting from HIF-2α inhibitor treatment, so far the only targeted treatment in hypoxia-related erythrocytosis disease.

Introduction

Erythrocytoses are characterized by an elevated red cell mass of more than 125% of the predicted value for the age and body mass of the subject, usually reflected by increased levels of hemoglobin (Hb) and/or hematocrit (Ht) values.[1] Erythrocytoses can be acquired as in the primary polycythemia vera, a myeloproliferative neoplasm as a consequence of a gain of function mutation in the *JAK2* gene (p.Val617Phe), or secondary to diverse pathological situations (pulmonary or heart disease, kidney cancer, carbon monoxide poisoning, etc.) due to an increased secretion of erythropoietin (EPO). On the other hand, erythrocytosis can be observed in a context of inherited disease, which can be primary when there is an intrinsic defect in the progenitor cells of the bone marrow (*EPOR* mutations), or secondary when the oxygen-sensing pathway is dysregulated and EPO is produced at a high level (*VHL, EGLN1/PHD2, EPAS1/HIF2A* mutations).

This study focuses on the *EPAS1* gene that encodes the hypoxia-inducible factor-2 α (HIF-2α), a major player in the oxygen-sensing pathway, also known as the hypoxia pathway.

HIF is a transcription factor that is stabilized when the oxygen concentration is reduced.[2] HIF is a α/β hetero-dimer consisting of a tightly regulated oxygen-labile α-subunit and a constitutive β-subunit. The HIF-α subunits (HIF-1α, 2α, 3α) contain an oxygen-dependent degradation (ODD) domain, and two independent transcriptional activation domains. In the presence of oxygen, the HIF-α subunits are hydroxylated by prolyl hydroxylases (PHD1-4)[3] that regulate their stability and an asparaginyl hydroxylase factor inhibiting HIF (FIH) that regulates their transcriptional activity.[4] The PHD hydroxylate proline residues are located within the HIF-α ODD (P402 and P564 for HIF-1α; P405 and P531 for HIF-2α). This hydroxylation allows the binding of the von Hippel-Lindau (VHL) protein, a recognition subunit of an E3 ubiquitin ligase multiprotein complex. Binding of VHL to HIF-α subunits induces ubiquitination, which targets them for degradation by the proteasome. Under hypoxic conditions, when the co-factor oxygen is limiting, hydroxylation of HIF-α subunits slows down and results in its stabilization. HIF-α then translocates to the nucleus, associates with the HIF-1α subunit and, upon recruiting appropriate co-activators, the HIF-α/β heterodimer binds to hypoxia response elements (HRE) within DNA and activates expression of HIF target genes.[5] HIF regulate the transcription of more than 200 genes involved in many pathways, notably erythropoiesis *via* the synthesis of erythropoietin (EPO)[6] and iron metabolism regulation (tranferrin, transferrin receptor, divalent metal transporter 1 [*DMT1*], ferroportin). HIF-2 is the main isoform that controls EPO expression which regulates the proliferation and differentiation of erythroid progenitors, thereby linking decreased tissue oxygenation to an adequate erythropoietic response.[7]

HIF-2α is an 870 amino acid protein encoded by the *EPAS1* gene located in 2p21, that contains 16 exons. The first mutation in the *EPAS1* gene has been described in 2008 associated with polycythemia, also called erythocytosis (ECYT) developed by patients over three generations.[8] Since then, more than 40 other cases have been published (*Online Supplementary Table S1*). These are always germline missense mutations in the heterozygous state. Patients described with a mutation in the *EPAS1* gene developed erythrocytosis frequently associated with high EPO levels and are at increased risk of pulmonary hypertension and thrombotic events (thrombosis, infarction, pulmonary embolism).[8-11] This disease has been classified as ECYT4 (OMIM#603349, MIM#611783).

Importantly, mutations in *EPAS1/HIF2A* gene are also responsible for the development of pheochromocytomas, paragangliomas (PPGL), somatostatinomas and ocular lesions[12] sometimes associated with erythrocytosis and rare cases of cyanotic congenital heart disease and hemangioblastomas (*Online Supplementary Table S2*).[13,14,12,15] In these cases, the mutations have been found at the somatic level within the tumor, but further investigations sometimes revealed that the mutation was acquired during development and was actually present in a mosaic state.[16]

The majority of these missense mutations are located in exon 12 and cluster close to the proline residues Pro531, hydroxylation of which regulates HIF-2α stability. Combined biochemical and cellular assays showed that the majority of these mutations may reduce both hydroxylation of HIF-2α by the PHD, and subsequent recognition of HIF-2α by pVHL. Functional studies showed that mutations associated with tumor development have a more severe gain-of-function than mutations associated with erythrocytosis alone.[17] These most deleterious mutations (amino acids 529-532) affect residues close to the key proline 531.

In the present collaborative study, we describe a unique series of 43 patients diagnosed with an erythrocytosis in whom we identified 33 different genetic variants in *EPAS1*. The aim of this study is to determine the causal role of these genetic variants in the pathogenesis of erythrocytosis and the potential risk for developing tumors.

In order to study the functional consequences of the different HIF-2α mutations, we performed luciferase reporter assays. We used a vector encoding the firefly luciferase driven by multiple proximal and distal regulatory elements of the HIF-2 target gene *EPO*, the key target gene linked to the development of erythrocytosis. A recent publication suggested that the *EPO* gene may contain complex regulatory elements in its proximal promoter based on the identification of a mutation located in a region (c.-136, upstream the ATG codon[18]) not included in

previously existing luciferase constructs.[19,20] We therefore generated a luciferase vector driven by a substantially extended region of the *EPO* promoter and distal enhancer regions. In order to accurately quantify subtle changes in HIF-2α activity real-time periodical luminescence measurements were performed.

Methods

Sequencing

Samples were obtained from laboratories specialized in diagnosis and research of idiopathic erythrocytosis after exclusion of classical causes of erythrocytosis (polycythemia vera or secondary erythrocytosis associated with particular renal, cardiac or pulmonary disorders). After receiving approval from the Ethics Committee and obtaining written informed consent from the proband and family members, blood samples were collected and DNA was extracted for genetic analysis. Molecular screening of genes associated with erythrocytosis was performed by high throughput sequencing with different technologies, depending on the sequencing center (see the *Online Supplementary Appendix*).

In silico analysis

Genetic variants located in the *EPAS1* gene with a global frequency lower than 5.10^{-4} in the gnomAD v3 database were selected. The MetaDome[21] analysis was performed on the *EPAS1* gene by using the GENCODE: ENST00000263734.3, RefSeq: NM_001430.4 and UniProt: Q99814. The effect of each non-synonymous variant was assessed using Protein Variation Effect Analyzer (PROVEAN)21 Protein Batch v1.1.3 prediction integrative webtool. ENSEMBL ID HIF-2α wild-type amino acid sequence (ENSP00000263734) was used as a reference. Cutoff of -2.5 and -4.1 were applied to PROVEAN predictions. *EPAS1* variants were studied by using MobiDetails,[22] an annotation platform for DNA variants. Values obtained by single-and meta-predictor tools were normalized (0-1), 0 being the less damaging and 1 the most damaging for each predictor. For the final interpretation and classification of the variants, we used the ACMG (American College of Medical Genetics and Genomics) criteria and guidelines.[23] The ACMG uses the following classification to describe variants identified in Mendelian disorders: class 1: benign; class 2: likely benign; class 3: variant of uncertain significance (VUS); class 4: likely pathogenic; class 5: pathogenic. Criteria used in our study are detailed in the *Online Supplementary Appendix*.

Full erythropoietin promoter plasmid generation

The full human core *EPO* promoter sequence was recovered from ENSEMBL (7 dna:chromosome:GRCh38:7:100720400:100721001:1, ENSR00000833692). Briefly, the pGL3-5'HRE290-FullProm-3'HRE126 *EPO* promoter-driven luciferase plasmid was generated after polymerase chain reaction (PCR) amplification of a 604 bp fragment of genomic DNA extracted from HeLa cells that was inserted into a pGL3-luciferase vector (details are provided in the *Online Supplementary Appendix*).

End point luciferase reporter assays

Cells were cultured as detailed in the *Online Supplementary Appendix*. Briefly, $4x10^5$, $3.5x10^5$ and $6.5x10^5$ cells were transiently transfected with 500 ng or 100 ng reporter plasmid in a 6-well format using $CaCl_2$ or JetOptimus (Polyplus), respectively for HEK293T and Hep3B or Kelly cells, and 1 μg of YFP-HIF-1α, YFP-HIF-2α or pcDNA3-HA-HIF-2α (Addgene) constructions. In order to control for differences in transfection efficiency and extract preparation, 50 ng or 75 ng pRL-SV40 *Renilla* luciferase reporter vector (Promega) was co-transfected, respectively for HEK293T and Hep3B or Kelly cells. The next day, cultures were evenly split onto 6-well plates, incubated for an additional 24 hours, under normoxic or hypoxic conditions (0.2% O_2, 5% CO_2 and 37°C). Cells were lysed with passive lysis buffer and luciferase activities of duplicated wells were determined using the Dual Luciferase Reporter Assay System (Promega) as described before.[24] Reporter activities were expressed as relative Firefly/*Renilla* luciferase activities normalized to control under hypoxic conditions. All reporter gene assays were performed at least three times independently. Proteins were extracted and immunoblotted to quantify the HIF-2α proteins as described before[25] (see the *Online Supplementary Appendix*).

Real-time luciferase reporter assays

Luciferase assays were performed on HEK293T cells seeded in 24-well Black Visiplate Perkin Elmer) ($1×10^5$ cells per well), 24 hours before transfection. Cells were transfected by using jetPRIME® (Ozyme Polyplus). The expression vectors pcDNA3-HA-HIF-2α (25 ng) were co-transfected with the pGL3-5'HRE-FullProm-3'HRE-*EPO* promoter-driven luciferase plasmid (100 ng), and empty vector for a total amount of 500 ng transfected DNA. Luciferase activity was monitored over 48 hours using the bioluminometer WSL-1565 Kronos HT® (ATTO). Cells were harvested and lysed in extraction buffer (Macherey Nagel) for quantification of transfected plasmids by PCR (for details see the *Online Supplementary Appendix*).

Statistical analysis

Values in the figures of the end point luciferase assays are presented as mean ± standard error of the mean (SEM). For the real-time luciferase reporter assay, differences in means among multiple groups were analyzed by using one-way ANOVA of Kruskal-Wallis and Dunn's *post hoc* tests. All statistics were performed with GraphPad Prism

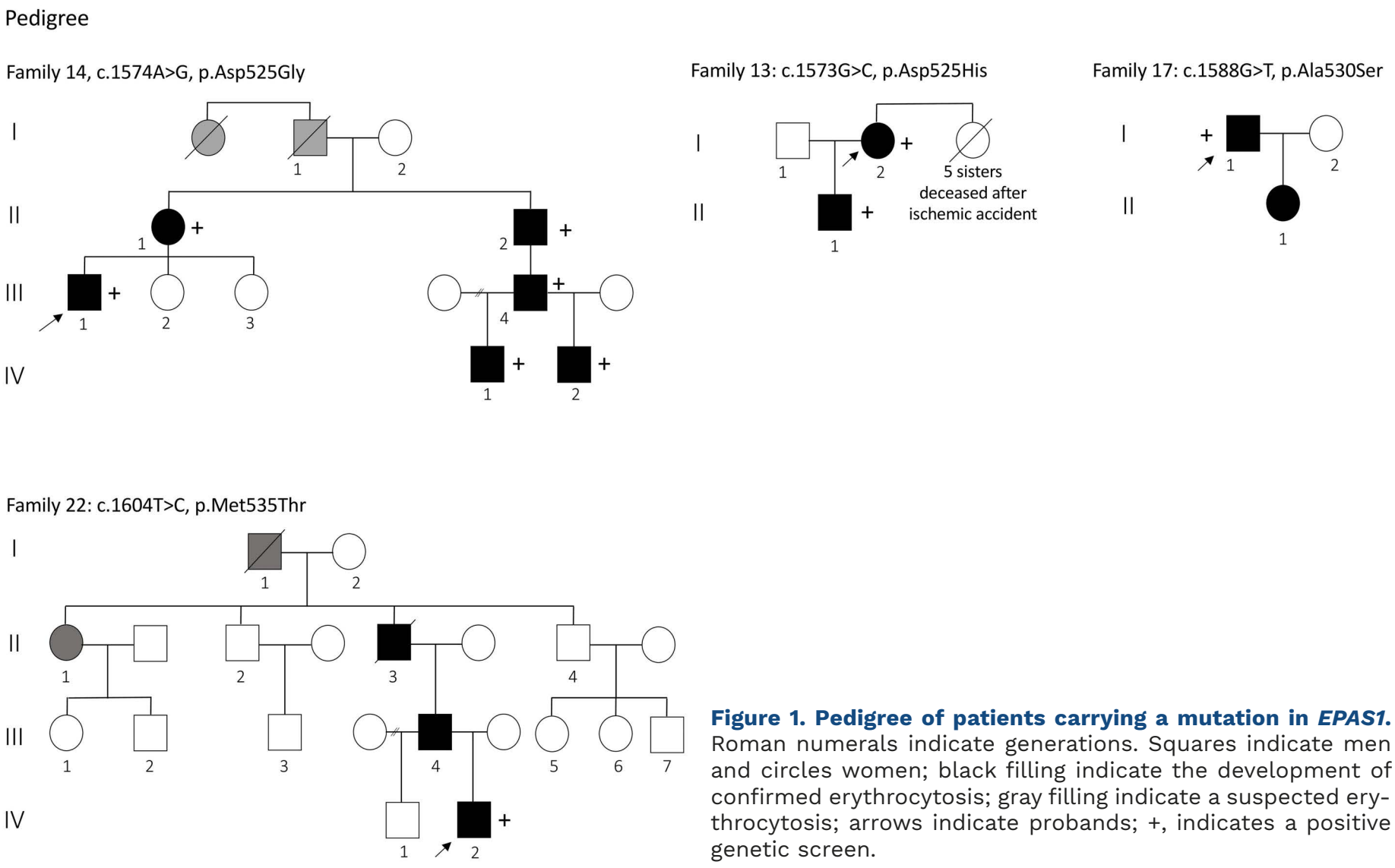

Figure 1. Pedigree of patients carrying a mutation in *EPAS1*. Roman numerals indicate generations. Squares indicate men and circles women; black filling indicate the development of confirmed erythrocytosis; gray filling indicate a suspected erythrocytosis; arrows indicate probands; +, indicates a positive genetic screen.

software 7.05. Values of $P \leq 0.05$ were considered statistically significant.

Results

Diagnosis and genetic screening

Patients with erythrocytosis were selected after exclusion of classical causes of erythrocytosis (polycythemia vera or secondary erythrocytosis associated with particular renal, cardiac or pulmonary disorders). Genetic screening was performed by using next-generation sequencing (NGS) panels dedicated to erythrocytosis, and a total of 1,450 patients were sequenced at all centers. Results are presented in Table 1 and show 33 missense genetic variants identified in a total of 43 patients (41 patients with heterozygous germline variant and 2 patients with mosaicism) and 28 relatives from four different European countries. Ten variants have already been described.[15,26-43] Twenty-four variants are located in exon 12 and extend from amino acids 525 to 658 surrounding the key proline in position 531. Family history of erythrocytosis is present in 14 families. The pedigrees are of variable size (Figure 1; *Online Supplementary Figure S1*) and can show a multigenerational history of erythrocytosis (family [F] 14, F22), but sometimes also display a family history related to the consequences of undiagnosed erythrocytosis, such as ischemic stroke (F13). Clinical manifestations are in general variable, ranging from mild to severe erythrocytosis (Ht up to 77.5% in patient 19) (Table 1 for probands; *Online Supplementary Table S3* for relatives). The serum EPO level is rarely elevated except in two patients (up to 7,500 UI/mL in patient 20). However, we were unable to obtain full information on the patient's phlebothomy time frame that may influence these factors.

Associated symptoms characteristic to *EPAS1* mutations have been observed in a limited number of patients. Pulmonary arterial hypertension (PAH) was developed by two patients and ocular lesions have been detected in three patients as previously reported in association with *EPAS1* mutations.[12] A history of thrombosis was reported in six patients and six relatives which mainly were of the stroke type. Close examination of the medical records of family 13 (variant D525H identified in the 76-year-old proband and her son with no history of thrombosis) found a stroke in five sisters at an age of onset between 60 and 80 years, however, no genetic test has been done in these relatives to indicate that the variant is associated with thrombosis. We also found a stroke in a child (patient 20, variant Y532C) with a Moya-Moya type cerebral vascular malformation. Unfortunately, no information could be ob-

Table 1. Clinical data of probands with a genetic variant identified in *EPAS1*.

ID	Ex	Pos. cDNA	Pos. protein	Age/age at diagnosis in years	Sex	Hb	Ht	RBC	EPO	Other symptoms	Family history	Ref	Diagnostic center
P#1	2	c.181A>G	p.Ile61Val	17	M	17.1	49.3	5.89	6.9	na	No	/	Coimbra
P#2	6	c.587C>T	p.Thr196Met	55	M	16.5	55	6	4.8	None	No	/	Nantes
P#3	6	c.734T>A	p.Leu245Gln	57	M	16	48.5	5.94	5.8	na	No	/	Dijon
P#4	7	c.818T>G	p.Leu273Arg	19	M	17.1	46	6.05	12.8	na	ne	/	Nantes
P#5	9	c.1046A>G	p.Lys349Arg	57	M	17.8	53.6	na	5.5	na	na	/	Dijon
P#6	9	c.1057G>C	p.Val353Leu	32	M	20	60	na	3	Pulmonary embolism at 30 yrs	No	/	Nantes
P#7	9	c.1121T>A	p.Phe374Tyr	57	M	17.6	50	5.71	na	None	na	G: 26-28 S: 15	Dijon
P#8	9	c.1121T>A	p.Phe374Tyr	57	M	17.5	50.5	5.88	5.6	None	No	G: 26-28 S: 15	Dijon
P#9	9	c.1121T>A	p.Phe374Tyr	57	M	18.5	55	6.2	5.7	Portal vein thrombosis at 1 yr	No	G: 26-28 S: 15	Paris
P#10	11	c.1478A>G	p.Asp493Gly	75/53	M	na	na	na	7	na	No	/	Nantes
P#11	11	c.1510C>G	p.Leu504Val	65	M	19	55.8	6.15	7.9	Tumors ne	Yes (2)	/	Dijon
P#12	11	c.1510C>G	p.Leu504Val	48/63	M	17.2	51	5.6	4.4	None	Yes (2)	/	Nantes
P#13	12	c.1573G>C	p.Asp525His	76	F	19.4	56	6.53	4.9	None; 5 deceased sisters after stroke	Yes (2)	G: 29	Nantes
P#14	12	c.1574A>G	p.Asp525Gly	44	M	20.5	56.9	6.54	16.8	ne	Yes (6)	/	Nantes/ Dijon
P#15	12	c.1578G>C	p.Leu526Phe	62/49	M	19.8	57.1	6.46	13.5	Stroke, myocardial infarction, pulmonary embolism	Yes (2)	/	Nantes
P#16	12	c.1579G>A	p.Glu527Lys	51	F	18.6	56.9	6.4	12.5	Congenital cataract	Yes (3)	/	Dijon
P#17	12	c.1588G>T	p.Ala530Ser	33/28	M	23	68	na	Norm.	None	Yes (2)	/	Nantes
P#18	12	c.1589C>A	p.Ala530Glu **1.5% reads**	6/9 months	M	19.5	65.6	10.04	na	na	No	S: 27, 30	Nantes
P#19	12	c.1591C>T	p.Pro531Ser **1.9% reads**	11/16 months	F	24.2	77.5	9.4	573	Paragangliomas, cardiac symptomes, ocular lesions[1]	No	S: 27, 31-36	Nantes
P#20	12	c.1595A>G	p.Tyr532Cys	9/7	M	14.8	53.7	8.33	>450-7,885	Stroke, PAH, ocular lesions,[2] tumors ne, Moya-Moya disease	No	S: 27, 37	Dijon
P#21	12	c.1597A>G	p.Ile533Val	32	M	20.7	58.5	6.87	15.3	None, tumors ne	Yes (3)	G: 38	Nantes
P#22	12	c.1604T>C	p.Met535Thr	17/6	M	18.2	52	5.89	na	None	Yes (5)	G: 39, 42	Nantes/ Créteil
P#23	12	c.1604T>C	p.Met535Thr	65/29	F	18.4	54	6.33	16.4	AHT	Yes (5)	G: 39, 42	Paris
P#24	12	c.1609G>C	p.Gly537Arg	26	M	20.2	59.8	6.7	13.6	ne	No	G: 9, 40, 41, 43	Nantes
P#25	12	c.1609G>A	p.Gly537Arg	11	M	18.2	53	6.13	10.1	None	No	G: 9, 40, 41, 43	Nantes
P#26	12	c.1609G>A	p.Gly537Arg	29	M	21.8	59	na	12.3	None	No	G: 9, 40, 41, 43	Nantes
P#27	12	c.1609G>A	p.Gly537Arg	23	F	19.0	55	6.4	13.1	PAH	Yes (3)	G: 9, 40, 41, 43	Nantes
P#28	12	c.1609G>A	p.Gly537Arg	/32	M	14.5	50	na	Norm.	na	No	G: 9, 40, 41, 43	Nantes
P#29	12	c.1609G>A	p.Gly537Arg	39/23	F	20	58	na	41	Cerebral thrombophlebitis	Yes (2)	G: 9, 40, 41, 43	Nantes
P#30	12	c.1612G>A	p.Glu538Lys	54/33	M	18.2	54	na	24	None	No	/	Dijon
P#31	12	c.1620C>A	p.Phe540Leu	45	M	20.3	60.5	6.7	8.3	None	na	G: 28, 42	Dijon
P#32	12	c.1642G>A	p.Glu548Lys	/28	M	19.5	62	6.52	17	None, tumors ne	Yes (3)	/	Utrecht
P#33	12	c.1671G>C	p.Gln557His	53	M	18.2	50.7	5.68	7.4	None, tumors ne	Yes (2)	/	Nantes/ Dijon
P#34	12	c.1679C>A	p.Pro560His	/32	M	17.9	51	6.23	na	na	na	/	Utrecht
P#35	12	c.1685A>T	p.His562Leu	/49	M	15.0	50	6.28	9	Stroke at 35 yrs, died of sepsis	na	/	Utrecht
P#36	12	c.1700T>C	p.Met567Thr	63	M	17.6	53	5.98	Norm.	Cardiac symptomes	No	S: 32	Nantes
P#37	12	c.1700T>C	p.Met567Thr	na	F	20.4	59.7	6.39	3.7	na	na	S: 32	Mondor
P#38	12	c.1705A>G	p.Asn569Asp	34/27	F	16.5	49	na	8.5	AHT, eclampsia	na	/	Paris
P#39	12	c.1750C>T	p.Leu584Phe	/43	M	18.4	53	5.88	7	na	No	/	Utrecht
P#40	12	c.1805G>A	p.Arg602Gln	/30	M	17.3	52	5.9	1.7	na	na	/	Nantes
P#41	12	c.1960G>A	p.Val654Ile	24	M	na	52	na	na	None	na	/	Nantes
P#42	12	c.1973G>A	p.Arg658His	73/68	M	18.6	56.5	6.55	14	Stroke, ocular lesions, tumors ne	No	/	Nantes
P#43	15	c.2474G>A	p.Arg825Gln	66	M	16.9	50	5.75	5.9	na	na	/	Dijon

Continued on following page

Position of the genetic variants are indicated. They have all been identified in the heterozygous state (approximatively 50% of the reads obtained by next-generation sequencing), except 2 mosaic patients (percentage of reads indicated in bold). All patients presented with erythrocytosis and additional clinical manifestations are indicated. ID: identification; P#: patient number; Ex: exon; Pos: position; diagnosis: the age is indicated in years or the number of months is specified; M: male; F: female; Hb: hemoglobin in g/dL (normal range, 13-16.5 g/dL for men; 12-16 g/dL for women); Ht: hematocrit in % (normal range, 40-49% for men; 37-48% for women); RBC: red blood cells in $x10^{12}$/L (normal range, 4.2-5.7$x10^{12}$/L for men, 4.2-5.2$x10^{12}$/L for women); EPO: erythropoietin in mU/mL (normal range, 5–25 mU/mL); PAH: pulmonary arterial hypertension; AHT: arterial hypertension; ne: not explored; Norm.: normal; yrs: years old; na: no data available. Ocular lesions[1]: bilateral major papilloedema with fibrotic aspect and vascular shunts; Ocular lesions[2]: monocular blindness due to retinal atrophy. Family history: the number of family members presenting erythrocytosis is indicated in brackets. Ref: references detailed in the bibliography of the main manuscript. G: germline mutation; S: somatic mutation identified in the tumor.

tained regarding the existence of any other cardiovascular risk factor in the relatives.

Of major importance, a paraganglioma has been detected in one mosaic patient (patient 19). Patient 19 presented at 16 months old with a strongly elevated Ht (77.5%), Hb (24.2 g/dL) and EPO (573 UI/mL). A first screening by NGS reads did not reveal germline mutations. In the presence of such a severe diagnosis, whole-exome and subsequently whole-genome sequencing were performed on the patient and parent's DNA, but no significant genetic abnormality was identified. At the age of 10 years, due to severe hypertension highly suggestive of the presence of a catecholamine secreting tumor, a deeper re-analysis of NGS reads allowed the identification of the c.1591C>T, p.Pro531Ser variant at a maximum rate of 1.9% of the reads that was confirmed by using droplet PCR (*Online Supplementary Figure S2*). Indeed, two abdominal paragangliomas were identified and resected. Sequencing of the tumor using droplet PCR showed a variant allele frequency of 60%, confirming the role of *EPAS1*-mutated cells in oncogenesis (*Online Supplementary Figure S2B*). After surgical resection of the paragangliomas, disappearance of the hypertrophic cardiomyopathy and the non-compaction syndrome initially found was noted.

A second patient presented at a very young age (patient 18, 9 months old) with severe erythrocytosis. The mutation c.1589C>A, p.Ala530Glu was detected at 1.5% of reads by NGS and confirmed by droplet PCR. The risk of tumor development in this patient is carefully monitored.

In silico analysis of the genetic variants

Subsequently, we employed the MetaDome web server to analyze the mutation tolerance at each position of the HIF-2α protein. The amino acids targeted by the missense genetic variants identified in our study were localized on the HIF-2α protein map (Figure 2). The resulted tolerance is reported in Table 2 for each variant, and the detailed score are indicated in the *Online Supplementary Table S4*. The majority of the variants target amino acids are located in the oxygen-dependent degradation (ODD) domain, an intolerant zone surrounding the key residue proline P531. The location of this hydroxylation site is shown in the 3D structure of HIF-2α (upper left panel of Figure 2).

We then used the Mobidetails online platform which gathers many sources of data for the interpretation of DNA variants in the context of molecular diagnosis. For each variant, we analyzed the scores obtained with single- and meta-predictors and classified the variants as benign when both scores were <0.5, and as deleterious when both scores were >0.7 (Figure 3A).

These scores are illustrated with a Radar view presenting the prediction results obtained from the different *in silico* tools (*Online Supplementary Figure S3*).

The analysis was completed by using the PROVEAN *in silico* tool. We classified the variants depending on their score regarding the cutoff: eutral >-2.5, possibly deleterious <-2.5 and >-4.1, deleterious <-4.1 (Figure 3B).

The final classification of the variants was based on a global analysis of *in silico* and functional studies (see below) with the ACMG (American College of Medical Genetics and Genomics) criteria which previously developed guidance for the interpretation of sequence variants (see the last column of Table 2, and details of scores in the *Online Supplementary Table S4*)[23].

Generation of an erythropoietin promoter-driven reporter vector

As HIF-2α-dependent regulation of EPO expression represents the key pathophysiological mechanism involved in the occurrence of erythrocytosis, we focused our functional studies on reporter assays using a luciferase gene driven by EPO regulatory elements. Previously existing luciferase constructs under the control of the EPO promoter contain the proximal region located between position -194 and -341 upstream the ATG codon (termed minimal promoter, labeled in black; *Online Supplementary Figure S4*).[19,20] A mutation located at position -136 in the *EPO* promoter has been recently described and linked to the development of erythrocytosis in two families.[18] We therefore introduced a larger promoter region into the reporter constructs from position -17 to -564 of the *EPO* promoter (construct termed Full *EPO* promoter, labeled in gray; *Online Supplementary Figure S4*).

Functional studies by using end point and real-time reporter luciferase assay

We performed luciferase reporter assays using this novel *EPO* promoter-driven construct in the absence or presence of the distal 5' and 3' hypoxia-responsive elements (HRE)

responsible for tissue-specific and hypoxia-inducible regulation of *EPO* gene expression in the kidney and the liver, respectively. Transient transfection of these constructs showed higher luciferase activity of the full promoter under hypoxic conditions compared to the minimal promoter. Furthermore, transient co-transfection of vectors encoding wild-type *HIF-1α* or *HIF-2α* displayed an increased luciferase signal of both the minimal and enlarged promoter, but resulted in preferential response to HIF-2α only when the enlarged promoter was present under both normoxic and hypoxic conditions. A similar preferential HIF-2α-dependent increase in luciferase activity was obtained with reporter

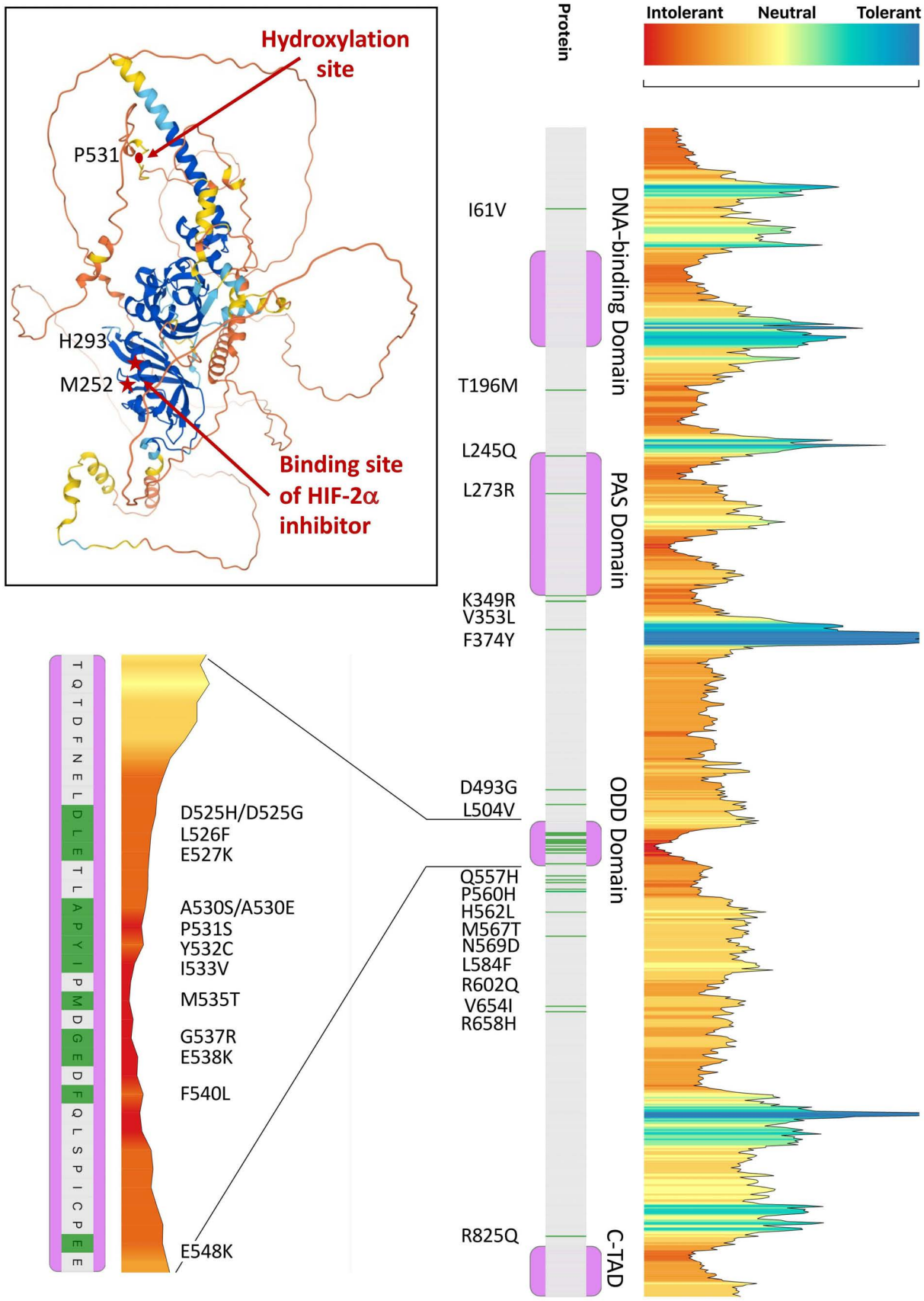

Figure 2. HIF-2α three-dimensional structure and HIF-2α protein tolerance landscape. The main domains of the HIF-2α protein are presented in purple in the middle of the figure. The severity of the impact induced by sequence variation targeting each amino acid was determined by using the MetaDome *in silico* analysis tool. It is presented with a color code, from blue (tolerant) to red (intolerance) on the right. The variants identified in our series are listed. PAS (Per-ARNT-Sim) domain: dimerization domain, ODD: oxygen dependent degradation domain, c-TAD: c-terminal transactivation domain. In the upper left panel, the three-dimensional structure of HIF-2α is presented (from AlphaFold Protein Structure Database). The HIF-2α inhibitor (MK-6482/Belzutifan/Welireg) binds to a region that causes a conformational change of the amino acids H293 and M252 (red stars in the figure), while the amino acids found mutated in patients target the close vicinity of the hydroxylation proline P531 located in the ODD domain.

Table 2. Results of *in silico* analysis and functional studies of a genetic variant identified in *EPAS1*.

ID	Ex	Pos. cDNA	Pos. protein	Pos. protein	Global gnomAD v3 frequency	Metadome prediction	Prediction PROVEAN	Radar Mobidetails prediction	Compiled luciferase functional test	ACMG class	Conclusion/ classification
P#1	2	c.181A>G	p.Ile61Val	I61V	$2.094x10^{-5}$	Slighly Int.	Neut.	Prob. Dam./Tol.	/	3	VUS
P#2	6	c.587C>T	p.Thr196Met	T196M	$7.678x10^{-5}$	Int.	Delet.	Prob. Dam./Tol.	/	3	VUS
P#3	6	c.734T>A	p.Leu245Gln	L245Q	0	Neut.	Delet.	Possibly Dam./Tol.	/	3	VUS
P#4	7	c.818T>G	p.Leu273Arg	L273R	0.0005	Int.	Delet.	Prob. Dam./Tol.	/	3	VUS
P#5	9	c.1046A>G	p.Lys349Arg	K349R	0	Int.	Neut.	Ben.	/	3	VUS
P#6	9	c.1057G>C	p.Val353Leu	V353L	0	Int.	Neut.	Possibly Dam./Tol.	/	3	VUS
P#7	9	c.1121T>A	p.Phe374Tyr	F374Y	0.0045	Tol.	Neut.	Ben.	No GOF	2	Likely Ben.
P#8	9	c.1121T>A	p.Phe374Tyr	F374Y	0.0045	Tol.	Neut.	Ben.	No GOF	2	Likely Ben.
P#9	9	c.1121T>A	p.Phe374Tyr	F374Y	0.0045	Tol.	Neut.	Ben.	No GOF	2	Likely Ben.
P#10	11	c.1478A>G	p.Asp493Gly	D493G	0	Int.	Prob. Delet.	Ben.	No GOF	3	VUS
P#11	11	c.1510C>G	p.Leu504Val	L504V	$6.977x10^{-6}$	Slighly Int.	Neut.	Prob. Dam./Tol.	No GOF	3	VUS
P#12	11	c.1510C>G	p.Leu504Val	L504V	$6.977x10^{-6}$	Slighly Int.	Neut.	Prob. Dam./Tol.	No GOF	3	VUS
P#13	12	c.1573G>C	p.Asp525His	D525H	0	Int.	Delet.	Prob. Dam.	/	4	Likely Path.
P#14	12	c.1574A>G	p.Asp525Gly	D525G	0	Int.	Delet.	Prob. Dam.	Sign. GOF	5	Path.
P#15	12	c.1578G>C	p.Leu526Phe	L526F	0	Int.	Prob. Delet.	Prob. Dam.	Sign. GOF	4	Likely Path.
P#16	12	c.1579G>A	p.Glu527Lys	E527K	0	Int.	Prob. Delet.	Prob. Dam.	Sign. GOF	4	Likely Path.
P#17	12	c.1588G>T	p.Ala530Ser	A530S	0	Int.	Prob. Delet.	Prob. Dam.	Sign. GOF	5	Path.
P#18	12	c.1589C>A	p.Ala530Glu **1.5% reads**	A530E	0	Int.	Delet.	Prob. Dam.	/	5	Path.
P#19	12	c.1591C>T	p.Pro531Ser **1.9% reads**	P531S	0	Highly Int.	Delet.	Prob. Dam.	Sign. GOF	5	Path.
P#20	12	c.1595A>G	p.Tyr532Cys	Y532C	0	Int.	Delet.	Prob. Dam.	Sign. GOF	5	Path.
P#21	12	c.1597A>G	p.Ile533Val	I533V	0	Highly Int.	Neut.	Prob. Dam.	/	5	Path.
P#22	12	c.1604T>C	p.Met535Thr	M535T	0	Highly Int.	Delet.	Prob. Dam.	/	5	Path.
P#23	12	c.1604T>C	p.Met535Thr	M535T	0	Highly Int.	Delet.	Prob. Dam.	/	5	Path.
P#24	12	c.1609G>C	p.Gly537Arg	G537R	0	Highly Int.	Prob. Delet.	Prob. Dam./Tol.	Sign. GOF	5	Path.
P#25	12	c.1609G>A	p.Gly537Arg	G537R	0	Highly Int.	Prob. Delet.	Prob. Dam./Tol.	Sign. GOF	5	Path.
P#26	12	c.1609G>A	p.Gly537Arg	G537R	0	Highly Int.	Prob. Delet.	Prob. Dam./Tol.	Sign. GOF	5	Path.
P#27	12	c.1609G>A	p.Gly537Arg	G537R	0	Highly Int.	Prob. Delet.	Prob. Dam./Tol.	Sign. GOF	5	Path.
P#28	12	c.1609G>A	p.Gly537Arg	G537R	0	Highly Int.	Prob. Delet.	Prob. Dam./Tol.	Sign. GOF	5	Path.
P#29	12	c.1609G>A	p.Gly537Arg	G537R	0	Highly Int.	Prob. Delet.	Prob. Dam./Tol.	Sign. GOF	5	Path.
P#30	12	c.1612G>A	p.Glu538Lys	E538K	0	Highly Int.	Prob. Delet.	Prob. Dam.	Sign. GOF	3	VUS
P#31	12	c.1620C>A	p.Phe540Leu	F540L	0	Int.	Delet.	Prob. Dam.	Sign. GOF	3	VUS
P#32	12	c.1642G>A	p.Glu548Lys	E548K	$1.396x10^{-5}$	Int.	Neut.	Prob. Dam./Tol.	/	3	VUS
P#33	12	c.1671G>C	p.Gln557His	Q557H	$6.983x10^{-5}$	Int.	Neut.	Possibly Dam./Tol.	Close to WT	3	VUS
P#34	12	c.1679C>A	p.Pro560His	P560H	$4.187x10^{-5}$	Int.	Neut.	Possibly Dam./Tol.	/	3	VUS
P#35	12	c.1685A>T	p.His562Leu	H562L	0	Int.	Neut.	Ben.	/	3	VUS
P#36	12	c.1700T>C	p.Met567Thr	M567T	0.0002	Int.	Neut.	Ben.	/	3	VUS
P#37	12	c.1700T>C	p.Met567Thr	M567T	0.0002	Int.	Neut.	Ben.	/	3	VUS
P#38	12	c.1705A>G	p.Asn569Asp	N569D	$4.192x10^{-5}$	Int.	Neut.	Ben.	/	3	VUS
P#39	12	c.1750C>T	p.Leu584Phe	L584F	$6.978x10^{-6}$	Neut.	Neut.	Ben.	/	3	VUS
P#40	12	c.1805G>A	p.Arg602Gln	R602Q	$1.397x10^{-5}$	Slighly Int.	Neut.	Ben.	/	3	VUS
P#41	12	c.1960G>A	p.Val654Ile	V654I	$2.094x10^{-5}$	Slighly Int.	Neut.	Ben.	/	3	VUS
P#42	12	c.1973G>A	p.Arg658His	R658H	0.0002	Slighly Int.	Neut.	Ben.	/	3	VUS
P#43	15	c.2474G>A	p.Arg825Gln	R825Q	$6.98x10^{-6}$	Neut.	Neut.	Ben.	/	3	VUS

The prediction of each bioinformatic tool is illustrated by a color: red for a pathogenic/damaging classification, green for benign/neutral impact and orange for intermediate/unknown signification. The prediction obtained with the Mobidetails single predictors are shown and additional information is added (/) when the prediction obtained with the meta-predictors is divergent. The classification obtained with the functional tests (end point luciferase assay performed in different cell lines; activity and/or velocity of the reaction obtained by real-time luciferase assay) is illustrated by a color: red for a significant gain of function obtained in at least 2 assays, orange for a significant gain of function obtained in at least 1 experiment and green for no significant gain of function or a function close to the wild-type (WT) protein. Pos: position; Int.: intolerant; Tol.: tolerated; Neut.: neutral; Del.: deleterious; Prob.: probably; Dam.: damaging; Ben.: benign; Sign.: significant; GOF: gain of function; PROVEAN: protein variation effect analyzer; ACMG: American College of Medical Genetics and Genomics; VUS: variant of unknown signification. Percentage of reads indicated in bold in 2 mosaic patients.

genes in the presence or absence of distal HREs (Figure 4; *Online Supplementary Figure S5*). The same experiments were performed with co-transfection of plasmids encoding the mutated HIF-2α proteins, including P531A as positive control in HEK293, Hep3B and Kelly cells (1 clone for each variant in triplicates). We observed a significant gain of function for the P531S mutant in all three cell lines and for A530S, G537R and F540L in Hep3B cells (Figure 5; Figure 6B; *Online Supplementary Figure S7A*). We also performed these experiments under hypoxic conditions confirming the results under normoxic conditions, with a significant gain of function for the P531S mutant in three

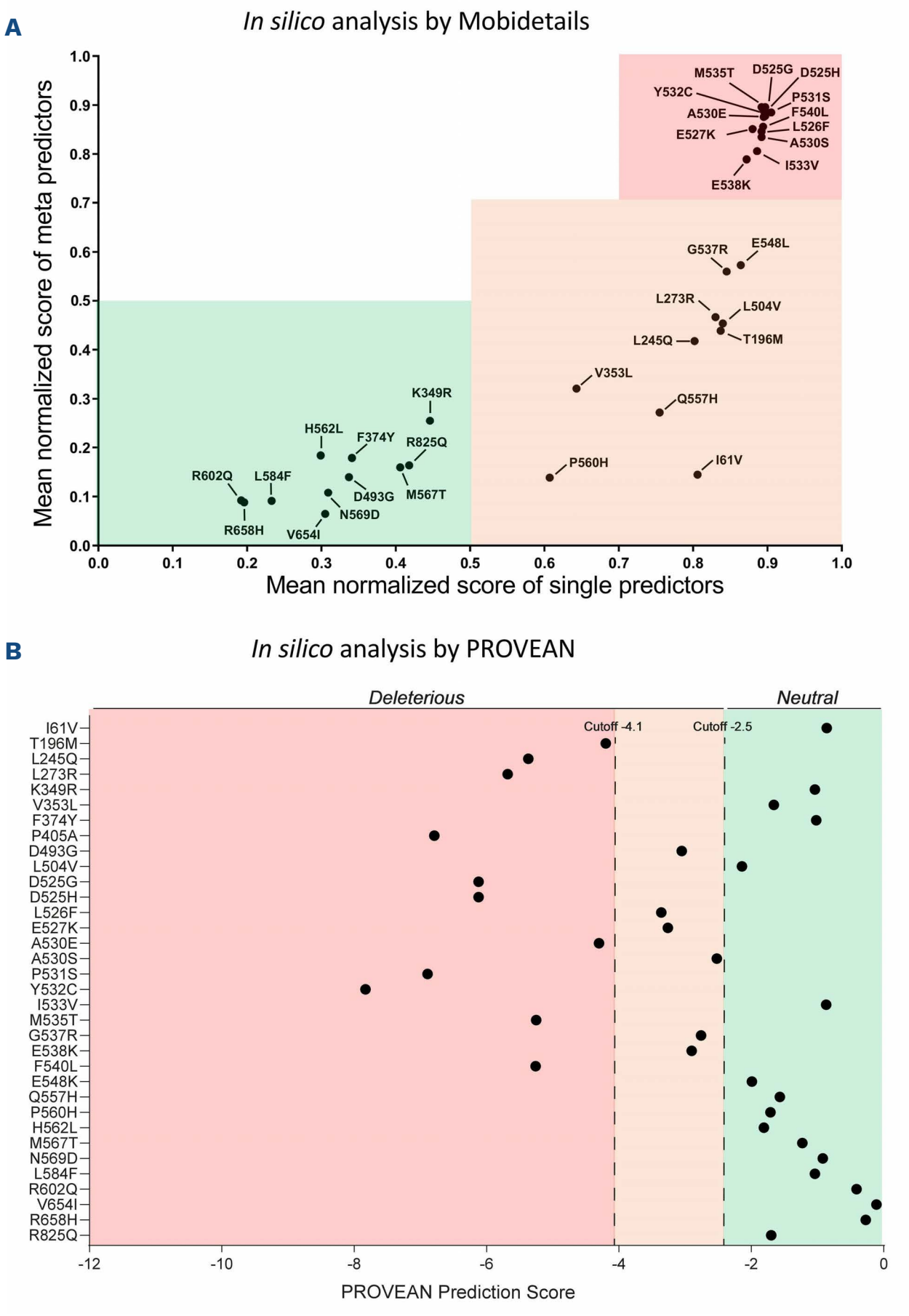

Figure 3. *In silico* prediction of the impact of an amino acid substitution on the HIF-2α protein. (A) Representation score obtained by the predictors analyzed by the Mobidetails annotation platform. Values are normalized (0-1), 0 being the less damaging and 1 the most damaging for each predictor. The graph indicates the mean normalized score obtained by single predictors (SIFT, Polyphen 2 HumDiv an0d HumVar) and meta-predictors (Fathmm, REVEL, ClinPred, Meta SVM, Meta LR, Mistic). (B) Representation of protein variation effect analyzer (PROVEAN) analysis performed on each non-synonymous HIF-2α variant. Variants with a score above -2.5 were predicted to be neutral. A threshold score below -2.5 was predicted to be deleterious and a more stringent threshold score of -4.1 is associated with increased specificity.

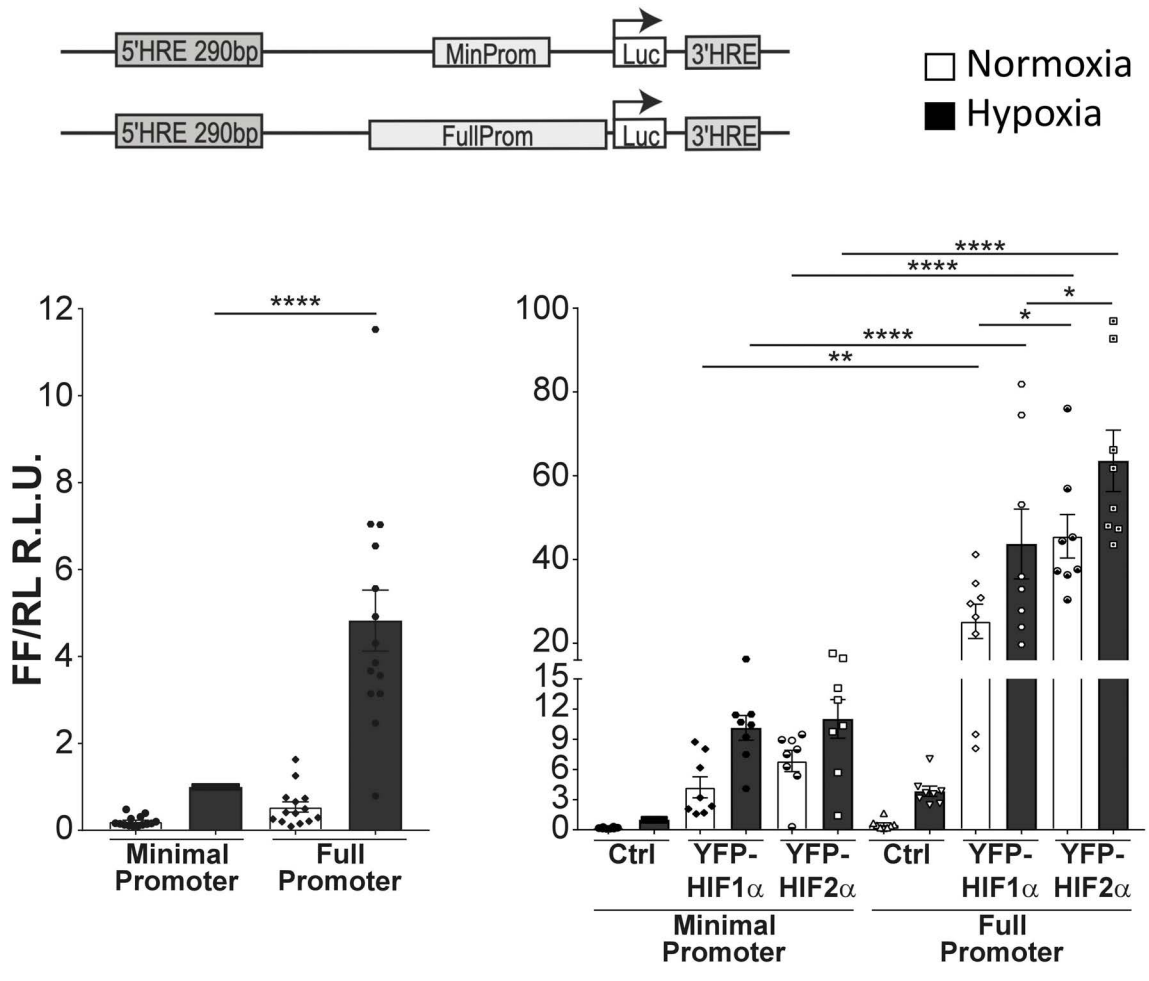

Figure 4. End point reporter gene assays demonstrate increased erythropoietin promoter-driven luciferase activity with a full erythropoietin promoter compared to minimal construct. HEK293 cells were co-transfected with the erythropoietin (*EPO*) minimal or full promoter-driven constructs containing the 5'HRE and 3'HRE and with HIF-1α or HIF-2α isoform overexpression plasmids, as indicated. The luciferase activity is reported as the induction compared to the control (Ctrl) under hypoxic condition and represents the ratio of firefly (FF) to *Renilla* (RL) relative light units (RLU). Each column represents the mean ± standard error of the mean of 4 to 14 different experiments performed in duplicate. One-way ANOVA (*P≤0.05; **P≤0.01; ****P≤0.0001).

cell lines and for the E538K in Hep3B (*Online Supplementary Figures S6* and *S7B*).

In order to increase the sensitivity of the reporter assay we reduced the amount of transfected HIF-2α expression vectors. We used HEK293 cells in order to avoid additional induction of the reporter vector by endogenous HIF-2α (expressed in Hep3B and Kelly cells). In order to determine the optimal time frame to quantify the luciferase activity after transfection, we followed it using a real-time luciferase measurement system for 48 h (Figure 6A). We set the 100% activity to the relative light unit (RLU) values obtained at the plateau of the wild-type protein after 30 h of expression from the beginning of the reaction (i.e., 34 h after transfection). We observed a large scale of activity from wild-type (around 100%) to elevated gain-of-function (150-300%). The mean of results derived from repeated experiments obtained at the plateau are shown in Figure 6B. We calculated the slope of the different curves to obtain an indicator of the velocity of the reaction (Figure 6C). The use of three independent clones for each variant shows consistent results with strong reliability of this test within an experiment, but some variants display variable behavior between experiments, demonstrating the importance to replicate assays. Altogether, a significant increase of the activity and/or the velocity of the reaction was found for the tested variants identified in patients targeting amino acids from D525 to E538. One variant presents an activity very close to the wild-type protein and was classified as benign (Q557H), the other variants have been classified as variant of unknown signification (VUS).

Discussion

We report here the largest collection of *EPAS1/HIF2A* gene variants from a large cohort of patients with idiopathic erythrocytosis. Variants were characterized using detailed *in silico* and new functional studies.

The use of a wide range of *in silico* prediction tools can be very useful for classifying variants in genetic diagnosis. However, the pathogenicity of a genetic variant can be difficult to classify as pathogenic when the disease is associated with a gain-of-function mutation. For this reason, functional studies need to be included in the diagnostic tools. In the case of erythrocytosis, the functional tests must be very sensitive because it is known that many mutations may be hypomorphic,[44–46] especially because the genes of the hypoxia pathway play a major role in physiology and a germline alteration that is too severe and would likely be incompatible with life. The construction of an improved, more sensitive *EPO* promoter-driven vector in combination with the use of a real-time luciferase reporter assay measurements allowed us to accurately measure gain of function of new HIF-2α mutants. Altogether, combined *in silico*, segregation and functional studies performed in this study allowed the classification of 11 variants as likely pathogenic or pathogenic (including 4 never previously described) in 17 patients and 23 relatives. Our study further expanded the region of HIF-2α associated with erythrocytosis. Thus far, this region spanned from amino acids 529 to 537, but the classification of pathogenic variants of amino acids 525, 526 and 527 prompt an expansion of this

area. Additional investigations will be necessary to definitely classify the E538K and F540L variants.

Our results will allow an appropriate follow-up of families at the clinical level, especially with regard to complications and associated diseases. It is indeed intriguing that so few patients carrying pathogenic variants present with additional symptoms associated with *EPAS1* mutations: PAH was only described in two patients (P#20 and 27), ocular lesions in two patients (P#19, 20) and thrombosis and ischemic accidents in five families (P#13, 15, 20, 22, 29). Interestingly, patient 20 carrying the variant p.Tyr532Cys also suffers from the Moya-Moya disease. This is a rare and chronic disease of mostly unknown causes that affects the blood vessels in the brain. No link has ever

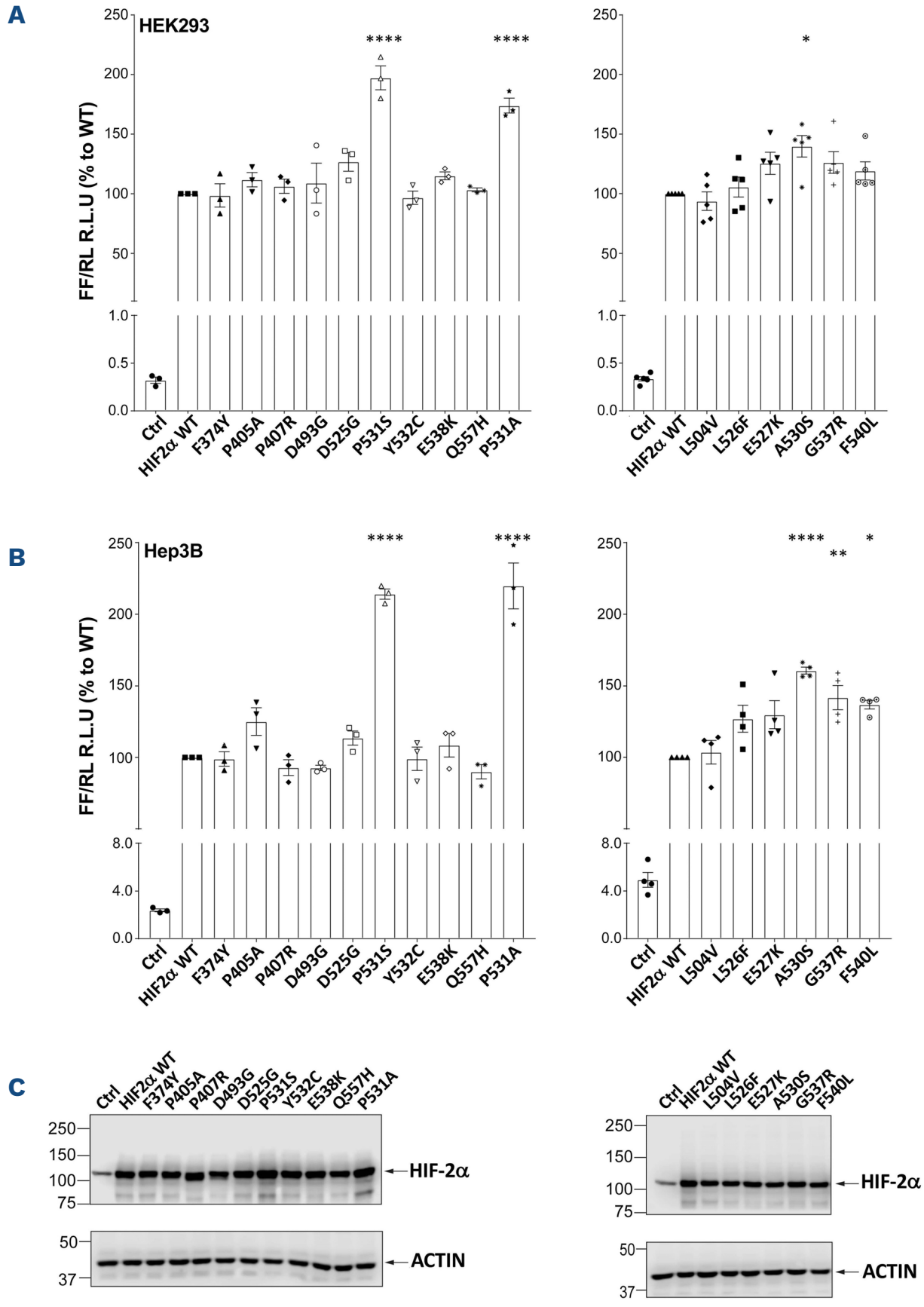

Figure 5. End point reporter gene assays demonstrate a significant gain of function of the P531S HIF-2α mutant. HEK293 cells (A) and Hep3B cells (B) were transfected with 5'HRE and 3'HRE-dependent erythropoietin (*EPO*) full promoter-driven construct and different HIF-2α mutants identified from patients, as well as with wild-type (WT) and positive control P531A HIF-2α constructs, as indicated. Luciferase activity is reported as the induction compared to the control (Ctrl) under normoxic conditions and represents the ratio of firefly (FF) to *Renilla* (RL) relative light units (RLU). Each column represents the mean ± standard error of the mean of 3 to 5 different experiments performed in duplicate. (C) Expression levels of the different HIF-2α proteins were assessed with a HIF-2α antibody. Actin and α-tubulin were used as loading controls. One-way ANOVA, compared to HIF-2α WT (*$P \leq 0.05$; ***$P \leq 0.001$; ****$P \leq 0.0001$).

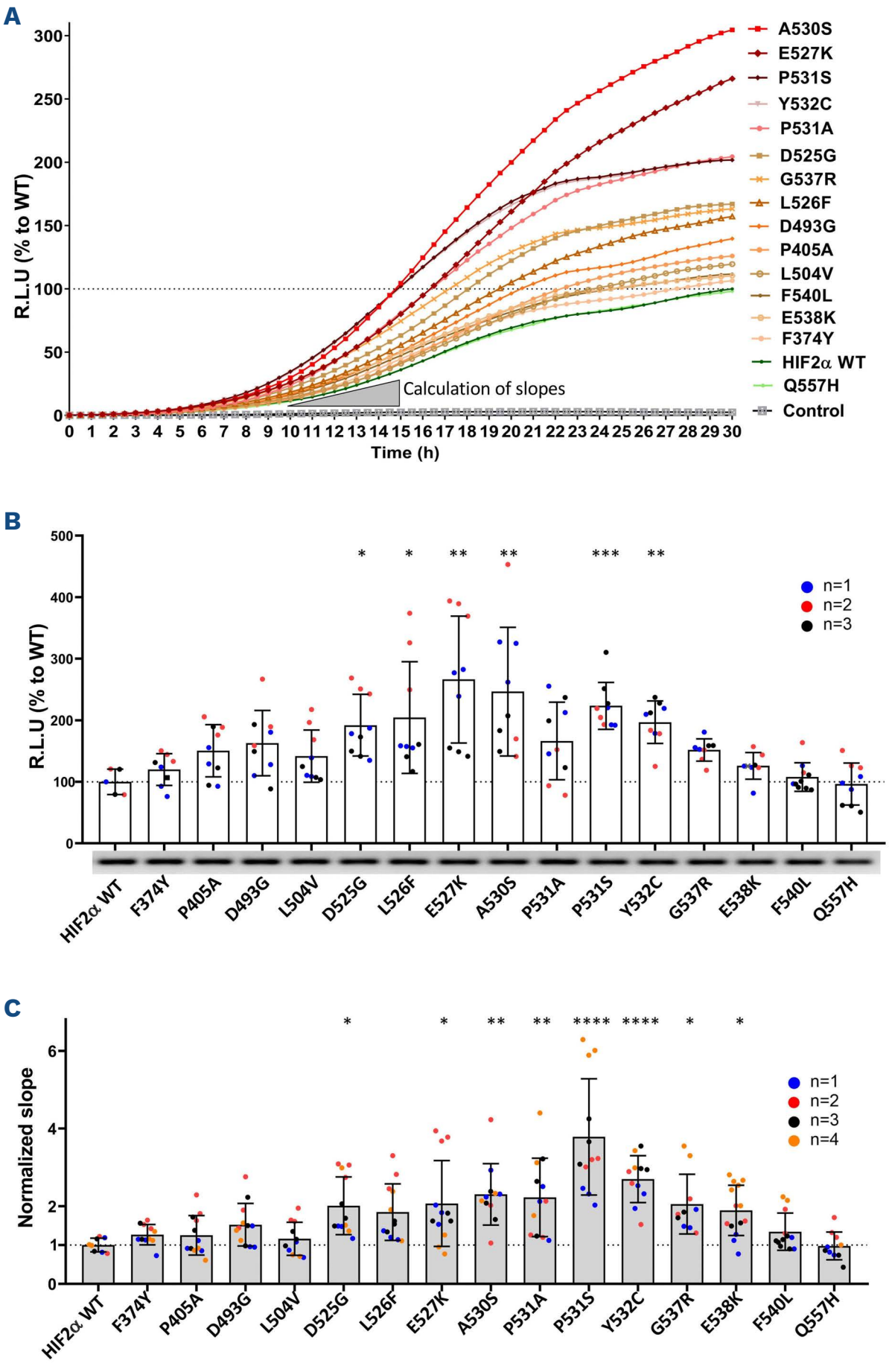

Figure 6. Real-time reporter gene assays demonstrate a significant gain of function for additional HIF-2α variants. (A) HEK293 cells were transfected with 3'HRE erythropoietin (*EPO*) full promoter luciferase reporter construct and different HIF-2α mutants. Relative light unit (RLU) was measured every 30 minutes for 48 hours (h). A plateau was obtained 30 h after the beginning of reaction (i.e., 34 h after transfection). The 100% was set on the activity of the wild-type (WT) protein at this plateau. Each mutant curve represents the mean of 3 independent clones. (B) Representation of luciferase activity obtained with the real-time luciferase assay, at the plateau obtained 30 h after the beginning of the reaction (i.e., 34 h after transfection); The 100% was set on the activity of the wild-type protein at this plateau. Each column represents the mean ± standard error of the mean of experiments performed in triplicate and each chip represents 1 clone, 3 different clones was tested each time. The quantity of transfected HA-plasmid was verified by polymerase chain reaction targeting the HA tag. The results obtained on agarose gel are shown. One-way ANOVA, Kruskal–Wallis tests with Dunn's *post hoc* analysis for multiple comparisons with HIF-2α WT was performed (**P*≤0.05; ***P*≤0.01; ****P*≤0.001). (C) Representation of curve slopes obtained from the real-time luciferase measurement between 10 h and 15 h after the beginning of the reaction (see grey triangle in Figure 6A), normalized to the WT protein. Each column represents the mean ± standard error of the mean of an experiment performed in quadruplicate, each color represents an experiment and each chip represents 1 clone, 3 different clones being were tested each time. One-way ANOVA, Kruskal–Wallis tests with Dunn's *post hoc* analysis for multiple comparisons with HIF-2α WT was performed (**P*≤0.05; ***P*≤0.01; ****P*≤0.001).

been found with the hypoxia pathway genes and additional investigations should be performed on the potential role of *EPAS1* in the etiology of Moya-Moya disease.

Some families are currently being closely monitored due to the high risk of evolution of tumor development. In the literature, somatic or mosaic pathogenic variants described in *EPAS1* target the amino acid from position 529 to 532. In our study, four families carry mutations that target these amino acids. Among them, only a single patient (P#19), carrying the p.Pro531Ser pathogenic variant at a mosaic state, developed multiple paragangliomas at the age of 10 years. Two other young patients at high risk carry *EPAS1* variants already described in tumors and will need close follow-up: patient 18 (mosaic p.Ala530Glu, 6 years old) and patient 20 (germline p.Tyr532Cys, 9 years old). Noteworthy, it would be important to extend the follow-up of tumor occurrence to the family carrying the variant p.Leu526Phe. This variant indeed targets the consensus hydroxylation motif "LXXLAP" ($L_{526}XXL_{529}A_{530}P_{531}$) that plays a major role in tumor occurrence when mutated.

It is interesting to note the complexity of genotype/phenotype correlations in this disease[17] with the example of

family 17 who carries the p.Ala530Ser variant (a substitution never described before this study) and do not present a tumor history despite a highly significant gain of function. A similar case has been described in a family who developed only erythrocytosis associated with the germline mutation variant p.Tyr532His.[40] The precise substitution targeting key amino acids may therefore be of major importance for the severity of the disease.

Of note, our study reintroduces the debate about the variant F374Y that was initially described as causal in patients with erythrocytosis and paraganglioma[15,26-28] and found to be deleterious in one functional study.[47] Here this variant, that could be considered as a polymorphism due to its high frequency in the population, has been clearly classified as benign. For patients in for whom variants have been classified as VUS with predictions in favor of a benign effect (i.e., patients 38 to 42), additional investigations should be conducted to identify the cause of the carrier's pathology. The search for associated variants in the same gene was carried out and only one was found: the c.1218C>G, p.Thr406Thr variant (P#41), whose high frequency provides little support for a gain of function. The association with other variants in the hypoxia pathway genes associated with erythrocytosis (*EGLN1, VHL*) should also be explored, including in intronic regions that may impact splicing.[45] As a matter of fact, substantial research efforts remain to be made as approximately 70% of erythrocytosis cases are still of unknown genetic cause.

The identification of two young cases carrying a variant at a mosaic state at a rate up to 1.9% of reads shows the potential failure to diagnose a number of patients with erythrocytosis. The development of a paraganglioma 8 years after the discovery of a major erythrocytosis, associated with very high EPO levels, is instructive: mosaicism may be present at a very low level that is not automatically detectable by usual NGS analysis. It is therefore necessary to lower the variant detection thresholds or to manually analyze the NGS results for the *EPAS1* gene, notably in young children with severe erythrocytosis.

Interestingly, our study opens up new research avenues related to HIF-2α-dependent *EPO* gene regulation. Indeed, expansion of the DNA sequence surrounding the minimal promoter previously described in the reporter vector not only resulted in substantially increased luciferase activity under both normoxic and hypoxic conditions, but also conferred more specificity towards HIF-2α (*vs*. HIF-1α) under overexpression conditions. This additional sequence does not contain any consensus HRE (G/ACGTG)[5] and strongly suggests the presence of additional regulatory sequences within the *EPO* promoter. It will be of interest to analyze which factors bind to this region and if these proximal elements further contribute to tissue-specific and conditional *EPO* gene regulation in co-operation with distal enhancers.

Overall, this study allowed the classification of *EPAS1* variants classified as causal mutations in 40 individuals (17 patients and 23 relatives). This classification is of major importance giving the new therapeutics that specifically target and inhibit the HIF-2α protein. Indeed, a clinical trial with the HIF-2α inhibitor (MK-6482/Belzutifan/Welireg) on a single patient with a HIF-2α mutation has been recently published.[48] This patient carried the p.Ala530Glu pathogenic variant at a mosaic state and the treatment led to a rapid resolution of the erythrocytosis, hypertension, headaches in addition to paraganglioma response.[48] This remarkable effect opens up the possibility to treat more patients carrying mutations in *EPAS1* which may potentially modify the drug binding capacity. Indeed, the chemical drug binds to a region that causes a conformational change of the amino acids H293 and M252 located in the HIF-2α-PAS-B domain. This drug-induced shift of the residue position, with a move of the side chains towards the binding surface, weakened the binding capacity of HIF-2α with HIF-β (ARNT-PASB) to form an active transcription factor.[49] Importantly, none of the mutations associated with erythrocytosis are located in the drug-targeted region (Figure 2, upper left panel). Therefore, all the patients carrying a causal mutation in *EPAS1*, located between amino acids 525 to 537, are theoretically eligible for HIF-2α inhibitor treatment.

Conclusions

Our collaborative study showed that enhanced functional assays in combination with *in silico* methods can improve the diagnosis in patients with erythrocytosis with an unclassified mutation in *EPAS1*. We also demonstrated the advantage of federating all diagnostic laboratories working on this rare pathology. This allows to increase the number of cases and the power of the analyses that can render genetic variants informative. It is also necessary, for a pathology linked to frequently hypomorphic mutations, to multiply the *in silico* analysis and to refine the functional approaches. A precise classification of mutations is indeed essential for a better diagnosis, clinical follow-up and access to a targeted treatment.

Disclosures

No conflicts of interest to disclose.

Contributions

VK, DM, AB, LS, VA, ALR, SC, VL, MD, DH and BG performed experiments. AR, FL, performed bioinformatics analyses. CG, NM, FA, BA, LM, MR, SB, BC, FG, AP GR, NB, HC, RvW, CB, FG and the consortium ECYT4 conducted the medical and diagnostic studies. BG, DH and FG wrote the manuscripts. BG, FG and DH designed the study. BG directed the study. All authors contributed to the research and approved the final manuscript.

Funding

This study was supported by grants from the Agence Nationale de la Recherche (ANR; PRTS 2015 "GenRED"; AAPG2020 "SplicHypoxia"), the Labex GR-Ex, reference ANR-11-LABX-0051, the Fondation Maladies Rares (FMR) and Kiwanis project FONDATION-GenOmics 2017, and the associations VHL Alliance USA, VHL France and Génavie. This work was also supported as a part of NCCR Kidney.CH, a National Center of Competence in Research, funded by the Swiss National Science Foundation (grant number 183774) and by Swiss National Science Foundation project grant 310030_207460.

Data-sharing statement

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

Appendix: ECYT-4 members

Marc Bernard, Charles Bescond, Anaïse Blouet, Juliette Bouteloup, Françoise Boyer, Aisha Bruce, Emilie Caysials, Nicole Casadevall, Brieuc Cherel, Marie Laure Couec, Guillaume Denis, Louis Devron, Cécile Dumesnil, Thierry Lamy de la Chapelle, Marion Gambart, Loïc Garçon, Brigitte Granel, Pierre Hirsch, Arnaud Hot, Agnès Lahary, Tabita Maia, Amira Mejri, Sandrine Meunier, Franck-Emmanuel Nicolini, Marie Nolla, Natalina Miguel, Nathalie Parquet, Emmanuel Raffoux, Dana Ranta, Laure Ricard, Nicolas Schleinitz, Jean-Baptiste Valentin and Mathieu Wemeau.

References

1. 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.
2. Wang GL, Semenza GL. Characterization of hypoxia-inducible factor 1 and regulation of DNA binding activity by hypoxia. J Biol Chem. 1993;268(29):21513-21518.
3. Appelhoff RJ, Tian YM, Raval RR, et al. Differential function of the prolyl hydroxylases PHD1, PHD2, and PHD3 in the regulation of hypoxia-inducible factor. J Biol Chem. 2004;279(37):38458-38465.
4. Lando D, Peet DJ, Whelan DA, Gorman JJ, Whitelaw ML. Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch. Science. 2002;295(5556):858-861.
5. Wenger RH, Stiehl DP, Camenisch G. Integration of oxygen signaling at the consensus HRE. Sci STKE. 2005;2005(306):re12.
6. Semenza GL, Koury ST, Nejfelt MK, Gearhart JD, Antonarakis SE. Cell-type-specific and hypoxia-inducible expression of the human erythropoietin gene in transgenic mice. Proc Natl Acad Sci U S A. 1991;88(19):8725-8729.
7. Scortegagna M, Ding K, Zhang Q, et al. HIF-2alpha regulates murine hematopoietic development in an erythropoietin-dependent manner. Blood. 2005;105(8):3133-3140.
8. Percy MJ, Furlow PW, Lucas GS, et al. A gain-of-function mutation in the HIF2A gene in familial erythrocytosis. N Engl J Med. 2008;358(2):162-168.
9. Gale DP, Harten SK, Reid CD, Tuddenham EG, Maxwell PH. Autosomal dominant erythrocytosis and pulmonary arterial hypertension associated with an activating HIF2 alpha mutation. Blood. 2008;112(3):919-921.
10. Martini M, Teofili L, Cenci T, et al. A novel heterozygous HIF2AM535I mutation reinforces the role of oxygen sensing pathway disturbances in the pathogenesis of familial erythrocytosis. Haematologica. 2008;93(7):1068-1071.
11. 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.
12. Pacak K, Chew EY, Pappo AS, et al. Ocular manifestations of hypoxia-inducible factor-2α paraganglioma-somatostatinoma-polycythemia syndrome. Ophthalmology. 2014;121(11):2291-2293.
13. Vaidya A, Flores SK, Cheng Z-M, et al. EPAS1 mutations and paragangliomas in cyanotic congenital heart disease. N Engl J Med. 2018;378(13):1259-1261.
14. Zhuang Z, Yang C, Lorenzo F, et al. Somatic HIF2A gain-of-function mutations in paraganglioma with polycythemia. N Engl J Med. 2012;367(10):922-930.
15. Taïeb D, Barlier A, Yang C, et al. Somatic gain-of-function HIF2A mutations in sporadic central nervous system hemangioblastomas. J Neurooncol. 2016;126(3):473-481.
16. Buffet A, Smati S, Mansuy L, et al. Mosaicism in HIF2A-related polycythemia-paraganglioma syndrome. J Clin Endocrinol Metab. 2014;99(2):E369-373.
17. Tarade D, Robinson CM, Lee JE, Ohh M. HIF-2α-pVHL complex reveals broad genotype-phenotype correlations in HIF-2α-driven disease. Nat Commun. 2018;9(1):3359.
18. Taylor JC, Martin HC, Lise S, et al. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. Nat Genet. 2015;47(7):717-726.
19. Orlando IMC, Lafleur VN, Storti F, et al. Distal and proximal hypoxia response elements cooperate to regulate organ-specific erythropoietin gene expression. Haematologica. 2020;105(12):2774-2784.
20. Storti F, Santambrogio S, Crowther L, et al. A novel distal upstream hypoxia response element regulating oxygen-dependent erythropoietin gene expression. Haematologica. 2014;99(4):e45-e48.
21. Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. MetaDome: pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. Human Mutation. 2019;40(8):1030-1038.
22. Baux D, Van Goethem C, Ardouin O, et al. MobiDetails: online DNA variants interpretation. Eur J Hum Genet. 2021;29(2):356-360.
23. 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.
24. Koay TW, Osterhof C, Orlando IMC, et al. Androglobin gene expression patterns and FOXJ1-dependent regulation indicate its functional association with ciliogenesis. J Biol Chem. 2021;296:100291.

25. De Backer J, Maric D, Bosman M, Dewilde S, Hoogewijs D. A reliable set of reference genes to normalize oxygen-dependent cytoglobin gene expression levels in melanoma. Sci Rep. 2021;11(1):10879.
26. Lorenzo FR, Yang C, Ng Tang Fui M, et al. A novel EPAS1/HIF2A germline mutation in a congenital polycythemia with paraganglioma. J Mol Med (Berl). 2013;91(4):507-512.
27. Welander J, Andreasson A, Brauckhoff M, et al. Frequent EPAS1/HIF2α exons 9 and 12 mutations in non-familial pheochromocytoma. Endocr Relat Cancer. 2014;21(3):495-504.
28. Oliveira JL, Coon LM, Frederick LA, et al. Genotype-phenotype correlation of hereditary erythrocytosis mutations, a single center experience. Am J Hematol. 2018;93:1029-1041.
29. Schelker RC, Herr W, Grassinger J. A new exon 12 mutation in the EPAS1 gene possibly associated with erythrocytosis. Eur J Haematol. 2019;103(1):64-66.
30. Toyoda H, Hirayama J, Sugimoto Y, et al. Polycythemia and paraganglioma with a novel somatic HIF2A mutation in a male. Pediatrics. 2014;133(6):e1787-1791.
31. Favier J, Buffet A, Gimenez-Roqueplo AP. HIF2A mutations in paraganglioma with polycythemia. N Engl J Med. 2012;367(22):2161; author reply 2161-2.
32. Comino-Mendez I, de Cubas AA, Bernal C, et al. Tumoral EPAS1 (HIF2A) mutations explain sporadic pheochromocytoma and paraganglioma in the absence of erythrocytosis. Hum Mol Genet. 2013;22(11):2169-2176.
33. Toledo RA, Qin Y, Srikantan S, et al. In vivo and in vitro oncogenic effects of HIF2A mutations in pheochromocytomas and paragangliomas. Endocr Relat Cancer. 2013;20(3):349-359.
34. Därr R, Nambuba J, Del Rivero J, et al. Novel insights into the polycythemia-paraganglioma-somatostatinoma syndrome. Endocr Relat Cancer. 2016;23(12):899-908.
35. Vaidya A, Flores SK, Cheng Z-M, et al. EPAS1 mutations and paragangliomas in cyanotic congenital heart disease. N Engl J Med. 2018;378(13):1259-1261.
36. Abdallah A, Pappo A, Reiss U, et al. Clinical manifestations of Pacak-Zhuang syndrome in a male pediatric patient. Pediatr Blood Cancer. 2020;67(4):e28096.
37. Pacak K, Jochmanova I, Prodanov T, et al. New syndrome of paraganglioma and somatostatinoma associated with polycythemia. J Clin Oncol. 2013;31(13):1690-1698.
38. Perrotta S, Stiehl DP, Punzo F, et al. Congenital erythrocytosis associated with gain-of-function HIF2A gene mutations and erythropoietin levels in the normal range. Haematologica. 2013;98(10):1624-1632.
39. Alaikov T, Ivanova M, Shivarov V. EPAS1 p.M535T mutation in a Bulgarian family with congenital erythrocytosis. Hematology. 2016;21(10):619-622.
40. 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.
41. Liu Q, Tong D, Liu G, et al. HIF2A germline-mutation-induced polycythemia in a patient with VHL-associated renal-cell carcinoma. Cancer Biol Ther. 2017;18(12):944-947.
42. Percy MJ, Chung YJ, Harrison C, et al. Two new mutations in the HIF2A gene associated with erythrocytosis. Am J Hematol. 2012;87(4):439-442.
43. Percy MJ, Beer PA, Campbell G, et al. Novel exon 12 mutations in the HIF2A gene associated with erythrocytosis. Blood. 2008;111(11):5400-5402.
44. 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.
45. 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.
46. Ladroue C, Carcenac R, Leporrier M, et al. PHD2 mutation and congenital erythrocytosis with paraganglioma. N Engl J Med. 2008;359(25):2685-2692.
47. Dwight T, Kim E, Bastard K, et al. Functional significance of germline EPAS1 variants. Endocr Relat Cancer. 2020;28(2):97-109.
48. 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.
49. 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.