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

Supplementary Data

Supplementary Table 1:

List of *EGLN1/PHD2* mutations and genetic variants described in the literature. Patients carrying variants located after the STOP codon (*) were not included in the list of the 64 different *PHD2* variants described in the introduction. VUS: variants of unknown significance. HA: Headache. Phleb: Phlebothomy. HFE: homeostatic iron regulator gene responsible for hereditary haemochromatosis. TIA: transient ischemic attack. Yrs: years.

Supplementary Table 2:

Compilation of hematological data identified in patients with erythrocytosis described in the present study. A) Patients carrying mutations classified as likely pathogenic or pathogenic. (B) Patients carrying variants of unknown significance (VUS) or likely benign (lik. Ben.). M: Male, F: Female.

Supplementary Table 3:

Compiled *in silico* studies of *EGLN1/PHD2* variants from the present study (in bold) and the literature. Frequency and detailed scores obtained by *in silico* studies using Mobidetails and Metadome are indicated. SPiP: Splicing Pipeline Prediction. Hmz: Homozygous. Ψ: Same proband as already published but with updated information in the present paper. Data obtained from UK Biobank studies are reported: nc: not conclusive; ns: hemoglobin (Hb) and hematocrit (Hct) levels in patients with *EGLN1* variants not significantly elevated compared to the wild type population; if the Hb and/or Hct are elevated, the percentile (>95th, 99th) are indicated. ACMG, American College of Medical Genetics and Genomics; BS, benign strong; BP, benign supporting; PM, pathogenic moderate; PP, pathogenic supporting; PS, pathogenic strong. Criteria used for the ACMG classification[29] are listed. Class1: benign; Class 2: likely benign; Class 3: variant of uncertain significance (VUS); Class 4: likely pathogenic; Class 5: pathogenic.
BP4: variant predicted as benign by all prediction software (here, when Mobidetails single and metapredictors scores <0.4),
BA1: high allelic frequency in the general population, scored here if frequency >1%,
BS1: allelic frequency too high compared to the frequency of the pathology, scored here if frequency is higher or equal to 5.10^{-4},
BS2 : variant observed in healthy individuals, here in the UK Biobank, associated with Hct and Ht normal levels,

PM1: variant located on a mutational hot-spot and/or a well established functional domain,
PM2: variant absent from population control databases, scored here if frequency comprised between 0 and $<1.10^{-5}$,
PM4 : variant affecting the length of the protein (Indels in phase, loss of stop codon),
PM5: variant causing a different amino acid change at the same position of a known pathogenic missense mutation,
PP1: variant co-segregating with the disease in several affected members of the same family (here, at least 3 members),
PP3: variant predicted as deleterious by all prediction software (here, when Mobidetails single and metapredictors scores >0.6),
PP: an expert pathogenic supporting criteria has been added to mutations associated with premature STOP codon, knowing that functional studies of truncated PHD2 proteins demonstrated a severe loss of function[13],
PS3: deleterious impact demonstrated by a functional study,
PS4: variant present in different unrelated patients and allelic frequency compatible with the presence and penetrance of the disease,
PVS1: pathogenic criterion weighted as very strong.

Supplementary Table 4:
Compilation of hematological data identified in patients with erythrocytosis described in the literature carrying mutations classified as likely pathogenic or pathogenic (A) or carrying variants of unknown significance (VUS) or likely benign (B). Variants p.Asp4Glu and p.Cys127Ser were not included in the count. M: Male, F: Female.

Supplementary Figure 1:
A) Localization of the *PHD2* missense variants on the protein. Protein's Tolerance Landscape of the PHD2 protein has been established from the MetaDome web server. MetaDome analyses the mutation tolerance at each position in a human protein. Major function domains of PHD2 are indicated in purple. The missense variants already described are listed on the left, the variants described in this study are listed on the right (all the variants are novel excepted the W334R already described in [9], indicated by Ψ). The impact of a mutation of each amino-acid is presented in color: red for high impact and intolerance, blue for tolerance. B) Schematic representation of D50H variant localization. The amino-acid D50 is part of the conserved zinc finger domain involved in the binding of the heat shock protein 90 (HSP90) co-chaperones p23

protein. PHD2 binds the p23 protein through the zinc finger domain and forms a complex (containing p23/HSP90 and FKBP38) which is suspected to facilitate HIF-1α hydroxylation (from Sinnema et al., 2018[4] and Song et al., 2013[40]). C) Schematic representation of the c.-410G>T variant localization. The human *PHD2* promotor region is presented in the upper panel and the region containing the nucleotide -410 is enlarged in the lower panel. The variant c.-410G>T is located in the HIF Binding Site (HBS) that contain the consensus CGTG sequence. The variant targeted the first G of the HIF binding sequence (from Metzen et al., 2005[30] and Wenger et al., 2005[31]).

Supplementary Figure 2:

Multiple sequence alignment of PHD2 homologues. Conservation of PHD2 sequences throughout various taxonomic classes. Grey-scale shading indicates conservation of iso-functional amino acid residues.

Supplementary Figure 3:

Prediction of the functional impact of the *PHD2* missense proteins with Mobidetails, an annotation platform dedicated to the interpretation of DNA variations. Representation of the results with a radar graph are shown. The higher is the score, the more surface is red and the more deleterious is the variant.

Supplementary Figure 4:

All the variants of the *EGLN1* gene have been analyzed and cases with hematocrit (Hct) and hemoglobin (Hb) values significantly elevated compared to the control population are presented. Scores obtained by *in silico* studies using Mobidetails are indicated. ID: position on chromosome 1. SPiP: Splicing Pipeline Prediction. Nb: number of carriers. Sd: standard deviation. Threshold used for the percentiles are indicated in the lower part of the figure.

Supplementary Figure 5:

In vitro measurement of PHD2 activity. Recombinant Flag-tagged proteins have been produced in insect cells and affinity purified using anti-Flag affinity. SDS-Page analysis of recombinant affinity purified human wild-type (WT) PHD2 and its variants W334R, G349S, G349C and R371H. Coomassie blue staining was used for visualization of the proteins. The arrow indicates the position of the enzymes. Molecular weight marker was run along the samples, the numbers indicate kDa.

Supplementary Figure 6:

Study of PHD2 protein stability by Cycloheximide (CHX) Chase Assay. A) Representative PHD2 western blots obtained for each variant. IB PHD2: immunoblot of transfected HA-PHD2 by using anti-HA antibody. B) Quantification of the amount of transfected HA-PHD2 protein after treatment with cycloheximide, reported in percentage of the initial HA-PHD2 protein level (100% at 0h of CHX treatment) and normalized to the intensity of ACTIN. Data are shown as mean ± standard error of the mean (SEM) from three independent experiments. The two-way ANOVA test has been used for statistics (****$p \leq 0.0001$).

Supplementary Figure 7:

Prediction of the impact of the *PHD2* variants on splicing. A) Prediction using Alamut Visual, an exploration software application for genomic variation. B) Prediction using the Splicing Pipeline Prediction (SPiP).

Supplementary Figure 8:

Study of hiPSC established from patient carrying the c.1152C>T, p.Y384Y variant and differentiated into the hepatocyte-like cells. A) Summary of the protocol of differentiation of hiPSC into hepatocyte-like cells. D: Day. B) RT-PCR performed on PHD2 mRNA (exon 3-5) have been performed and PCR products have been loaded on a TapeStation® migration system a for quantification. For the wild-type hiPSC, one or two clones from two different lines were cultured (clones 1 and 2 from Line 1, and clone 4 from Line 2). For the mutant hiPSC, two clones from one reprogrammed line were cultured (clones 1 and 4 from Line 1). Differentiation have been performed in replicates (number of the differentiation is indicated).

Supplementary Methods

NGS sequencing

Series of patients presenting with erythrocytosis were collected for sequencing in laboratories of genetic diagnosis: 405 patients from Dijon (France); 236 patients from Nantes (France); 280 patients from Créteil (Hopital Henri Mondor), France ; 90 from Paris (St Louis Hospital), France; 316 patients from Munich, Germany; 250 patients in Utrecht, Netherland; 220 patients in Coimbra, (patients from Viseu, Savilha and Salamanca), Portugal; 100 patients from Belfast, UK; 143 patients from Barcelona, Spain, 123 patients, Padova, Italy. DNA was extracted from whole blood on EDTA by using the QuickGene DNA Whole blood kit L (Kurabo) on the QuickGene610L® (in Nantes), QIAGEN spin-columns on the QIAcube (St Louis Hospital, France), QIAGEN gDNA Blood Kit on the QIAsymphony SP (Coimbra, Portugal), Roche MagNA Pure System (Roche LifeScience) with the MagNAPure96 DNA and Viral NA LV Kit (Munich, Germany), GenoVision M48 -Qiagen (Barcelona, Spain), EuroGold Blood DNA Mini Kit Plus (EuroClone, Italy).

Molecular screening was performed by high throughput sequencing with different technologies, depending on the sequencing center: KAPA HyperPlus" (Roche) associated with IDT probes for the capture provided by Sophia Genetics (Nantes and Dijon, France); KAPA Hyperchoice Max 3MB T3 - 12 RXN (Roche) (Créteil, France); HaloPlex (Agilent) with sequencing on MiSeq and bioinformatic analysis with SureSelect (St Louis Hospital, France); Agilent SureSelectXT capture library and sequence analysis on the Illumina sequencing platform (Netherland); AmpliSeq Library with sequencing on Plataforma IonS5 (Thermo Fisher Scientific) (Coimbra, Portugal); library preparation with the Illumina DNA Prep Kit using Unique Dual Indices (UDI) and DNA target regions (enriched using the IDT Hybridization Capture Protocol) sequenced with Illumina NovaSeq 6000 instruments (Munich, Germany), QIAseq Custom DNA Panels (Qiagen) sequenced with Illumina MiSeq or NextSeq (Barcelona, Spain); "AmpliSeq for Illumina On-Demand, (Illumina, Inc.), Padova, Italy.

The NGS panel contained the following genes: *VHL, EGLN1, EGLN2, EGLN3, HIF1A, EPAS1, EPOR, SH2B3, JAK2, BPGM*, *EPO* (Nantes and Dijon, France); *BPGM, CYB5R3, EGLN1, EGLN2, EGLN3, HIF2A, HIF1A, HIF1AN, HIF3A, EPO, EPOR, SH2B3, VHL, JAK2, HBB, HBA1, HBA2 , MPL, CALR , KDM6A, GFI1B, BHLHE41, OS9, ZNF197, PIEZO1, MVK , THRA, FH, HIKESHI, HSF1, HSPA4, HSPA8, HSPB1, HSPH1, MITF, P4HTM, USP20, VHLL, XPO1* (Créteil, France); *BPGM, EGLN1, EGLN2, EGLN3, EPAS1, EPO, EPOR, FH, HIF1A, HIF1AN, HIF3A, JAK2, MITF, P4HTM, SH2B3, USP20, VHL, VHLL, GFI1B* (St Louis

Hospital, France); *BPGM*, *EGLN1*, *EGLN2*, *EGLN3*, *EPAS1*, *EPO*, *EPOR*, *HBA1*, *HBA2*, *HBB*, *HIF1A*, *HIF3A*, *JAK2*, *SH2B3*, *VHL* (Coimbra, Portugal); *BHLHE41, BPGM, EGLN1, EGLN2, EGLN3, EPAS, EPO, EPOR, GFI1B, HBA1, HBA2, HBB, HIF1A, HIF1AN, HIF3A, JAK2, KDM6A, OS9, SH2B3, VHL,* and *ZNF197* (Munich, Germany); *VHL, EPAS1, EGLN1, EGLN2, EPO-R, BPGM*, *SH2B3, JAK2, HBA1*, *HBA2*, *HBB* (Belfast, UK); *EPOR, EPAS1, EGLN1, HBA1, HBA2, HBB, JAK2, MPL, THPO, BPGM, VHL, SH2B3* (Barcelona, Spain); *EGNL1, EPAS1, VHL, EPOR, BPGM, JAK2, HFE, HJV, TFR2, FTH, FTL, SLC40A1, SLC11A2, HAMP,* Padova, Italy.

The NGS panel covers all the exons, intron/exon junctions (minimum 25 base pairs of the intronic sequences) and partial sequences of 5' and 3'UTR (size depending on the gene, detailed sequences and bed files available upon request. The NM_022051.2 was used for nomenclature.

Quantitative PCR to validate EGLN1 deletion.

Search for molecular rearrangements and relative copy number quantification was performed by real-time PCR (qPCR) technique as described in DD Ct described by Livak et al (Methods 25, 402-408 (2001)). A SYBR premix Ex Taq (Takara) and Primers located in exons 3 and 4 of the *EGLN1* gene were used. The PCR was performed with the Light Cycler 480 (ROCHE A.S.) apparatus.

In silico analysis:

Links for the websites used to analyze the genetic variants are indicated below:

MetaDome: https://stuart.radboudumc.nl/metadome/dashboard

MobiDetails: https://mobidetails.iurc.montp.inserm.fr/MD/genes

The frequency of variants was informed from the values reported by gnomAD on the Mobidetails site. This site was also used for the localization of variants with AlphaFold Protein Structure Database ; for the predictions of mutation impact by single and meta predictors; and for the analysis of splicing with SPiP-Splicing Pipeline Prediction. SPiP is a randomForest model running a cascade of bioinformatics tools. Briefly, SPiP uses SPiCE tool for the consensus splice sites (donor and acceptor sites), MES for polypyrimidine tract between -13 and -20, BPP for branch point area between -18 and -44, an homemade score to research cryptic/de novo activation and ΔtESRseq for exonic splicing regulatory element until to 120 nt in exon.

UK Biobank analysis:

The UK Biobank resource was used to analyze the hematological parameters of the carriers of *EGLN1* genetic variants (application number 49823). The UK Biobank study, described in detail previously[32], is a population-based prospective cohort in the United Kingdom in which >500 000 individuals aged between 40 and 69 years were included from 2006 to 2010. The study has been approved by the North West Multi-Centre Research Ethics Committee for the United Kingdom, from the National Information Governance Board for Health and Social Care for England and Wales, and by the Community Health Index Advisory Group for Scotland (REF: https://www.ukbiobank.ac.uk/media/0xsbmfmw/egf.pdf) and all participants have given informed consent (REF: https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf). The records of 77 individuals (last update February 22nd 2022) who have withdrawn from UK Biobank were remove from the analyses.

Whole-Exome Sequencing

Full details of the whole-exome sequencing (WES) in the UK Biobank have been reported previously[33]. In short, WES was performed using IDT xGen Exome Research Panel v1.0, targeting 38 997 831 bases in 19 396 genes. Exomes were captured including 100 bp flanking regions. At the time of analysis, WES data were available for 200 643 participants. WES and phenotypic data were available for 184 896 participants (83 147 men and 101 749 women).

Genetic analysis, data processing and plotting

Variant filtering (GRCh38) was performed using vcf files (format VCFv4.2) with bcftools (v1.14) (REF: doi:10.1093/gigascience/giab008). Genetic and phenotypic data were combined and processed using RStudio (v.2022.02.1). Plots were generating using RStudio using ggplot2 and ggridges R libraries.

Hemoglobin and hematocrit

Hemoglobin, in g/dL, was measured on a Beckman Coulter LH750 analyser (Beckman Coulter, Ltd, United Kingdom). Hematocrit, in percentage, has been calculated by the formula: (red blood cells x mean corpuscular volume) / 10 both measured on a Beckman Coulter LH750 analyser (Beckman Coulter, Ltd, United Kingdom).

Expression and purification of PHD2 protein to perform in vitro activity assays

Baculoviruses were generated for the wild type and mutated PHD2 with C-terminal FLAG and His tags in the pVL1393 plasmid as described before.[34] Sf9 cells were infected with the PHD2 viruses and infected cells were harvested 72 h after infection. Cells were washed with PBS and resuspended in lysis buffer containing 10 mM Tris-HCl pH 7.8, 0.1 M NaCl, 0.1 M glycine, 5

μM $FeSO_4$, and 0.1% Triton X-100. Cells were lysed by homogenization, and insoluble debris was separated by centrifugation. Soluble lysate was incubated with anti-FLAG M2 affinity gel (Sigma), the gel was washed with wash buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, and 5 μM $FeSO_4$, and bound proteins were eluted with wash buffer containing additionally 150 μg/ml FLAG peptide. The recombinant HIF-2α ODDD was expressed and purified as described previously.[35]

In vitro enzyme activity assays

Flag-tagged PHD2 and its variants W334R, G349S, G349C and R371H were expressed in insect cells and affinity purified using anti-Flag and the His-tagged HIF-2α ODDD protein was expressed in *E*. Coli and affinity purified using NiNTA as described previously.[34,33] Synthetic peptide representing the HIF-1α C-terminal hydroxylation site (DLDLEMLAPYIPMDDDFQL) was used as another substrate. 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.[36] To determine the kinetic constants for the substrates and co-substrates, PHD2 wild-type and variant enzymes were incubated in reaction mixtures in which other components had a saturating concentration while the concentration of the (co-)substrate to be studied was varied. The 2-oxoglutarate constants were determined using the HIF-1α peptide as substrate.

Cell culture

HEK293T (human embryonic kidney) and Hep3B (hepatocarcinoma) cell lines were maintained in DMEM supplemented with 10% FCS, 1 % L-Glutamine, and 1% penicillin/streptomycin. UT-7 (megakaryoblastic leukemia) cells were maintained in α-MEM supplemented with 20% FCS and 1μg/mL EPO. Lymphoblastoid cell lines (LCL) were maintained in RPMI supplemented with 10% FCS, 1% L-Glutamine, and 1% penicillin/streptomycin. LCLs from each patient were grown and harvested independently three times for triplicate experiments. Two separate flasks were seeded for each LCL, with puromycin (InvivoGen) added to one flask at a final concentration of 200μg/ml and incubated overnight. Cells were harvested the next day.

Mutagenesis:

PHD2 mutations were introduced into a pcDNA3-HA-PHD2 vector or pCAS2 minigene vector using the Q5 Site-directed Mutagenesis Kit® (New England Biolabs) or the In-Fusion® kit

(Takara) according to the manufacturer's instructions. Following transformation into chemically competent bacteria, mutated plasmids were extracted using the NucleoSpin® Plasmid kit (Macherey-Nagel), sequenced, amplified, and purified using the Nucleobond® Xtra Maxi Plus EF kit (Macherey-Nagel).

Luciferase reporter assay:

End point Luciferase assays were performed on HEK293T (human embryonic kidney 293T) cells seeded in 12-well plates (2×10^5 cells per well) 24 hours before transfection. The transfection reagent jetPRIME® (Ozyme Polyplus) was used for transient transfections. The expression vectors pcDNA3-HA-PHD2, wild type, or mutants, (25 to 400 ng) were co-transfected with pcDNA3-HA- HIF-2α (100 ng), pGL3 3xHRE-luciferase reporter plasmid (200 ng), pRenilla vector (20 ng) and pCMV-HA-empty vector for a total amount of 800 ng transfected DNA. Luciferase activity was measured 24 hours after transfection by using a Dual-Luciferase Reporter Assay System (Promega).

Real-time Luciferase assays were performed 24 hours before transfection on HEK 293T cells seeded in 24-well Black Visiplate Perkin Elmer) (1×10^5 cells per well). The transfection reagent jetPRIME® (Ozyme Polyplus) was used for transient transfections. Expression vectors pcDNA3-HA-PHD2 (wild-type, or mutants, 20 to 80 ng) were co-transfected with pcDNA3-HA-HIF-2α (200 ng), pGL3 3xHRE-luciferase reporter plasmid (100 ng), and pCMV-HA-empty vector for a total amount of 500 ng of transfected DNA. Luciferase activity was measured over 24 hours using the bioluminometer WSL-1565 Kronos HT® (ATTO). Cells were harvested at different time points and lysed in passive lysis buffer (Invitrogen) for immunoblot detection.

Protein stability study by Cycloheximide Chase Assay:

Experiments were performed on HEK 293T cells seeded 24 hours before transfection in 12-well plates (2×10^5 cells per well). Cells were co-transfected using the jetPRIME® (Ozyme Polyplus) reagent with the plasmids encoding pcDNA3-HA-PHD2 wild-type or mutated (100 to 800 ng) in addition to pCMV-HA-empty vector for a total amount of 800 ng of transfected DNA. 24 hours after transfection, cells were treated with cycloheximide (CHX) (Sigma-Aldrich) at a final concentration of 100µg/ml. Cells were harvested at different time points. A equal volume of protein lysates was analyzed by Western blot assay.

Western Blotting

Western blotting was performed using the NuPAGE® kit from Invitrogen according to the manufacturer's instructions. Briefly, 20µL of cell lysates from the luciferase reporter assay was mixed with to Sample Loading Buffer and Reducing Agent, then loaded into a Bis-Tris Mini Bolt Gel (4-12%). Migration was performed in MES SDS Running Buffer and proteins were transferred to a nitrocellulose membrane (0.2 µM Hybond ECL; GE Healthcare) using the Invitrogen Transfer system. The membrane was blotted with a mouse anti-HA antibody (clone 16B12, BioLegend), anti-actin (Sigma) and a goat anti-mouse secondary antibody (Jackson Immuno Research). Bound antibody was then revealed using the Clarity™ Western ECL Substrate (BioRad). Quantification was performed using the ImageLab software.

Splicing reporter assay and study of PHD2 exon 4 skipping in patient's cells

Minigene constructs were prepared according to Cooper and Gaildrat.[37] Briefly, mutant fragments containing the exon of interest, along with flanking intron sequences, were generated by PCR amplification from patient genomic DNA, using oligonucleotides engineered to contain the BamH1 and MluI restriction enzyme sites at their 5' end. Following a double digest and gel purification, the PCR product was ligated into a pCas2 plasmid containing two artificial exons from *SERPING* gene (A and B) that had been cut using BamH1/MluI and dephosphorylated. DH5α-competent bacteria were then transformed with the plasmids. Clones from each transformation were collected and amplified, with plasmid DNA harvested using the NucleoSpin® Plasmid kit (Macherey-Nagel). Plasmid DNA was then sequenced, and clones carrying the mutated plasmids amplified. The cloned intron sequences flanking PHD2-Exon 3 (137 base pairs, pb) contain approximately 140 bp and 230 bp in the 5' and 3' end respectively. The cloned intronic sequences on each side of PHD2-Exon 4 (68 base bp) contain approximately 250 bp and 220 bp. Cells were transfected with 2 µg of each pCas2 plasmid containing the exon of interest, wild type sequence, or mutant. Chemical transfection was performed with the jetPRIME® reagent (Polyplus) according to the manufacturer's instructions. UT-7 human megakaryoblastic leukemia cells were transfected by nucleofection using the Amaxa technology (Kit V, Lonza). Transfection efficiency was assessed using 2µg of a GFP-containing plasmid. 24 hours after transfection, cells were harvested and total RNA extracted using the NucleoSpin® RNA kit (Macherey-Nagel). Reverse transcription reactions were carried out with 0.5µg or 1µg of RNA using Maxima First Strand cDNA Synthesis Kit (ThermoScientific). PCR amplification was performed with 50ng of cDNA using the PCR GoTaqQ2 kit (Promega) or TaqDNA polymerase native (Invitrogen). Primers against artificial

exons (A and B) of the pCAS2 minigene was used to quantify the mRNA expressed by the construct. PCR products were resolved on a 2% agarose gel.

To assess the endogenous splicing, primers surrounding exon 4 have been used: primers in PHD2 exon 3: TTGCTGACATTGAACCCAAA and primer in *PHD2* exon 5: CCTTCAGATTCGGTCGGTAA). Primers against RPL13A was used as control housekeeping gene (Forward: AAAAGCGGATGGTGGTTCCT, Reverse: GCTGTCACTGCCTGGTACTT). PCR products were resolved on a 2% agarose gel. Quantification of the *PHD2*-exon 4 skipping was performed with primers located in endogenous *PHD2* exons, and PCR products were analyzed using the TapeStation® migration system (Agilent), an automated electrophoresis method for fine quantification.

Generation and characterization of human induced pluripotent stem cell (hiPSC) lines

The hiPSC lines from patient carrying the c.1152C>T, p.Tyr384Tyr variant were generated from peripheral blood mononuclear cells (PBMCs) in the iPSC core facility of Nantes University. PBMCs were reprogrammed by Sendai viruses expressing *OCT4, SOX2, KLF4* and *c-MYC* (CytoTune^{TM}-IPS 2.0 Sendai Reprogramming kit, Life Technologies). The hiPSC clones were picked and expanded on mouse embryonic fibroblasts (MEFs) feeder cells in KSR-FGF2 medium (DMEM/F12 supplemented with 0.1% β-mercaptoethanol, 20% knockout serum replacement, 10 ng/mL basic fibroblast growth factor, 2 mmol/L l-glutamine and 1% NEAA). Mycoplasma detection was realized by using the MycoAlert™ kit (LONZA, LT07-318). RT-qPCR has been performed to validate the reprogrammation. Total RNA was extracted using RNeasy® columns and DNAse-treated using RNase-free DNase (Qiagen). For quantitative PCR, first-strand cDNAs were generated using 500ng of RNA, SuperScript™ II Reverse Transcriptase (Invitrogen), 25µg/ml polydT and 9.6µg/ml random primers (Invitrogen). To quantitate transcripts, absolute quantitative PCR was performed on a StepOne (Applied Biosystems) using power SYBR green PCR master mix (Applied Biosystems), for genes listed in the primers table (end of Experimental Procedures). For each sample, the ratio of specific mRNA level relative to GAPDH levels was calculated. Experimental results are shown as levels of mRNA relative to the highest value. SeV expression was measured using primers recommended by the manufacturer.

PCR primers used to validate hiPSC:

Gene Name	Primer sequence 5'-3'	Amplicon size (bp)	Melting Temp. (°C)	Position
GAPDH	AATCCCATCACCATCTTCCA TGGACTCCACGACGTACTCA	82	80.5	494-576
OCT4	TGGGTGGAGGAAGCTGACAACAAT TTCGGGCACTGCAGGAACAAATTC	142	82.1	1005-1147
SOX2	CCTACTCGCAGCAGGGCACC CTCGGCGCCGGGGAGATACA	169	xx	1114-1283
NANOG	ATAGCAATGGTGTGACGCAGAAGG CTGGTTGCTCCACATTGGAAGGTT	116	82	701-816

All primers have a hybridization temperature of 60°C. *GAPDH, OCT4* and *NANOG* amplicons span two adjacent exons.

Differentiation of hiPSC into hepatocyte-like cells (HLCs)

HiPSCs from passage 21-25 were differentiated into HLCs as previously described.[38,39] HiPSCs were plated in tissue-culture dishes previously coated with Matrigel at 0.05 mg/ml for 1 h, and cultured in StemMACS iPS-Brew. Once cells reached 80-85% confluence, differentiation was performed with RPMI 1640 (Life Technologies) and B27 (Life Technologies) containing Activin A (AA; Miltenyi), fibroblast growth factor 2 (FGF2; Miltenyi), bone morphogenetic protein 4 (BMP4; Miltenyi) and hepatocyte growth factors (HGFs; Miltenyi) with the following sequence: AA/BMP4/FGF2 (2 days); AA (3 days), BMP4/FGF2 (5 days) and HGF (5 days). Cells were then incubated with Hepatocyte Culture Medium (Lonza) supplemented with oncostatin M (Miltenyi) for 5 days.

After differentiation of hiPCS, hepatic markers are studied by Real-time RT-PCR performed with Taqman gene expression assays (Applied Biosystems) using a 7900HT Fast Real-Time PCR System (Applied Biosystems) or on a StepOne (Applied Biosystems). Relative quantification of selected genes: *ALB* (albumin, Hs00910225_m1), *HNF4* (hepatocyte nuclear factor 4, Hs00230853_m1), *AFP* (alpha fetoprotein, Hs00173490_m1), *FOXA2* (forkhead box A2, Hs00232764_m1) was performed against a standard curve and the values were normalized against the input determined for the housekeeping genes, *RPLP0* (ribosomal protein lateral stalk subunit P0, Hs99999902_m1), *ACTB* (actin beta, Hs99999903_m1) or *RPL13A* (ribosomal protein L13a, Hs04194366_g1). TaqMan Gene Expression Assay IDs (Applied Biosystems) are shown in parentheses after the gene names.

References of supplementary data

(1) Xiang, K., Ouzhuluobu, Peng, Y., Yang, Z., Zhang, X., Cui, C., Zhang, H., Li, M., Zhang, Y., Bianba, Gonggalanzi, Basang, Ciwangsangbu, Wu, T., Chen, H., Shi, H., Qi, X., and Su, B. (2013) Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. *Mol Biol Evol 30*, 1889–98.
(2) Oliveira, J. L., Coon, L. M., Frederick, L. A., Hein, M., Swanson, K. C., Savedra, M. E., Porter, T. R., Patnaik, M. M., Tefferi, A., Pardanani, A., Grebe, S. K., Viswanatha, D. S., and Hoyer, J. D. (2018) Genotype–phenotype correlation of hereditary erythrocytosis mutations, a single center experience. *American Journal of Hematology 93*, 1029–1041.
(3) Jalowiec, K. A., Vrotniakaite-Bajerciene, K., Capraru, A., Wojtovicova, T., Joncourt, R., Rovó, A., and Porret, N. A. (2021) NGS Evaluation of a Bernese Cohort of Unexplained Erythrocytosis Patients. *Genes (Basel) 12*, 1951.
(4) Sinnema, M., Song, D., Guan, W., Janssen, J. W. H., van Wijk, R., Navalsky, B. E., Peng, K., Donker, A. E., Stegmann, A. P. A., and Lee, F. S. (2018) Loss-of-function zinc finger mutation in the EGLN1 gene associated with erythrocytosis. *Blood 132*, 1455–1458.
(5) Provenzano, A., Chetta, M., De Filpo, G., Cantini, G., La Barbera, A., Nesi, G., Santi, R., Martinelli, S., Rapizzi, E., Luconi, M., Maggi, M., Mannelli, M., Ercolino, T., and Canu, L. (2022) Novel Germline PHD2 Variant in a Metastatic Pheochromocytoma and Chronic Myeloid Leukemia, but in the Absence of Polycythemia. *Medicina (Kaunas) 58*, 1113.
(6) Lorenzo, F. R., Huff, C. D., Myllymaki, M., Swierczek, S., Semenza, G. L., Xing, J., Koivunen, P., and Prchal, P. T. (2014) EGLN1 variant associated with the Tibetan adaptation in high altitude. *Keystone Symposia 2014. Sensing and signaling of hypoxia : interfaces with biology and medicine.*
(7) Lorenzo, F. R., Simonson, T. S., Yang, Y., Ge, R., and Prchal, J. T. (2010) A Novel PHD2 Mutation Associated with Tibetan Genetic Adaptation to High Altitude Hypoxia. *ASH Annual Meeting Abstracts. Paper344020*.
(8) Petousi, N., Croft, Q. P., Cavalleri, G. L., Cheng, H. Y., Formenti, F., Ishida, K., Lunn, D., McCormack, M., Shianna, K. V., Talbot, N. P., Ratcliffe, P. J., and Robbins, P. A. (2013) Tibetans living at sea level have a hyporesponsive hypoxia-inducible factor (HIF) system and blunted physiological responses to hypoxia. *J Appl Physiol (1985)*.
(9) Bento, C., Almeida, H., Maia, T. M., Relvas, L., Oliveira, A. C., Rossi, C., Girodon, F., Fernandez-Lago, C., Aguado-Diaz, A., Fraga, C., Costa, R. M., Araujo, A. L., Silva, J., Vitoria, H., Miguel, N., Silveira, M. P., Martin-Nunez, G., and Ribeiro, M. L. (2013) 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 91*, 361–8.
(10) Bonnin, A., Gardie, B., Girodon, F., Airaud, F., Garrec, C., Bézieau, S., Vignon, G., Mottaz, P., Labrousse, J., and Lellouche, F. (2020) [A new case of rare erythrocytosis due to EGLN1 mutation with review of the literature]. *Rev Med Interne 41*, 196–199.
(11) Gangat, N., Oliveira, J. L., Porter, T. R., Hoyer, J. D., Al-Kali, A., Patnaik, M. M., Pardanani, A., and Tefferi, A. (2022) Erythrocytosis associated with EPAS1(HIF2A), EGLN1(PHD2), VHL, EPOR or BPGM mutations: The Mayo Clinic experience. *Haematologica 107*, 1201–1204.
(12) Albiero, E., Ruggeri, M., Fortuna, S., Bernardi, M., Finotto, S., Madeo, D., and Rodeghiero, F. (2011) Analysis of the oxygen sensing pathway genes in familial chronic myeloproliferative neoplasms and identification of a novel EGLN1 germ-line mutation. *Br J Haematol 153*, 405–8.
(13) Ladroue, C., Hoogewijs, D., Gad, S., Carcenac, R., Storti, F., Barrois, M., Gimenez-Roqueplo, A. P., Leporrier, M., Casadevall, N., Hermine, O., Kiladjian, J. J., Baruchel, A., Fakhoury, F., Bressac-de Paillerets, B., Feunteun, J., Mazure, N., Pouyssegur, J., Wenger, R.

H., Richard, S., and Gardie, B. (2012) Distinct deregulation of the hypoxia inducible factor by PHD2 mutants identified in germline DNA of patients with polycythemia. *Haematologica 97*, 9–14.
(14) Al-Sheikh, M., Moradkhani, K., Lopez, M., Wajcman, H., and Prehu, C. (2008) 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 40*, 160–5.
(15) Albiero, E., Ruggeri, M., Fortuna, S., Finotto, S., Bernardi, M., Madeo, D., and Rodeghiero, F. (2012) Isolated erythrocytosis: study of 67 patients and identification of three novel germ-line mutations in the prolyl hydroxylase domain protein 2 (PHD2) gene. *Haematologica 97*, 123–7.
(16) Bento, C., Percy, M. J., Gardie, B., Maia, T. M., van Wijk, R., Perrotta, S., Della Ragione, F., Almeida, H., Rossi, C., Girodon, F., Aström, M., Neumann, D., Schnittger, S., Landin, B., Minkov, M., Randi, M. L., Richard, S., Casadevall, N., Vainchenker, W., Rives, S., Hermouet, S., Ribeiro, M. L., McMullin, M. F., Cario, H., ECE-Consortium, Chauveau, A., Gimenez-Roqueplo, A.-P., Bressac-de-Paillerets, B., Altindirek, D., Lorenzo, F., Lambert, F., Dan, H., Gad-Lapiteau, S., Catarina Oliveira, A., Rossi, C., Fraga, C., Taradin, G., Martin-Nuñez, G., Vitória, H., Diaz Aguado, H., Palmblad, J., Vidán, J., Relvas, L., Ribeiro, M. L., Luigi Larocca, M., Luigia Randi, M., Pedro Silveira, M., Percy, M., Gross, M., Marques da Costa, R., Beshara, S., Ben-Ami, T., Ugo, V., and ECE-Consortium. (2014) Genetic basis of congenital erythrocytosis: mutation update and online databases. *Hum. Mutat. 35*, 15–26.
(17) Moore, J. A., Hubbi, M. E., Wang, C., Wang, Y., Luo, W., Hofmann, S., and Rambally, S. (2020) Isolated Erythrocytosis Associated With 3 Novel Missense Mutations in the EGLN1 Gene. *Journal of Investigative Medicine High Impact Case Reports 8*, 2324709620947256.
(18) Gardie, B., Percy, M. J., Hoogewijs, D., Chowdhury, R., Bento, C., Arsenault, P. R., Richard, S., Almeida, H., Ewing, J., Lambert, F., McMullin, M. F., Schofield, C. J., and Lee, F. S. (2014) The role of PHD2 mutations in the pathogenesis of erythrocytosis. *Hypoxia (Auckl) 2*, 71–90.
(19) Yang, C., Zhuang, Z., Fliedner, S. M., Shankavaram, U., Sun, M. G., Bullova, P., Zhu, R., Elkahloun, A. G., Kourlas, P. J., Merino, M., Kebebew, E., and Pacak, K. (2015) Germ-line PHD1 and PHD2 mutations detected in patients with pheochromocytoma/paraganglioma-polycythemia. *J Mol Med (Berl) 93*, 93–104.
(20) Ansar, S., Malcolmson, J., Farncombe, K. M., Yee, K., Kim, R. H., and Sibai, H. (2022) Clinical implementation of genetic testing in adults for hereditary hematologic malignancy syndromes. *Genetics in Medicine*.
(21) Jang, J. H., Seo, J. Y., Jang, J., Jung, C. W., Lee, K. O., Kim, S. H., and Kim, H. J. (2014) Hereditary gene mutations in Korean patients with isolated erythrocytosis. *Ann Hematol*.
(22) Camps, C., Petousi, N., Bento, C., Cario, H., Copley, R. R., McMullin, M. F., van Wijk, R., Ratcliffe, P. J., Robbins, P. A., and Taylor, J. C. (2016) Gene panel sequencing improves the diagnostic work-up of patients with idiopathic erythrocytosis and identifies new mutations. *Haematologica 101*, 1306–1318.
(23) Percy, M. J., Zhao, Q., Flores, A., Harrison, C., Lappin, T. R., Maxwell, P. H., McMullin, M. F., and Lee, F. S. (2006) A family with erythrocytosis establishes a role for prolyl hydroxylase domain protein 2 in oxygen homeostasis. *Proc Natl Acad Sci U S A 103*, 654–9.
(24) Talbot, N. P., Smith, T. G., Balanos, G. M., Dorrington, K. L., Maxwell, P. H., and Robbins, P. A. (2017) Cardiopulmonary phenotype associated with human PHD2 mutation. *Physiol Rep 5*, e13224.
(25) Wilson, R., Syed, N., and Shah, P. (2016) Erythrocytosis due to PHD2 Mutations: A

Review of Clinical Presentation, Diagnosis, and Genetics. *Case Rep Hematol 2016*, 6373706.
(26) Barradas, J., Rodrigues, C. D., Ferreira, G., Rocha, P., Constanço, C., Andrade, M. R., Bento, C., and Silva, H. M. (2018) Congenital erythrocytosis – discover of a new mutation in the EGLN1 gene. *Clinical Case Reports 6*, 1109–1111.
(27) Percy, M. J., Furlow, P. W., Beer, P. A., Lappin, T. R., McMullin, M. F., and Lee, F. S. (2007) A novel erythrocytosis-associated PHD2 mutation suggests the location of a HIF binding groove. *Blood 110*, 2193–6.
(28) Ladroue, C., Carcenac, R., Leporrier, M., Gad, S., Le Hello, C., Galateau-Salle, F., Feunteun, J., Pouyssegur, J., Richard, S., and Gardie, B. (2008) PHD2 mutation and congenital erythrocytosis with paraganglioma. *N Engl J Med 359*, 2685–92.
(29) Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., and ACMG Laboratory Quality Assurance Committee. (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med 17*, 405–424.
(30) Metzen, E., Berchner-Pfannschmidt, U., Stengel, P., Marxsen, J. H., Stolze, I., Klinger, M., Huang, W. Q., Wotzlaw, C., Hellwig-Burgel, T., Jelkmann, W., Acker, H., and Fandrey, J. (2003) Intracellular localisation of human HIF-1 alpha hydroxylases: implications for oxygen sensing. *J Cell Sci 116*, 1319–26.
(31) Wenger, R. H., Stiehl, D. P., and Camenisch, G. (2005) Integration of oxygen signaling at the consensus HRE. *Sci STKE 2005*, re12.
(32) Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., and Collins, R. (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med 12*, e1001779.
(33) Van Hout, C. V., Tachmazidou, I., Backman, J. D., Hoffman, J. D., Liu, D., Pandey, A. K., Gonzaga-Jauregui, C., Khalid, S., Ye, B., Banerjee, N., Li, A. H., O'Dushlaine, C., Marcketta, A., Staples, J., Schurmann, C., Hawes, A., Maxwell, E., Barnard, L., Lopez, A., Penn, J., Habegger, L., Blumenfeld, A. L., Bai, X., O'Keeffe, S., Yadav, A., Praveen, K., Jones, M., Salerno, W. J., Chung, W. K., Surakka, I., Willer, C. J., Hveem, K., Leader, J. B., Carey, D. J., Ledbetter, D. H., Geisinger-Regeneron DiscovEHR Collaboration, Cardon, L., Yancopoulos, G. D., Economides, A., Coppola, G., Shuldiner, A. R., Balasubramanian, S., Cantor, M., Regeneron Genetics Center, Nelson, M. R., Whittaker, J., Reid, J. G., Marchini, J., Overton, J. D., Scott, R. A., Abecasis, G. R., Yerges-Armstrong, L., and Baras, A. (2020) Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature 586*, 749–756.
(34) Hirsilä, M., Koivunen, P., Xu, L., Seeley, T., Kivirikko, K. I., and Myllyharju, J. (2005) Effect of desferrioxamine and metals on the hydroxylases in the oxygen sensing pathway. *FASEB J. 19*, 1308–1310.
(35) Koivunen, P., Hirsila, M., Kivirikko, K. I., and Myllyharju, J. (2006) The length of peptide substrates has a marked effect on hydroxylation by the hypoxia-inducible factor prolyl 4-hydroxylases. *J Biol Chem 281*, 28712–20.
(36) Koivunen, P., and Myllyharju, J. (2018) Kinetic Analysis of HIF Prolyl Hydroxylases. *Methods Mol. Biol. 1742*, 15–25.
(37) Gaildrat, P., Killian, A., Martins, A., Tournier, I., Frebourg, T., and Tosi, M. (2010) Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol 653*, 249–57.
(38) DeLaForest, A., Nagaoka, M., Si-Tayeb, K., Noto, F. K., Konopka, G., Battle, M. A.,

and Duncan, S. A. (2011) HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. *Development 138*, 4143–4153.
(39) Si-Tayeb, K., Noto, F. K., Nagaoka, M., Li, J., Battle, M. A., Duris, C., North, P. E., Dalton, S., and Duncan, S. A. (2010) Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology 51*, 297–305.
(40) 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):9662–74.

Supplementary Table 1

Patient ID	Nucleotide exchange	Protein effect	Exon	Age at prognosis	Age at diag	Sex	Hb g/dl	Ht %	RBC M/mm³	EPO mU/ml	Clinical manifestations	Ref
	c.12C>A	p.Asp4Glu#	1								No polycythemia but adaptation to high altitude	1
	c.112A>G	p.Ser38Gly	1				17.6			5	Erythrocytosis but VUS, Late Onset	2
	c.122A>G	p.Tyr41Cys	1								Erythrocytosis	2
Patient 3	c.122_124delACT	p.Tyr41del	1		80	M	20.7	60		7.6	Erythrocytosis	3
Proband	c.124T>C	p.Cys42Arg, Hmz	1	14		F	18	54	6.46	6	Erythrocytosis, abdominal pain, fatigue	4
Brother	c.124T>C	p.Cys42Arg, Hmz	1			M	17.9	52	6.4	6.3	Erythrocytosis	
Father	c.124T>C	p.Cys42Arg, Ht	1			M	15.9	48	5.8	10	Healthy	
Mother	c.124T>C	p.Cys42Arg, Ht	1			F	13.9	44	4.4	6.7	Healthy	
	c.153G>A	p.Trp51*	1	17		F	14.5	43.4	5.09	-	No erythrocytosis, pheochromocytoma, chronic myeloid leukemia	5
	c.165G>C	p.Lys55Asn	1				14.1-16.2			9.5	Erythrocytosis but VUS, Late Onset, HA, Phleb (2 patients)	2
	c.298C>T	p.Arg100Trp	1				19				Erythrocytosis but VUS	2
	c.311C>T	p.Ser104Phe	1								Erythrocytosis but VUS	2
	c.359C>T	p.Pro120Leu	1				17.1			4.4-13	Erythrocytosis but VUS, (2 patients)	2
	c.380G>C	p.Cys127Ser#	1								No erythrocytosis, adaptation to high altitude	1,6–8
10	c.380G>C	p.Cys127Ser#	1	36		F	16.8	45.4	4.98	6	Erythrocytosis	9
58	c.380G>C	p.Cys127Ser#	1	34		M	18.8	52.5	5.54	11	Erythrocytosis, High ferritin level Heterozygous *HFE* H63D	9
69	c.380G>C	p.Cys127Ser#	1	23		M	18.2	51.1	5.7	5	Erythrocytosis	9
	c.400C>T	p.Gln134*	1		56	M	18.4	54	6.1	6.3-10.6	Erythrocytosis, plethoric, familial history, Phleb/Aspirin	10
	c.461C>A	p.Ser154*	1	67		M	17.8	50.7		10.3	Erythrocytosis, cerebrovascular accident, familial history, no history of thrombosis, Aspirin	2, 11
	c.471G>C	p.Gln157His#	1								Erythrocytosis	12
8416	c.471G>C	p.Gln157His#	1	44	43	M	19.3-20.2	57		24	Erythrocytosis, Renal cysts	13
	c.494delC	p.Pro165GlnfsX9	1				17.1			51	Erythrocytosis, PV thrombosis, HA, Splenomegaly, Phleb	2
	c.494dupC	p.Ser166LysfsX81	1				16.3			17.6	Erythrocytosis	2
	c.586G>C	p.Glu196Gln	1				20.2			10.9	Erythrocytosis but VUS, Late Onset/Acquired	2
424	c.599C>A	p.Pro200Gln	1	34	22	M	17.9	56	5.9	90	Erythrocytosis, Renal transplantation	13
Father	c.606delG	p.Met202Ilefs*72	1	Died at 76		M					Erythrocytosis	14
Brother 1			1	54		M	19.2	56		20	Erythrocytosis, Inflammatory arthomyalgia, visual symptoms, arterial hypertension, no history of thrombosis, Phleb	
Brother 2			1									
Patient 1	c.609C>G	p.Asn203Lys	1			M	23	68	6.03	2	Erythrocytosis, PV= *JAK2* exon 12 547insL+154F547dup8, no familial history, no history of thrombosis, Phleb/Aspirin	15
	c.610G>A	p.Lys204Glu	1	49	46	M	22.1	64		20.1	Erythrocytosis, Cardiac disease	16
	c.616G>T	p.Gly206Cys	1				16.5			14	Erythrocytosis but VUS, Late Onset/Acquired, congenital hemangioma	2
Patient 3	c.616G>T	p.Gly206Cys	1	28		F	17.2			14	Erythrocytosis, extremity hemangioma, asymptomatic	17
	c.661C>T	p.Gln221*	1	52	51	M	20.2	56.3	6.3	15	Erythrocytosis, Klinefelter syndrome, hypertension, type II diabetes	18
Pt71	c.678dupG	p.Arg227Alafs*20	1	34	24	M	18.3	55	6.3	8.13	Erythrocytosis, Headaches	
Brother				48		M	17.2	50	5.5			
Sister				43		F	17.4	52	5.26			
	c.682G>T	p.Ala228Ser	1	60	16	F	13.9	44.5	5.9	40.5	Erythrocytosis, pheochromocytoma and multiple paragangliomas, Headaches, chest pain, 3 miscarriages, Phleb	19
Pt72	c.715C>T	p.Gln239*	1	58	52	M	20	54		11	Erythrocytosis and hypertension	18
	c.715C>T	p.Gln239*	1			F					Erythrocytosis and adrenal adenoma	20
	c.729_730dup	p.Lys244ArgfsX31	1	73							Erythrocytosis	2
2412	c.760G>C	p.Asp254His	1	48	25	M	19.2	57.2	6.3	36	Erythrocytosis	13
	c.806T>C	p.Ile269Thr	1				18.8			14.1	Erythrocytosis but VUS, Asymptomatic, Neuroendocrine tumor	2
	c.815T>C	p.Leu272Pro	1				17.1			9.4	Erythrocytosis, Phleb	2
	c.826A>G	p.Met276Val	1				18.1				Erythrocytosis but VUS	2
IE-45	c.835del14	p.L279Tfs43*	1	73		M	18.8			1.3	Erythrocytosis	21
	c.836T>C	p.Leu279Pro	1		52	M	19.8	61.2	6.42	12.2	Erythrocytosis, Headaches diziness, familial history	22
Case 2	c.840dupA	p.Arg281Thrfs*4	1	22		M	17.8	50.9			Erythrocytosis, Tinnitus, no history of thrombosis	14
	c.853G>C	p.Gly285Arg	1	68	65	M	16.6	49		N	Erythrocytosis	16

Supplementary Table 1 (continued)

Patient ID	Nucleotide exchange	Protein effect	Exon	Age at prognosis	Age at diag	Sex	Hb g/dl	Ht %	RBC M/mm³	EPO mU/ml	Clinical manifestations	Ref
	c.867C>G	p.Ser289Arg	1				17			9.1	Erythrocytosis, Familial history, Phleb	2
	c.872A>T	p.Lys291Ile	1	38	29	M	17.6	52		5	Erythrocytosis, Familial history, no history of thrombosis	15
	c.892-3C>Gn /c.1011+3A>G		/								Erythrocytosis but VUS, Asymptomatic, familial history, Phleb	
	c.896T>C	p.Met299Thr	2				14.1-15.8			38	Erythrocytosis, cerebrovascular accident, Familial history, Splenomegaly, Phleb/Aspirin (2 patients)	2
	c.911C>T	p.Pro304Leu	2	Died at 70	48	M	20.8	63	5.8	268	Erythrocytosis, Leucoclastis vasculitis	18
					48	M	18.3	55	4.3	60	Erythrocytosis	
	c.911C>T	p.Pro304Leu	2							20	Erythrocytosis	2
	c. 949_950delinsAG	p.Pro317Arg	2				16.5			7.1	Asymptomatic	2
Father	c.950C>G	p.Pro317Arg	2	Died at 61	45	M	18	53	6.4		Erythrocytosis, Esophageal carcinoma (smoker), Phleb	23, 24
Daughter			2		26	F	18			6.2	Erythrocytosis concomittant with menorrhagia, Superficial thrombophlebitis	
Son			2		30	M	17.5-19.1	52-54	6.1	8.6	Erythrocytosis, Parasthesia, enhanced pulmonary vascular and ventilatory response to acute hypoxia	
	c.977G>A	p.Cys326Tyr	2				18.7			6	Erythrocytosis but VUS, Asymptomatic, Phleb	2
26	c.1000T>C	p.Trp334Arg	2	31		F	17.4	51	5.15	53	Erythrocytosis	9
	c.1001G>A	p.Trp334*	2				17.8				Erythrocytosis	2
	c.1001G>A	p.Trp334*	2	49	55	M	21.5			4-16	Erythrocytosis, Familial history (sister), Phleb	25
	c.1010dup	p.Val338Glyfs*18	2	24	21	M	17.1	47	5.5	9.9	Erythrocytosis, TIA	16
	c.1011+3A>G		/				18.1				Erythrocytosis but VUS	2
	c.1016G>C	p.Ser339Thr	3								Erythrocytosis but VUS	2
	c.1024A>G	p.Ile342Val	3				18.2				Erythrocytosis but VUS	2
	c.1030C>T	p.Arg344*	3	60		F	17			30	Erythrocytosis, hypertension, no history of thrombosis, Phleb/Aspirin	2, 11
	c.1031G>A	p.Arg344Gln	3				17				Erythrocytosis but VUS, Asymptomatic	2
	c.1079T>C	p.Phe360Ser	3				16.8				Erythrocytosis but VUS	2
Patient 2	c.1088T>G	p.Leu363Arg	3	34		M	19.5	57	5.89	8.45	Erythrocytosis, no history of thrombosis, Phleb	3
	c.1096T>C	Phe366Leu	3				18.2	58		N	Erythrocytosis, Affecting three generations (grandfather, father (propositus), and son), hypertension, headache, plethoric face, Phleb/anti-platelet aggregation therapy	26
	c.1106A>C	p.Asp369Ala	3				18.9				Erythrocytosis but VUS	2
	c.1109G>A	p.Arg370His	3				18.4				Erythrocytosis but VUS	2
	c.1111C>T	p.Arg371Cys	3	47		M	16.8	53	6.6	9.5	Erythrocytosis	18
	c.1111C>T	p.Arg371Cys	3			F	17.2	52.6		11.2	Familial history, no history of thrombosis, Phleb	11
	c.1112G>A	p.Arg371His	3	44	29	M	18.8	56		12	Erythrocytosis, Sagittal sinus thrombosis, Phleb	27
2403	c.1112G>A	p.Arg371His	3	25	17	M	19.1	56.7	6.4	N	Erythrocytosis	13
2295	c.1121A>G	p.His374Arg	3	52	30	M	20.2	61.6	6.2	18	Erythrocytosis, Hypertension, Paraganglioma, Homozygous HFE C282Y	28
Case 3	c.1129C>T	p.Gln377*	3	35		F	17.8	54.7		10.7	Erythrocytosis, no history of thrombosis, Phleb	14
	c.1132C>T	p.Pro378Ser	3				15.6			11.5	Erythrocytosis	2
Patient 1	c.1132C>T	p.Pro378Ser	3	60	47	F	16-17			11	Erythrocytosis, No thrombotic event	17
	c.1153G>A	p.Ala385Thr	4				17.3-18.6			9.2	Erythrocytosis (2 patients)	2
Patient 2	c.1153G>A	p.Ala385Thr	4	52		M	18-19			9	Erythrocytosis, headaches, parasthesias, Phleb	17
	c.1167G>T	p.Trp389Cys	4				17.5			4.9	Erythrocytosis	2
1406	c.1192C>T	p.Arg398*	4	41	26	M	19.3	53.8	5.9	6.5	Erythrocytosis (4 patients)	13
1406's mother		p.Arg398*, mosaic		67	64	F	16.1	49.5	5.2		Uterin leiomyoma, (51 yrs), Suspected renal and liver angioma	
	c.1192C>T	p.Arg398*	4				16-16.4			10.3-17.8	Erythrocytosis	2
Patient 3	c.1267A>G	p.Lys423Glu	5	80	60	M	16.4	51.8		23	Erythrocytosis, no familial history, no history of thrombosis, Phleb/Aspirin	15
	c.*16A>G		/				17.7			8.9	Erythrocytosis but VUS, Asymptomatic, Thalassemia minor	2
	c.*68G>A		/								Erythrocytosis but VUS	2
	c.*92G>A		/				15.3-17.7			9.4-10	Erythrocytosis but VUS, HA, (3 patients)	2

Supplementary Table 2

A Clinical data of patients carrying variants classified as pathogenic

ID	Position cDNA	Position Protein	sex	Hb (g/L)	Hct %	RBC M/mm3	EPO mU/ml	Classification
P#3	c.122A>G	p.Tyr41Cys	M	19.2	57		9	Likely Pathogenic
P#7	c.400C>T	p.Gln134*	M	18.4	54	6.1	6.03	Pathogenic
P#9	c.489C>A	p.Tyr163*	M	20.5	58	6.83	9.2	Pathogenic
P#12	c.661C>T	p.Gln221*	F	16.7	53	5.09		Likely Pathogenic
P#15	c.715C>T	p.Gln239*	F	18	55.4	5.7	10	Pathogenic
P#19	c.806T>C	p.Ile269Thr	M	20.4	61.5	7	6.6	Pathogenic
P#20			M	17.1	52.6	6.23	3.5	
P#21			M	18.3	55.6	6.11	6	
P#22			M	18.6	52	6.11	17.9	
P#23			M	17.1	51.5		5.8	
P#24	c.808_811dup	p.Leu271Argfs*15	M	16.8	48.5	5.87		Pathogenic
P#26	c.891+1G>A	NA	M	18.8	54	5.58	14.6	Likely Pathogenic
P#28	c.935G>A	p.Arg312His	M	18.5	52	5.95	5	Likely Pathogenic
P#29			F	15.6	47	5.85	45.5	
P#30			M	20.1	61.5	6.54	14.6	
P#31			M					
P#33	c.990dup	p. Asn331*	M	20.7	62		N	Pathogenic
P#34	c.1000T>C	p.Trp334Arg	M	17.6	49.9			Pathogenic
P#40	c.1129C>T	p.Gln377*	F	18	53.3	5.87	69	Pathogenic
P#41	c.1152C>T	p.Tyr384Tyr	M	13.4	42.8		91	Likely Pathogenic
P#42	c.1165T>C	p.Trp389Arg	M	20	56	6.08	2.4	Pathogenic
P#43			F	18.1	54		50	
P#44	c.1216+1G>T	NA	F	12.9	44	5.91	24.1	Likely Pathogenic
P#47	Deletion of the entire gene	NA	F	15.6	47.4			Pathogenic
		MEAN VALUES MEN:		**18.47**	**54.3**	**6.2**	**14.74**	n=16
		RANGE MEN:		13.4-20.7	42.8-62	5.58-7	2.4-91	
		Normal Values Men:		>18	>52	>5.7	5-25	
		MEAN VALUES WOMEN		**16.4**	**50.58**	**5.68**	**39.72**	n=7
		RANGE WOMEN:		12.9-18.1	44-55.4	5.09-5.9	10-69	
		Normal Values Women:		>16	>47	>5.2	5-25	

B Clinical data of patients carrying variants classified as VUS/Benign

ID	Position cDNA	Position Protein	sex	Hb (g/L)	Hct %	RBC M/mm3	EPO mU/ml	Classification
P#1	c.-410G>T	NA	M	17	51	6.05	17	VUS
P#2	c.104G>A	p.Arg35His	M	17.1	51.2	5.71	N	VUS
P#4	c.148G>C	p.Asp50His	M	19.9	54.8	6.17	17	VUS
P#5	c.230C>T	p.Pro77Leu	M					VUS
P#6	c.287C>T	p.Ala96Val	F	22.6	65.6	6.97	6.3	Lik. Ben.
P#8	c.470A>G	p.Gln157Arg	M	20.2	57.9	5.07	3.2	Lik. Ben.
P#10	c.568_569delinsTT	p.Ala190Leu	M	15	51.8		30	VUS
P#11	c.629_631delTGG	p.Val210del	F	16.7	49			VUS
P#13	c.665T>G	p.Ile222Ser	M					VUS
P#14	c.698G>C	p.Gly233Ala	M	17.4	49		10.4	VUS
P#16	c.763A>G	p.Lys255Glu	F	16.9	52			VUS
P#17	c.794_795delinsTT	p.Gly265Val	M	15.4	46.6			VUS
P#18	c.803C>T	p.Thr268Ile	M					VUS
P#25	c.887C>T	p.Thr296Met	M	18.6	53.8	6.48	11.9	VUS
P#27	c.908A>G	p.Tyr303Cys	M	19.8	56	7.1		VUS
P#32	c.985T>C	p.Tyr329His	F	16.8	49.7	5.75	6.2	VUS
P#35	c.1045G>T	p.Gly349Cys	F	17.6	48-58	6.22	16.5	VUS
P#36	c.1045G>A	p.Gly349Ser	M	16.9	50.1			VUS
P#37	c.1072C>A	p.Pro358Thr	F	18.9	53.4	5.87	8.07	VUS
P#38	c.1096T>C	p.Phe366Leu	M	18.8	55.8	6.66	15	VUS
P#39	c.1108C>G	p.Arg370Gly	M	19.6	53.8		5	VUS
P#45	c.1257T>G	p.Asp419Glu	M		49.6	5.33	9.7	VUS
P#46	c.1259C>T	p.Ser420Leu	M	18	53			VUS
		MEAN VALUES MEN:		**18**	**52.4**	**6.07**	**13.2**	n=13
		RANGE MEN:		15-20.2	46.6-57.9	5.03-7.1	3.2-30	
		Normal Values Men:		>18	>52	>5.7	5-25	
		MEAN VALUES WOMEN		**18.3**	**54.6**	**6.2**	**9.3**	n=6
		RANGE WOMEN:		16.7-22.6	49-65.6	5.75-6.97	6.2-16.5	
		Normal Values Women:		>16	>47	>5.2	5-25	

Supplementary Table 3

Patient# in Present paper/ Ref	Exon	Nucleotide exchange	Protein effect	Frequency Gnomad V3	Prediction Metadome	Mobidetails Single predictors	Mobidetails Meta predictors	SPIP score	UK bio bank	ACMG criteria	Classification	Conclusion
P#1	5UTR	**c.-410G>T**	NA	0	NA	NA	NA	NA		PM1, PM2	Class 3	VUS
1	1	c.12C>A	p.Asp4Glu [#]	0	Highly tolerant (1.53)	0.295	0.302	2.51 %	ns	PM2, BP4	Class 3	VUS
P#2	1	**c.104G>A**	**p.Arg35His**	0	Intolerant (0.39)	0.870	0.776	2.51 %		PM2, PP3	Class 3	VUS
2	1	c.112A>G	p.Ser38Gly	0	Intolerant (0.27)	0.848	0.674	2.51 %		PM2, PP3	Class 3	VUS
P#3/ 2	1	**c.122A>G**	**p.Tyr41Cys**	0	Intolerant (0.23)	0.442	0.842	2.51 %		PM2, PS3	Class 4	Likely Pathogenic
3	1	c.122_124delACT	p.Tyr41del	6.995e-06	Highly intolerant (0.14)	NA	NA	2.51 %	>90th	PM2, PM4	Class 3	VUS
4	1	c.124T>C	p.Cys42Arg, Hmz	0	Highly intolerant (0.14)	0.471	0.867	2.51 %		PM2, PS3, PP1	Class 4	Likely Pathogenic
P#4	1	**c.148G>C**	**p.Asp50His**	0	Highly intolerant (0.05)	0.442	0.662	2.51 %		PM2	Class 3	VUS
5	1	c.153G>A	p.Trp51*	0	NA	NA	NA	2.51 %		PM2, PVS1, PP	Class 5	Pathogenic
2	1	c.165G>C	p.Lys55Asn	0	Highly intolerant (0.08)	0.391	0.693	2.51 %		PM2	Class 3	VUS
P#5	1	**c.230C>T**	**p.Pro77Leu**	1.976e-05	Slighly Intolerant (0.62)	0.292	0.235	2.51 %	nc		Class 3	VUS
P#6	1	**c.287C>T**	**p.Ala96Val**	1E-04	Tolerated (1.2)	0.304	0.204	2.51 %	ns	BS2, BP4	Class 2	Likely Benign
2	1	c.298C>T	p.Arg100Trp	0	Tolerated (1.19)	0.702	0.357	2.51 %	ns	PM2	Class 3	VUS
2	1	c.311C>T	p.Ser104Phe	0.0011	Tolerated (1.1)	0.384	0.200	2.51 %		BS1, BP4	Class 2	Likely Benign
2	1	c.359C>T	p.Pro120Leu	0.0001	Tolerated (1.24)	0.333	0.245	2.51 %	nc	BP4	Class 3	VUS
1,6–8	1	c.380G>C	p.Cys127Ser [#]	0.1536	Slightly intolerant	0.184	0.105	2.51 %	ns	BA1	Class 1	Benign
P#7/10Ψ	1	**c.400C>T**	**p.Gln134***	0	NA	NA	NA	2.51 %		PM2, PVS1, PP	Class 5	Pathogenic
2, 11	1	c.461C>A	p.Ser154*	6.983e-06	NA	NA	NA	2.51 %		PM2, PVS1, PP	Class 5	Pathogenic
P#8	1	c.470A>G	p.Gln157Arg	4E-04	Intolerant (0.34)	0.209	0.183	2.51 %	ns	BP4, BS2	Class 2	Likely Benign
12, 13	1	c.471G>C	p.Gln157His [#]	0.0180	Intolerant (0.34)	0.417	0.157	2.51 %	ns	BA1	Class 1	Benign
P#9	1	**c.489C>A**	**p.Tyr163***	0	NA	NA	NA	2.51 %		PM2, PVS1, PP	Class 5	Pathogenic
2	1	c.494delC	p.Pro165Glnfs*9	6.993e-06	NA	NA	NA	2.51 %	>99th	PM2, PVS1, PP	Class 5	Pathogenic
2	1	c.494dupC	p.Ser166Lysfs*81	0	NA	NA	NA	2.51 %		PM2, PVS1, PP	Class 5	Pathogenic
P#10	1	**c.568_569delinsTT**	**p.Ala190Leu**	0	Intolerant (0.42)	NA	NA	2.51 %		PM2	Class 3	VUS
2	1	c.586G>C	p.Glu196Gln	0	Intolerant (0.25)	0.316	0.236	2.51 %		PM2, BP4	Class 3	VUS
13	1	c.599C>A	p.Pro200Gln	0	Intolerant (0.19)	0.763	0.418	2.51 %		PM2, PS3	Class 3	VUS
14	1	c.606delG	p.Met202Ilefs*72	0	NA	NA	NA	2.51 %		PM2, PVS1, PP	Class 5	Pathogenic
15	1	c.609C>G	p.Asn203Lys	0	Highly intolerant	0.381	0.207	2.51 %		PM2	Class 3	VUS
16	1	c.610A>G	p.Lys204Glu	0	Highly intolerant (0.1)	0.249	0.345	2.51 %		PM2	Class 3	VUS
2, 17	1	c.616G>T	p.Gly206Cys	0	Highly intolerant	0.832	0.480	2.51 %		PM2	Class 3	VUS
P#11	1	**c.629_631delTGG**	**p.Val210del**	0	Highly intolerant	NA	NA	2.51 %		PM2, PP3	Class 3	VUS
P#12/16	1	**c.661C>T**	**p.Gln221***	6.979e-06	NA	NA	NA	2.51 %	>99th	PM2, PVS1, PP, PS4 moderate	Class 5	Pathogenic
P#13	1	**c.665T>G**	**p.Ile222Ser**	0	Intolerant (0.34)	0.796	0.475	2.51 %		PM2	Class 3	VUS
18	1	c.678dupG	p.Arg227Alafs*20	0	NA	NA	NA	3.48 %		PM2, PVS1, PP	Class 5	Pathogenic
19	1	c.682G>T	p.Ala228Ser	0	Intolerant (0.42)	0.230	0.414	3.48 %		PM2, PS3	Class 4	Likely Pathogenic
P#14	1	**c.698G>C**	**p.Gly233Ala**	6.977e-06	Intolerant (0.25)	0.829	0.423	3.48 %		PM2	Class 3	VUS
P#15/16, 18	1	**c.715C>T**	**p.Gln239***	0	NA	NA	NA	3.48 %		PM2, PVS1, PP, PS4 moderate	Class 5	Pathogenic
2	1	c.729_730dup	p.Lys244Argfs*31	0	NA	NA	NA	3.48 %		PM2, PVS1, PP	Class 5	Pathogenic
13	1	c.760G>C	p.Asp254His	0	Intolerant (0.24)	0.833	0.568	3.48 %		PM2, PS3	Class 4	Likely Pathogenic
P#16	1	**c.763A>G**	**p.Lys255Glu**	0	Intolerant (0.25)	0.329	0.128	3.48 %	nc	PM2, BP4	Class 3	VUS
P#17	1	**c.794_795delinsTT**	**p.Gly265Val**	0	Intolerant (0.28)	0.78	0.768	26.62 %		PM2, PP3	Class 3	VUS
P#18	1	**c.803C>T**	**p.Thr268Ile**	0	Intolerant (0.32)	0.701	0.492	3.43 %		PM2	Class 3	VUS
P#19-23/2	1	**c.806T>C**	**p.Ile269Thr**	6.985e-06	Intolerant (0.33)	0.866	0.828	3.43 %	>99th	PS3 , PP3, PS4 strong	Class 5	Pathogenic
P#24	1	**c.808_811dup**	**p.Leu271Argfs*15**	0	NA	NA	NA	3.43 %		PM2, PVS1, PP	Class 5	Pathogenic
2	1	c.815T>C	p.Leu272Pro	0	Intolerant (0.35)	0.873	0.748	5.1 %		PM2, PP3	Class 3	VUS
2	1	c.826A>G	p.Met276Val	0	Intolerant (0.48)	0.199	0.430	5.1 %		PM2	Class 3	VUS
21	1	c.835del14	p.L279Tfs43*	0	NA	NA	NA	NA		PM2, PVS1, PP	Class 5	Pathogenic

Supplementary Table 3 (continued)

Patient# in Present paper/ Ref	Exon	Nucleotide exchange	Protein effect	Frequency Gnomad V3	Prediction Metadome	Mobidetails Single predictors	Mobidetails Meta predictors	SPIP score	UK bio bank	ACMG criteria	Classification	Conclusion
22	1	c.836T>C	p.Leu279Pro	0	Intolerant (0.52)	0.861	0.870	4.35 %		PM2, PP3	Class 3	VUS
14	1	c.840dupA	p.Arg281Thrfs*4	0	NA	NA	NA	NA		PM2, PVS1, PP	Class 5	Pathogenic
16	1	c.853G>C	p.Gly285Arg	0	Slightly intolerant	0.439	0.677	5.03 %	nc	PM2	Class 3	VUS
2	1	c.867C>G	p.Ser289Arg	1.397e-05	Intolerant (0.36)	0.358	0.289	69.33 %	nc	PM2	Class 3	VUS
15	1	c.872A>T	p.Lys291Ile	0	Intolerant (0.36)	0.260	0.420	8.25 %		PM2	Class 3	VUS
P#25	1	**c.887C>T**	**p.Thr296Met**	0	Intolerant (0.32)	0.868	0.814	35.81 %	>95th	PM2, PP3	Class 3	VUS
P#26	Intron 1	c.891+1G>A	NA	0	NA	NA	NA	98.41 %		PM2, PVS1 strong	Class 4	Likely Pathogenic
2	Intron 1	c.892-3C>G	NA	0	NA	NA	NA	98.41 %		PM2, PP3	Class 3	VUS
2	2	c.896T>C	p.Met299Thr	0	Intolerant (0.39)	0.826	0.699	9.63 %		PM2, PP3	Class 3	VUS
P#27	2	**c.908A>G**	**p.Tyr303Cys**	0	Intolerant (0.49)	0.85	0.79	9.76 %		PM2, PP3	Class 3	VUS
2, 18	2	c.911C>T	p.Pro304Leu	0	Intolerant (0.52)	0.871	0.841	30.67 %		PM2, PP3, PS4 moderate	Class 3	VUS
P#28-31	2	**c.935G>A**	**p.Arg312His**	0	Neutral (0.77)	0.696	0.521	5.41 %		PM2, PS4, PP1	Class 4	Likely Pathogenic
2	2	c. 949_950delinsAG	p.Pro317Arg	0	Intolerant (0.34)	0.841	0.552	30.67 %		PM2, PP3, PP1, PS4 moderate	Class 5	Pathogenic
23	2	c.950C>G	p.Pro317Arg	0	Intolerant (0.34)	0.841	0.616	7.81 %		PM2, PP3, PP1, PS3, PS4 moderate	Class 5	Pathogenic
2	2	c.977G>A	p.Cys326Tyr	0	Highly intolerant	0.840	0.521	7.62 %		PM2	Class 3	VUS
P#32	2	**c.985T>C**	**p.Tyr329His**	0	Highly intolerant	0.85	0.822	7.62 %		PM2, PP3	Class 3	VUS
P#33	2	**c.990dup**	**p.Asn331***	0	NA	NA	NA	NA		PM2, PVS1, PP	Class 5	Pathogenic
P#34/9Ψ	2	**c.1000T>C**	**p.Trp334Arg**	0	Intolerant (0.19)	0.85	0.769	8.04 %		PM2, PS3, PP1 Supporting, PP3	Class 5	Pathogenic
2, 25	2	c.1001G>A	p.Trp334*	0	NA	NA	NA	8.04 %		PM2, PVS1, PP, PS4 moderate	Class 5	Pathogenic
16	2	c.1010dup	p.Val338Glyfs*18	0	NA	NA	NA	9.3 %		PM2, PVS1, PP	Class 5	Pathogenic
2	Intron 2	c.1011+3A>G	NA	0	NA	NA	NA	99.33 %		PM2, PP3	Class 3	VUS
2	3	c.1016G>C	p.Ser339Thr	0	Intolerant (0.2)	0.182	0.298	9.63 %		PM2, BP4	Class 3	VUS
2	3	c.1024A>G	p.Ile342Val	0	Intolerant (0.25)	0.359	0.163	6.8 %		PM2, BP4	Class 3	VUS
2, 11	3	c.1030C>T	p.Arg344*	0	NA	NA	NA	NA		PM2, PVS1, PP	Class 5	Pathogenic
2	3	c.1031G>A	p.Arg344Gln	0	Intolerant (0.21)	0.790	0.496	26.62 %		PM2	Class 3	VUS
P#35	3	**c.1045G>T**	**p.Gly349Cys**	0	Intolerant (0.31)	0.845	0.593	3.59 %		PM2	Class 3	VUS
P#36	3	**c.1045G>A**	**p.Gly349Ser**	0	Intolerant (0.31)	0.706	0.471	3.59 %		PM2	Class 3	VUS
P#37	3	**c.1072C>A**	**p.Pro358Thr**	0	Intolerant 0.27	0.853	0.805	7.81 %		PM2, PP3	Class 3	VUS
2	3	c.1079T>C	p.Phe360Ser	0	Intolerant (0.32)	0.752	0.578	7.8 %		PM2	Class 3	VUS
3	3	c.1088T>G	p.Leu363Arg	0	Intolerant (0.24)	0.852	0.793	5.1 %		PM2, PP3	Class 3	VUS
P#38/26	3	**c.1096T>C**	**p.Phe366Leu**	0	Intolerant (0.28)	0.827	0.791	4.35 %		PM2, PS4 moderate, PP3	Class 3	VUS
2	3	c.1106A>C	p.Asp369Ala	0	Intolerant (0.24)	0.868	0.838	5.03 %		PM2, PP3	Class 3	VUS
P#39	3	**c.1108C>G**	**p.Arg370Gly**	0	Intolerant (0.2)	0.743	0.573	5.03 %		PM2	Class 3	VUS
2	3	c.1109G>A	p.Arg370His	0	Intolerant (0.2)	0.756	0.543	5.03 %		PM2	Class 3	VUS
11, 18	3	c.1111C>T	p.Arg371Cys	0	Highly intolerant	0.848	0.692	5.03 %	>95th	PM2, PP3, PM5	Class 3	VUS
13, 27	3	c.1112G>A	p.Arg371His	0	Highly intolerant	0.814	0.656	5.03 %	>95th	PM2, PP3, PS3, PS4 moderate	Class 4	Likely Pathogenic
28	3	c.1121A>G	p.His374Arg	0	Intolerant (0.25)	0.868	0.879	7.62 %		PM2, PP3, PS3	Class 4	Likely Pathogenic
P#40/13	3	**c.1129C>T**	**p.Gln377***	0	NA	NA	NA	8.25 %		PM2, PVS1, PP, PS4 moderate	Class 5	Pathogenic
2, 17	3	c.1132C>T	p.Pro378Ser	0	Intolerant (0.35)	0.817	0.641	8.04 %		PM2, PP3	Class 3	VUS
P#41	4	**c.1152C>T**	**p.Tyr384Tyr**	0	NA	NA	NA	47.89 %		PM2, PS3	Class 4	Likely Pathogenic
2, 17	4	c.1153G>A	p.Ala385Thr	0	Intolerant (0.29)	0.850	0.631	9.63 %		PM2, PP3	Class 3	VUS
P#42-42	4	**c.1165T>C**	**p.Trp389Arg**	0	Intolerant (0.27)	0.85	0.77	9.76 %		PM2, PS3, PS4 Moderate, PP1, PP3	Class 5	Pathogenic
2	4	c.1167G>T	p.Trp389Cys	0	Intolerant (0.27)	0.849	0.736	9.76 %		PM2, PP3	Class 3	VUS
13	4	c.1192C>T	p.Arg398*	0	NA	NA	NA	NA		PM2, PS3, PP, PS4 moderate	Class 5	Pathogenic
P#44	4	**c.1216+1G>T**	NA	0	NA	NA	NA	98.41 %		PM2, PS3, PP	Class 5	Pathogenic
P#45	5	**c.1257T>G**	**p.Asp419Glu**	1.6E-05	Intolerant (0.38)	0.288	0.236	5.41 %	nc	PM2, BP4	Class 3	VUS
P#46	5	**c.1259C>T**	**p.Ser420Leu**	5E-04	Intolerant (0.47)	0.699	0.39	5.41 %	nc	BS1	Class 3	VUS
15	5	c.1267A>G	p.Lys423Glu	0	Intolerant (0.52)	0.831	0.648	7.81 %		PM2, PP3	Class 3	VUS
P#47	1-5	**Deletion of the entire gene**	**NA**	0	NA	NA	NA	NA	NA	PM2, PVS1, PP	Class 5	Pathogenic

Supplementary Table 4

A Clinical data of published patients carrying variants classified as pathogenic

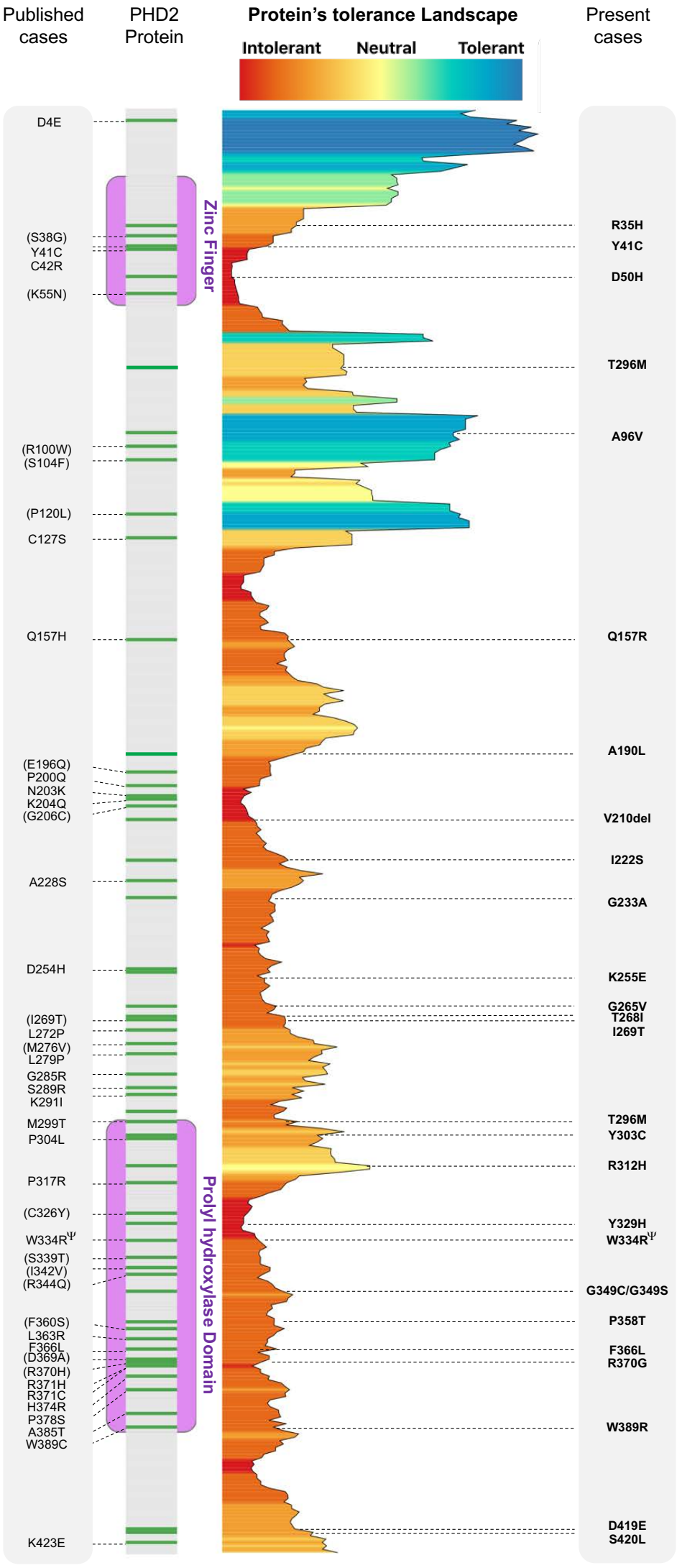
Position cDNA	Position Protein	Sex	Hb g/dl	Ht %	RBC M/mm^3	EPO mU/ml	Classification
c.122A>G	p.Tyr41Cys						Likely Pathogenic
c.124T>C	p.Cys42Arg, Hmz	F	18	54	6.46	6	Likely Pathogenic
	p.Cys42Arg, Hmz	M	17.9	52	6.4	6.3	
	p.Cys42Arg, Ht	M	15.9	48	5.8	10	
	p.Cys42Arg, Ht	F	13.9	44	4.4	6.7	
c.153G>A	p.Trp51*	F	14.5	43.4	5.09		Pathogenic
c.461C>A	p.Ser154*	M	17.8	50.7		10.3	Pathogenic
c.494delC	p.Pro165Glnfs*9		17.1			51	Pathogenic
c.494dupC	p.Ser166Lysfs*81		16.3			17.6	Pathogenic
c.606delG	p.Met202Ilefs*72	M					Pathogenic
		M	19.2	56		20	
c.661C>T	p.Gln221*	M	20.2	56.3	6.3	15	Pathogenic
c.678dupG	p.Arg227Alafs*20	M	18.3	55	6.3	8.13	Pathogenic
		M	17.2	50	5.5		
		F	17.4	52	5.26		
c.682G>T	p.Ala228Ser	F	13.9	44.5	5.9	40.5	Likely Pathogenic
c.715C>T	p.Gln239*	M	20	54		11	Pathogenic
c.729_730dup	p.Lys244Argfs*31						Pathogenic
c.760G>C	p.Asp254His	M	19.2	57.2	6.3	36	Likely Pathogenic
c.806T>C	p.Ile269Thr		18.8			14.1	Pathogenic
c.835del14	p.L279Tfs43*	M	18.8			1.3	Pathogenic
c.840dupA	p.Arg281Thrfs*4	M	17.8	50.9			Pathogenic
c. 949_950delinsAG	p.Pro317Arg		18.7			6	Pathogenic
c.950C>G	p.Pro317Arg	M	18	53	6.4		Pathogenic
		F	18			6.2	
		M	17.5-19.1	52-54	6.1	8.6	
c.1001G>A	p.Trp334*		17.8				Pathogenic
		M	21.5			4-16	
c.1010dup	p.Val338Glyfs*18	M	17.1	47	5.5	9.9	Pathogenic
c.1030C>T	p.Arg344*	F	17			30	Pathogenic
c.1112G>A	p.Arg371His	M	18.8	56			Likely Pathogenic
		M	19.1	56.7	6.4		
c.1121A>G	p.His374Arg	M	20.2	61.6	6.2	18	Likely Pathogenic
c.1129C>T	p.Gln377*	F	17.8	54.7		10.7	Pathogenic
c.1192C>T	p.Arg398*	M	19.3	53.8	5.9	6.5	Pathogenic
	MEAN VALUES Men:		**18.75**	**53.6**	**6.1**	**12.2**	n=19
	RANGE Men:		15.9-21.5	48-61.6	4.4-6.46	1.3-36	
	Normal Men:		>18	>52	>5.7	5-25	
	MEAN VALUES Women:		**16.1**	**48.76**	**5.42**	**16.68**	
	RANGE Women:		13.9-18	43.4-54.7	4.4-6.46	6-40.5	n=8
	Normal Women:		>16	>47	>5.2	5-25	

B Clinical data of published patients carrying variants classified as VUS/Benign

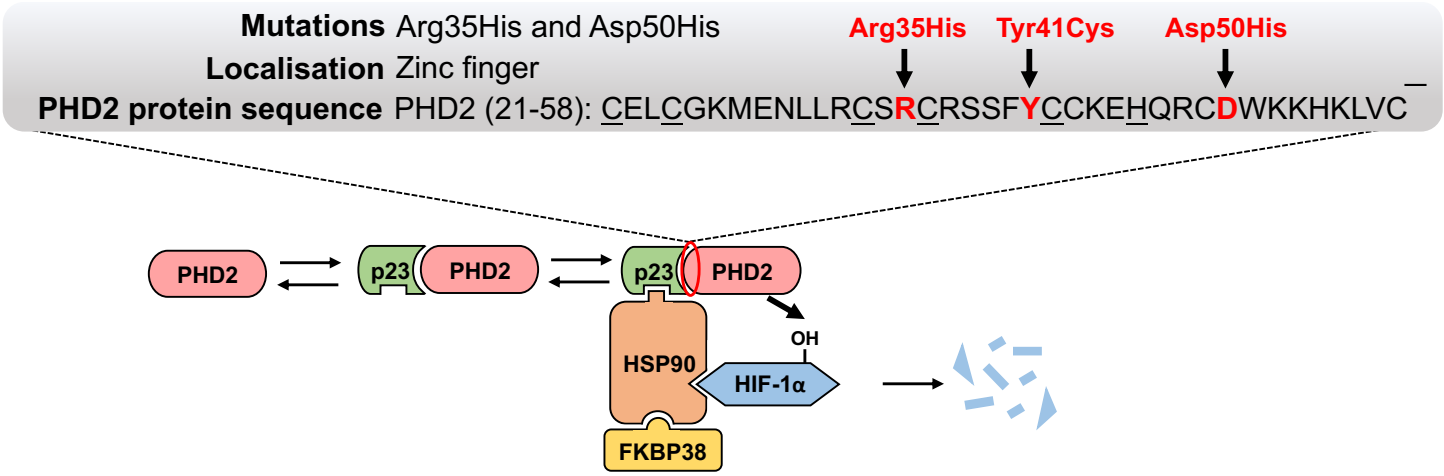
Position cDNA	Position Protein	Sex	Hb g/dl	Ht %	RBC M/mm^3	EPO mU/ml	Classification
c.112A>G	p.Ser38Gly		17.6			5	VUS
c.122_124delACT	p.Tyr41del	M	20.7	60		7.6	VUS
c.165G>C	p.Lys55Asn		14.1-16.2			9.5	VUS
c.298C>T	p.Arg100Trp		19				VUS
c.311C>T	p.Ser104Phe						Likely Benign
c.359C>T	p.Pro120Leu		17.1			4.4-13	VUS
c.471G>C	p.Gln157His	M	19.3-20.2	57		24	Likely Benign
c.586G>C	p.Glu196Gln		20.2			10.9	VUS
c.599C>A	p.Pro200Gln	M	17.9	56	5.9	90	VUS
c.609C>G	p.Asn203Lys	M	23	68	6.03	2	VUS
c.610G>A	p.Lys204Glu	M	22.1	64		20.1	VUS
c.616G>T	p.Gly206Cys		16.5			14	VUS
c.616G>T	p.Gly206Cys	F	17.2			14	
c.815T>C	p.Leu272Pro		17.1			9.4	VUS
c.826A>G	p.Met276Val		18.1				VUS
c.836T>C	p.Leu279Pro	M	19.8	61.2	6.42	12.2	VUS
c.853G>C	p.Gly285Arg	M	16.6	49		N	VUS
c.867C>G	p.Ser289Arg		17			9.1	VUS
c.872A>T	p.Lys291Ile	M	17.6	52		5	VUS
c.896T>C	p.Met299Thr		14.1-15.8			38	VUS
c.911C>T	p.Pro304Leu	M	20.8	63	5.8	268	VUS
		M	18.3	55	4.3	60	
c.911C>T	p.Pro304Leu					20	
c.977G>A	p.Cys326Tyr		18.7			6	VUS
c.1024A>G	p.Ile342Val		18.2				VUS
c.1031G>A	p.Arg344Gln		17				VUS
c.1079T>C	p.Phe360Ser		16.8				VUS
c.1088T>G	p.Leu363Arg	M	19.5	57	5.89	8.45	VUS
c.1096T>C	Phe366Leu		18.2	58		N	VUS
c.1106A>C	p.Asp369Ala		18.9				VUS
c.1109G>A	p.Arg370His		18.4				VUS
c.1111C>T	p.Arg371Cys	M	16.8	53	6.6	9.5	VUS
c.1111C>T	p.Arg371Cys	F	17.2	52.6		11.2	
c.1132C>T	p.Pro378Ser		15.6			11.5	VUS
c.1132C>T	p.Pro378Ser	F	16-17			11	
c.1153G>A	p.Ala385Thr		17.3-18.6			9.2	VUS
c.1153G>A	p.Ala385Thr	M	18-19			9	
c.1167G>T	p.Trp389Cys		17.5			4.9	VUS
c.1267A>G	p.Lys423Glu	M	16.4	51.8		23	VUS
	MEAN VALUES Men:		**19.12**	**57.7**	**5.84**	**43.2**	n=13
	RANGE Men:		16.4-23	49-54	4.3-6.6	2-268	
	Normal Men:		>18	>52	>5.7	5-25	
	MEAN VALUES Women:		**19.9**	**/**	**/**	**12.06**	n=3
	RANGE Women:		16.5-17.2			11-14	
	Normal Women:		>16	>47	>5.2	5-25	

Supplementary Figure 1

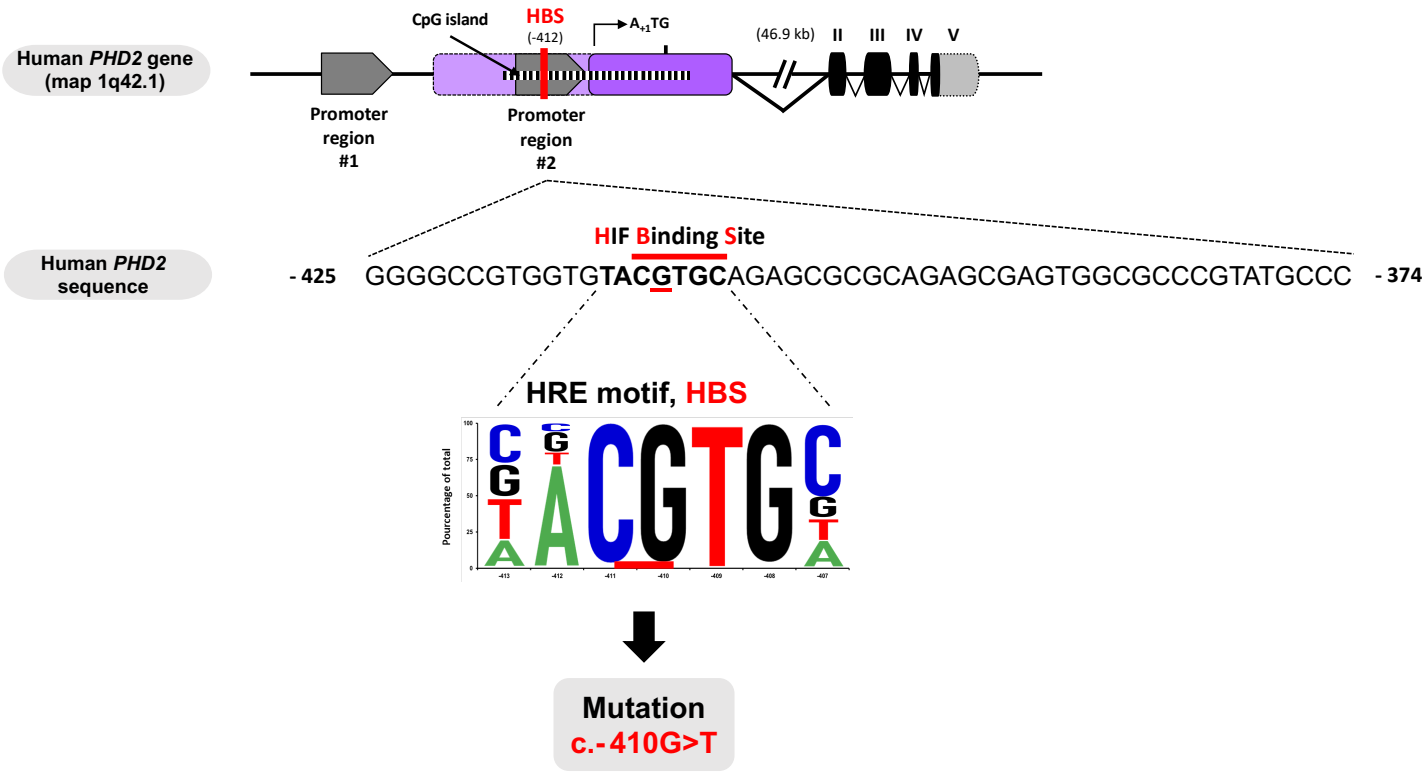
A



B Arg35His, Tyr41Cys and Asp50His: located in the Zinc finger domain

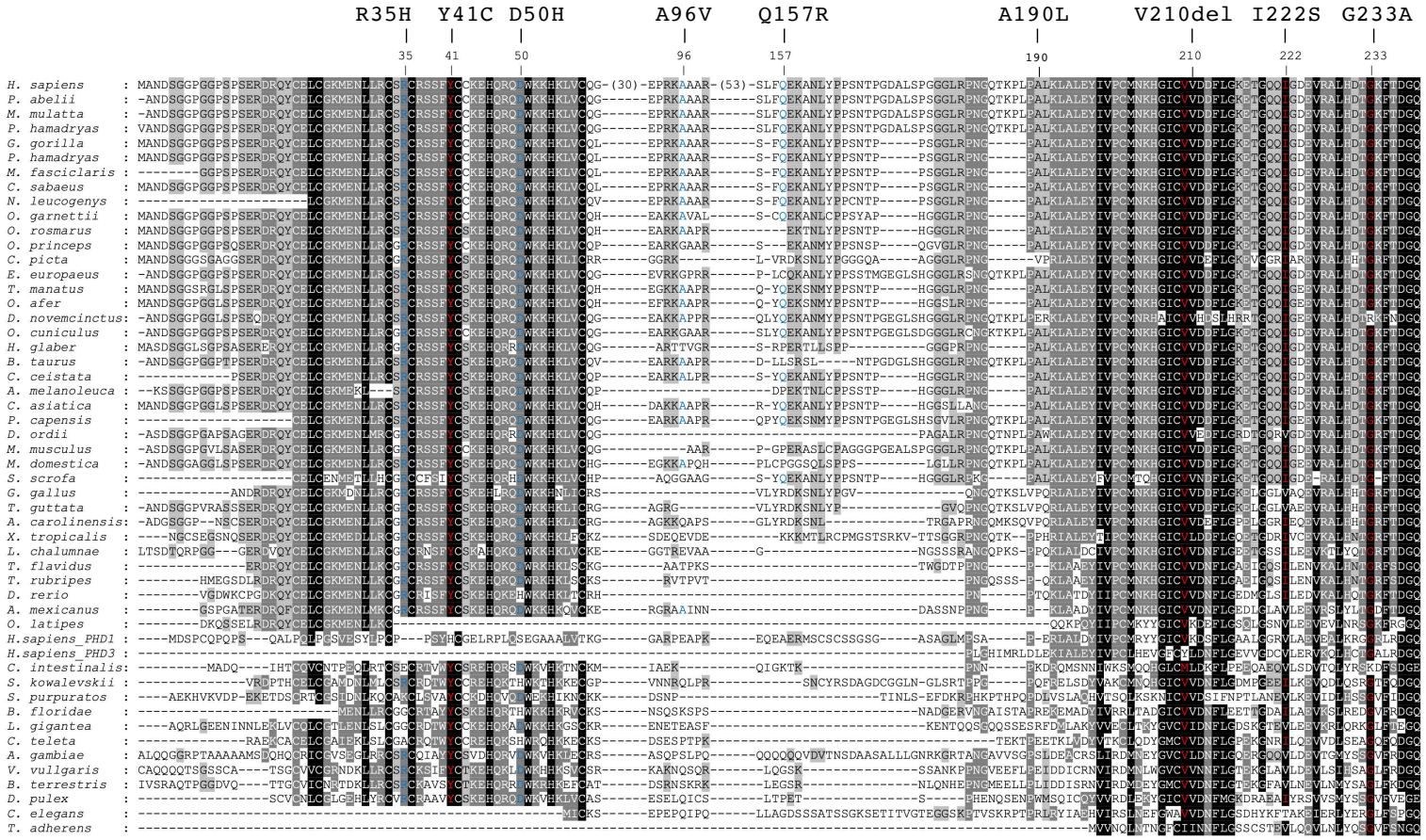


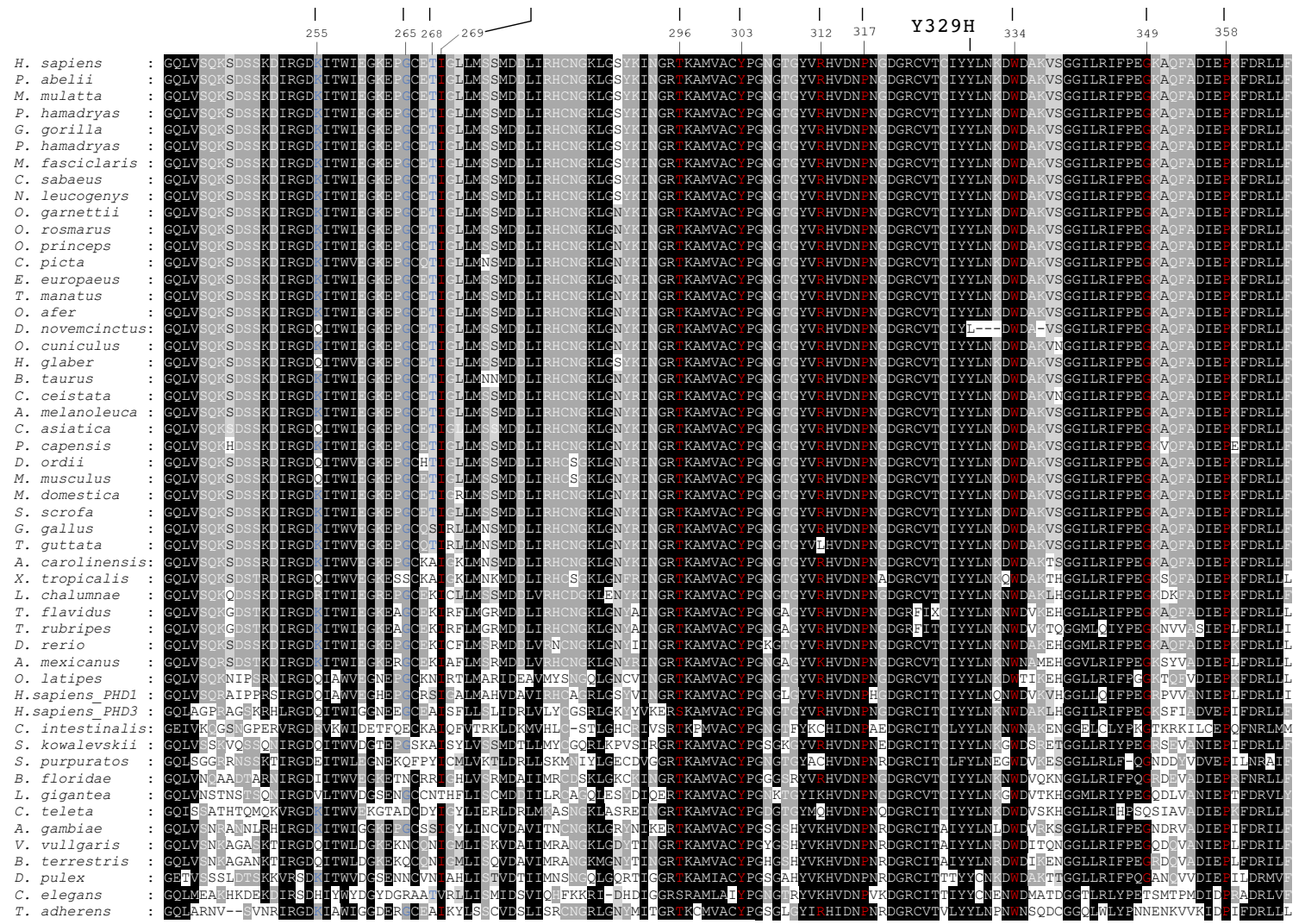
C c.-410G>T: located in the promotor

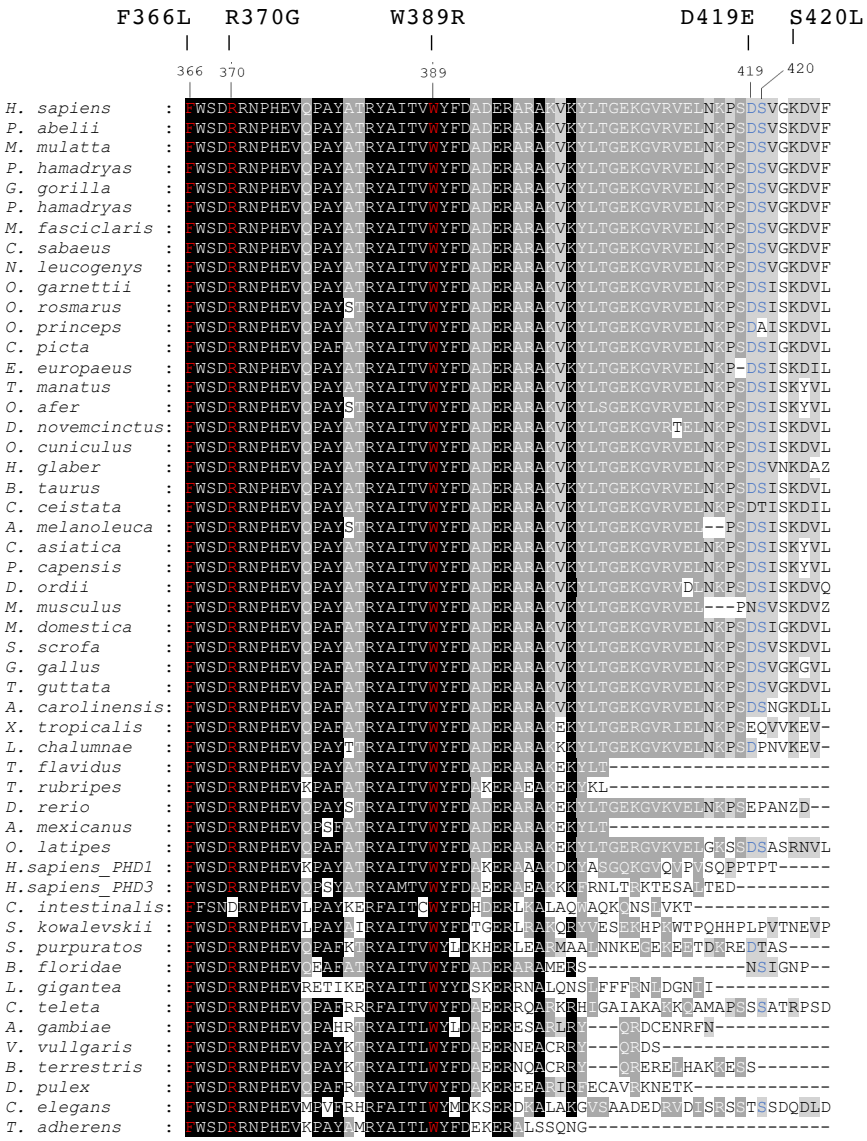


Supplementary Figure 2

Conservation

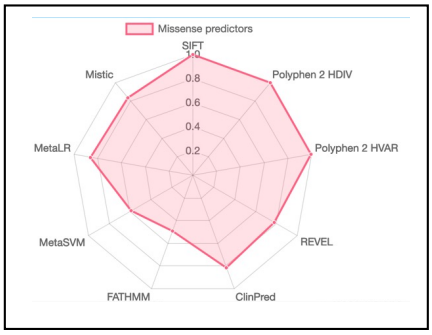




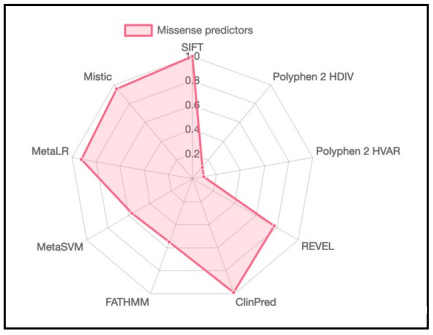

F366L
R370G
W389R
D419E
S420L
366
370
389
419
420
H. sapiens
P. abelii
M. mulatta
P. hamadryas
G. gorilla
P. hamadryas
M. fasciclaris
C. sabaeus
N. leucogenys
O. garnettii
O. rosmarus
O. princeps
C. picta
E. europaeus
T. manatus
O. afer
D. novemcinctus
O. cuniculus
H. glaber
B. taurus
C. ceistata
A. melanoleuca
C. asiatica
P. capensis
D. ordii
M. musculus
M. domestica
S. scrofa
G. gallus
T. guttata
A. carolinensis
X. tropicalis
L. chalumnae
T. flavidus
T. rubripes
D. rerio
A. mexicanus
O. latipes
H.sapiens_PHD1
H.sapiens_PHD3
C. intestinalis
S. kowalevskii
S. purpuratos
B. floridae
L. gigantea
C. teleta
A. gambiae
V. vullgaris
B. terrestris
D. pulex
C. elegans
T. adherens

Supplementary Figure 3

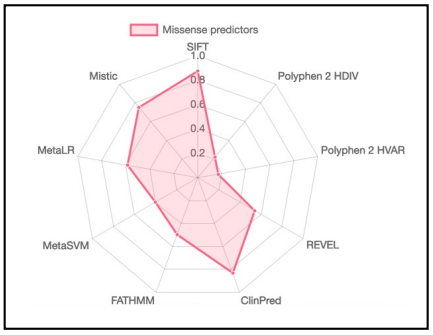
p.Arg35His (R35H)



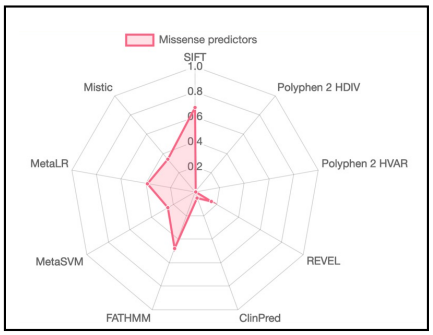
p.Tyr41Cys (Y41C)



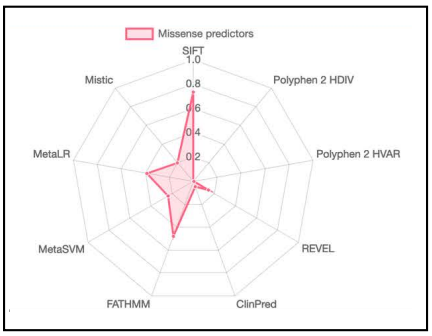
p.Asp50His (D50H)



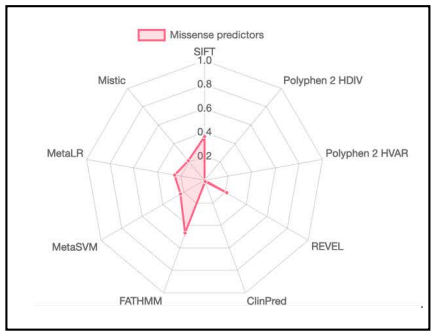
p.Pro77Leu (P77L)



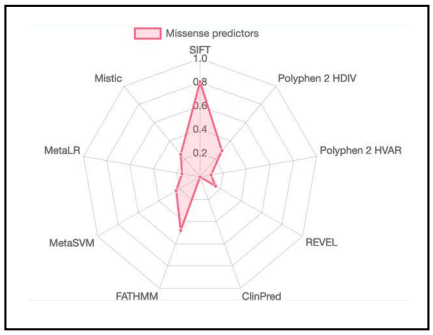
p.Ala96Val (A96V)



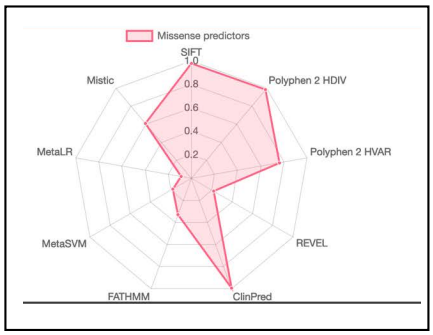
p.Gln157Arg (Q157R)



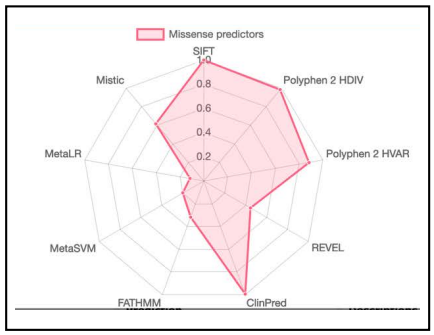
p.Gln157His (Q157H)



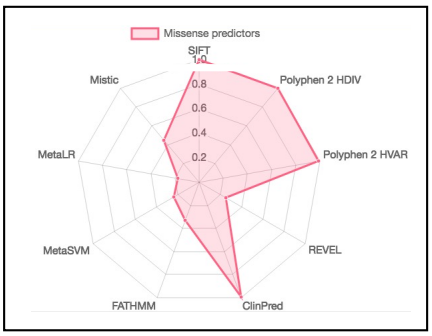
p.Pro200Gln (P200Q)



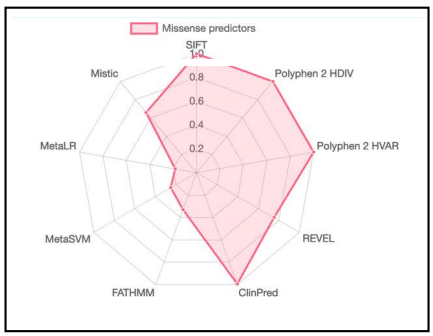
p.Ile222Ser (I222S)



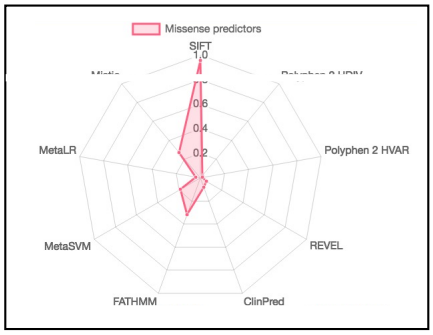
p.Gly233Ala (G233A)



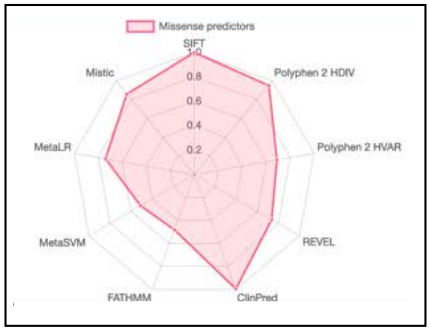
p.Asp254His (D254H)



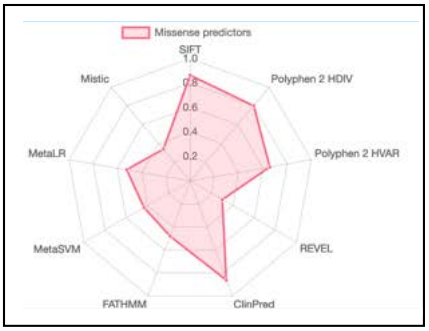
p.Lys255Glu (K255E)



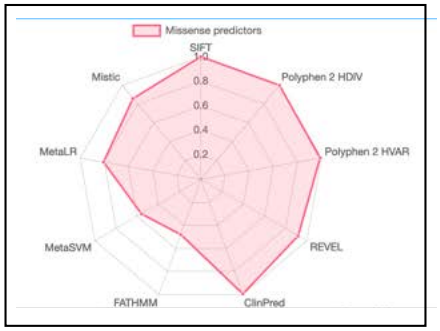
p.Gly265Val (G265V)



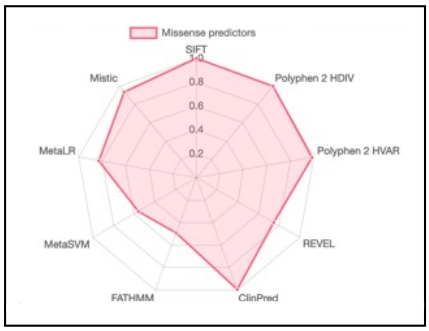
p.Thr268Ile (T268I)



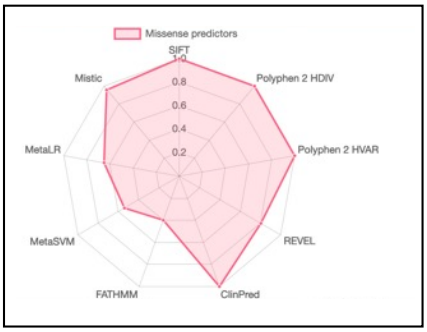
p.Ile269Thr (I269T)



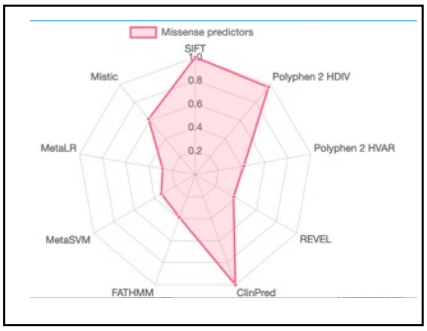
p.Thr296Met (T296M)



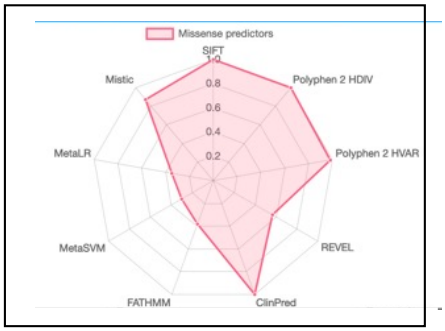
p.Tyr303Cys (Y303C)



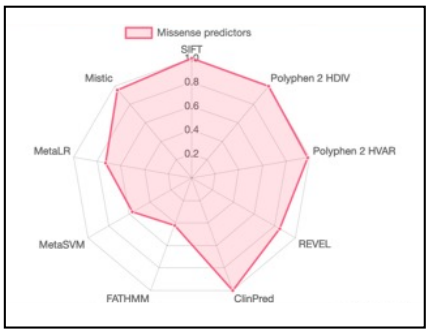
p.Arg312His (R312H)



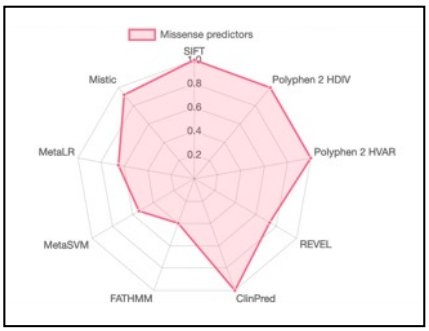
p.Pro317Arg (P317R)



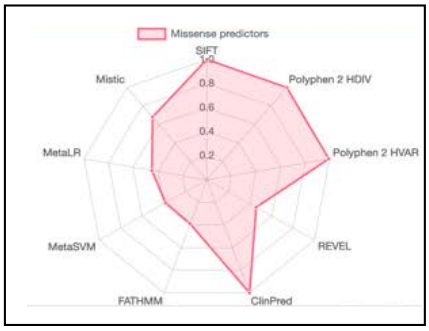
p.Tyr329His (Y329H)



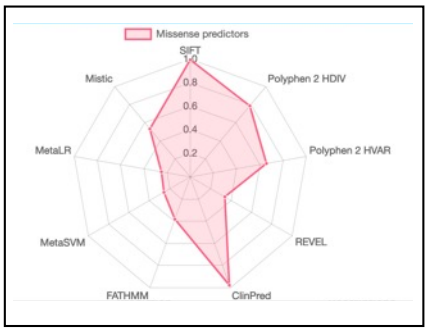
p.Trp334Arg (W334R)



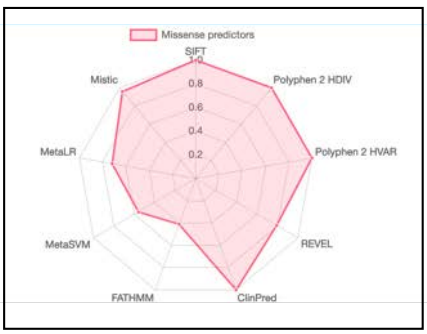
p.Gly349C (G349C)



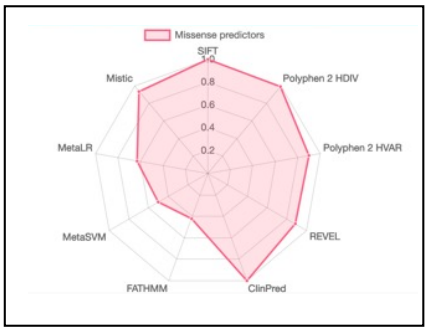
p.Gly349Ser (G349S)



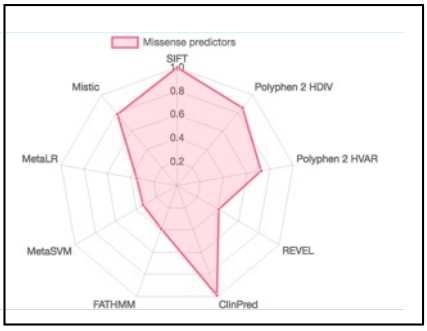
p.Pro358Thr (P358T)



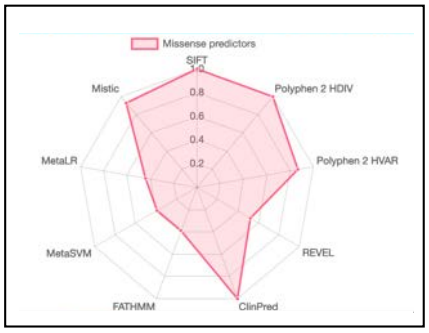
p.Phe366Leu (F366L)



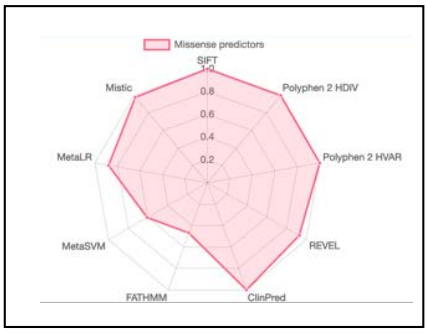
p.Arg370Gly (R370G)



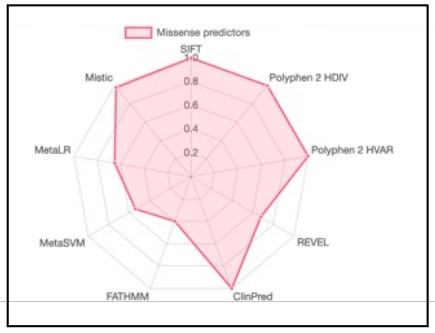
p.Arg371His (R371H)



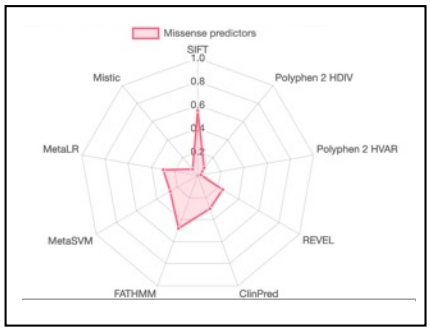
p.His374Arg (H374R)



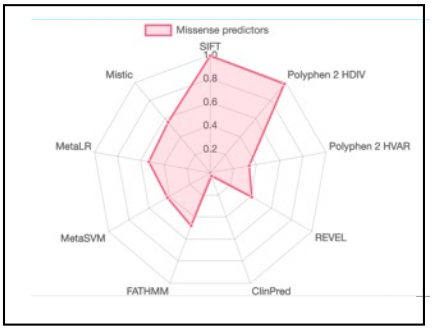
p.Trp389Arg (W389R)



p.Asp419Glu (D419E)



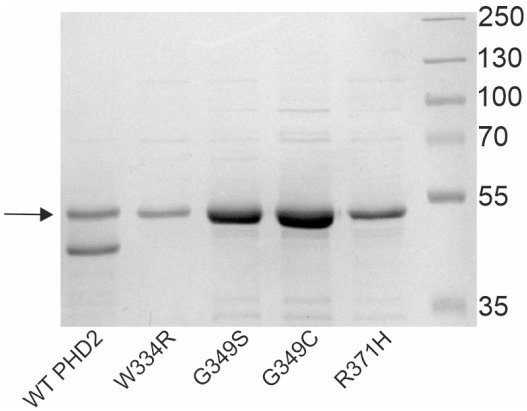
p.Ser420Leu (S420L)



Supplementary Figure 4

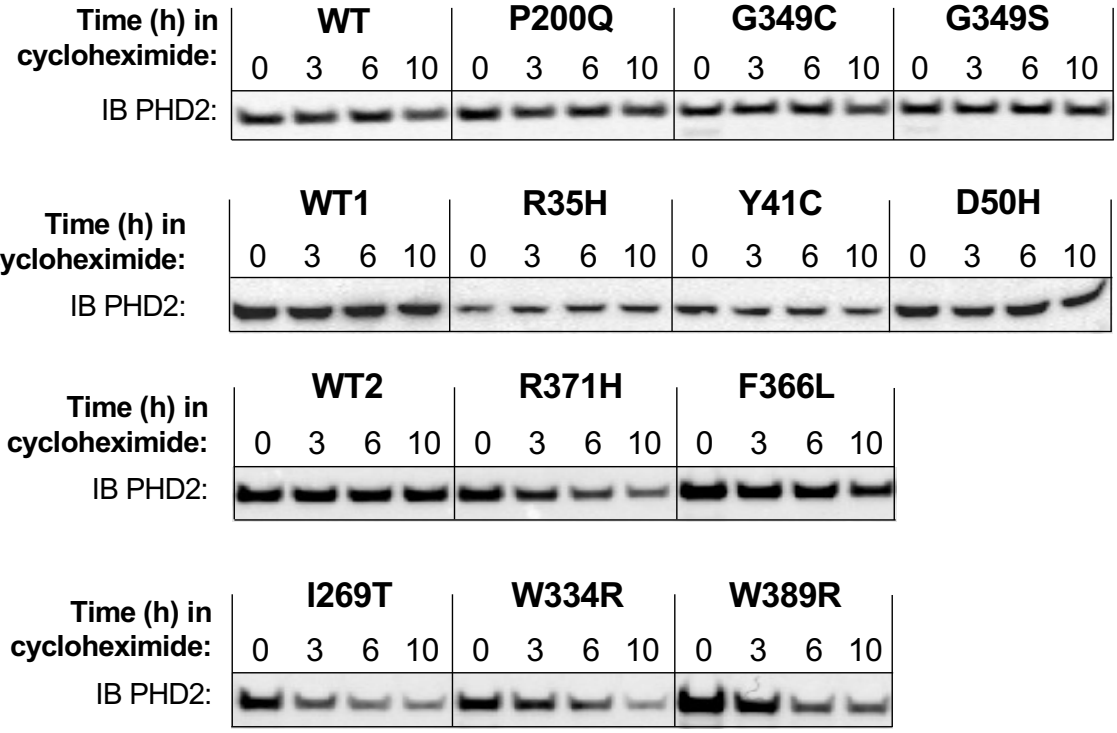
					In silico prediction			Hematocrit			Hemoglobin		
ID	Frequency gnomAD	Nb	Location	Coordinates	Mobidetails Single predictors	Mobidetails Meta predictors	SPiP score	Mean_Hct	Sd_Hct	p value_Hct	Means_Hb	Sd_Hb	p value_Hb
chr1_231421984_T_C		2	UTR5					47,56	0,099	0,009	16,075	0,007	1,43E-05
chr1_231421973_G_T		1	UTR5					47,23	NA	NA	15,66	NA	NA
chr1_231421888_TGGC(		1	UTR5					47,86	NA	NA	16,06	NA	NA
chr1_231421888_TGGC(		2	UTR5					47,16	2,319	0,261	16,275	1,167	0,366
chr1_231421888_TGGC(		1	UTR5					48,39	NA	NA	16,53	NA	NA
chr1_231421856_CGGC(	6.601e-06	1	Exon 1	c.C32T, p.P11L	02.51 %	0.299	0.239	47,6	NA	NA	16,5	NA	NA
chr1_231421836_C_A	6.592e-06	1	Exon 1	c.G53T, p.R18L	02.51 %	0.650	0.607	45,66	NA	NA	16,29	NA	NA
chr1_231421786_G_C	6.589e-06	1	Exon 1	c.C103G, p.R35G	02.51 %	0.850	0.664	47,17	NA	NA	15,66	NA	NA
chr1_231421781_G_A	2.635e-05	1	Exon 1	c.C108T, p.C36C	02.51 %	NA	NA	47,35	NA	NA	16,45	NA	NA
chr1_231421764_CAGT_	6.588e-06	1	Exon 1	c.122_124del, p.Y41del	02.51 %	NA	NA	47,54	NA	NA	16,4	NA	NA
chr1_231421757_C_T	0	1	Exon 1	c.G132A, p.K44K	02.51 %	NA	NA	45,5	NA	NA	16,4	NA	NA
chr1_231421678_GGCC(	0	1	Exon 1	c.A200G, p.H67R	02.51 %	0.221	0.152	48,18	NA	NA	16,7	NA	NA
chr1_231421633_GCGG	0	2	Exon 1	c.241_255del, p.A81_P85del	02.51 %	NA	NA	46,19	2,927	0,406	16,36	1,188	0,353
chr1_231421566_G_A	6.593e-06	2	Exon 1	c.C323T, p.A108V	02.51 %	0.311	0.226	47,89	2,143	0,207	16,425	0,064	0,019
chr1_231421552_T_C	0	1	Exon 1	c.A337G, p.K113E	02.51 %	0.329	0.214	47,58	NA	NA	16,24	NA	NA
chr1_231421543_G_T	4.615e-05	1	Exon 1	c.C346A, p.P116T	02.51 %	0.297	0.242	47,98	NA	NA	16,24	NA	NA
chr1_231421507_G_C	1.982e-05	2	Exon 1	c.C382G, p.R128G	02.51 %	0.280	0.268	47,37	1,520	0,168	16	0,283	0,126
chr1_231421394_TG_T	6.588e-06	1	Exon 1	c.494delC, p.P165Qfs*9	02.51 %	NA	NA	50,3	NA	NA	17,5	NA	NA
chr1_231421373_C_T	0	1	Exon 1	c.G516A, p.A172A	02.51 %	NA	NA	48,55	NA	NA	16,67	NA	NA
chr1_231421327_G_A	1.314e-05	1	Exon 1	c.C562T, p.L188L	02.51 %	NA	NA	48,98	NA	NA	16,99	NA	NA
chr1_231421290_G_A	0	2	Exon 1	c.C599T, p.P200L	02.51 %	0.605	0.412	45,3	2,878	0,521	16,25	0,339	0,121
chr1_231421228_G_A	6.573e-06	1	Exon 1	c.C661T, p.Q221X	02.51 %	NA	NA	50,55	NA	NA	17,51	NA	NA
chr1_231421191_C_T	0	1	Exon 1	c.G698A, p.G233E	02.51 %	0.832	0.451	53,43	NA	NA	18,46	NA	NA
chr1_231421154_A_C	7.275e-05	2	Exon 1	c.T735G, p.S245R	03.48 %	0.344	0.156	47,45	3,719	0,366	16,145	1,138	0,390
chr1_231421141_C_A	0	1	Exon 1	c.G748T, p.D250Y	03.48 %	0.667	0.518	48,83	NA	NA	17,83	NA	NA
chr1_231421105_T_C	0	1	Exon 1	c.A784G, p.K262E	03.43 %	0.369	0.502	49,45	NA	NA	16,76	NA	NA
chr1_231421083_A_G	1.316e-05	1	Exon 1	c.T806C, p.I269T	03.43 %	0.866	0.828	54,6	NA	NA	18,76	NA	NA
chr1_231421025_G_C	0	1	Exon 1	c.C864G, p.G288G	07.62 %	NA	NA	47,1	NA	NA	16,1	NA	NA
chr1_231421007_G_A	0	1	Exon 1	c.C882T, p.G294G	05.05 %	NA	NA	50,1	NA	NA	17,2	NA	NA
chr1_231421002_G_A	0	1	Exon 1	c.C887T, p.T296M	35.81 %	0.868	0.814	48,35	NA	NA	16,72	NA	NA
chr1_231420964_GCCTG		2	intronic					46,85	2,192	0,268	16,24	0,651	0,226
chr1_231420936_C_G		1	intronic					47,5	NA	NA	15,3	NA	NA
chr1_231420929_T_C		1	intronic					47,2	NA	NA	15,7	NA	NA
chr1_231420906_A_G		1	intronic					48	NA	NA	17,1	NA	NA
chr1_231374074_T_C	2.628e-05	1	Exon 2	c.A917G, p.N306S	30.67 %	0.315	0.209	46,8	NA	NA	16,9	NA	NA
chr1_231373910_T_C		1	intronic					49	NA	NA	16,88	NA	NA
chr1_231370724_C_T		1	intronic					49	NA	NA	16,8	NA	NA
chr1_231370599_G_A	0	1	Exon 3	c.C1111T, p.R371C	05.03 %	0.848	0.692	47,81	NA	NA	16,95	NA	NA
chr1_231370598_C_T	0	1	Exon 3	c.G1112A, p.R371H	05.03 %	0.814	0.656	49,09	NA	NA	16,91	NA	NA
chr1_231370593_G_A	0	1	Exon 3	c.C1117T, p.P373S	07.62 %	0.847	0.689	46,93	NA	NA	16,82	NA	NA
chr1_231370475_C_A		1	intronic					49,09	NA	NA	16,43	NA	NA
chr1_231369653_A_T		1	intronic					49,8	NA	NA	16,5	NA	NA
chr1_231369461_G_C		1	intronic					49,4	NA	NA	16,6	NA	NA
chr1_231369460_G_A		1	intronic					47,88	NA	NA	16,2	NA	NA
chr1_231367723_A_C		1	intronic					47,56	NA	NA	15,73	NA	NA
chr1_231367655_T_G		1	intronic					45,89	NA	NA	16,28	NA	NA
chr1_231367600_C_T	0	1	Exon 4	c.G1185A, p.E395E	07.62 %	NA	NA	48,8	NA	NA	17	NA	NA
chr1_231367496_G_C	3.229e-05	1	intronic					47,39	NA	NA	16,25	NA	NA
chr1_231366408_G_A		1	UTR3					48,14	NA	NA	16,66	NA	NA

Supplementary Figure 5

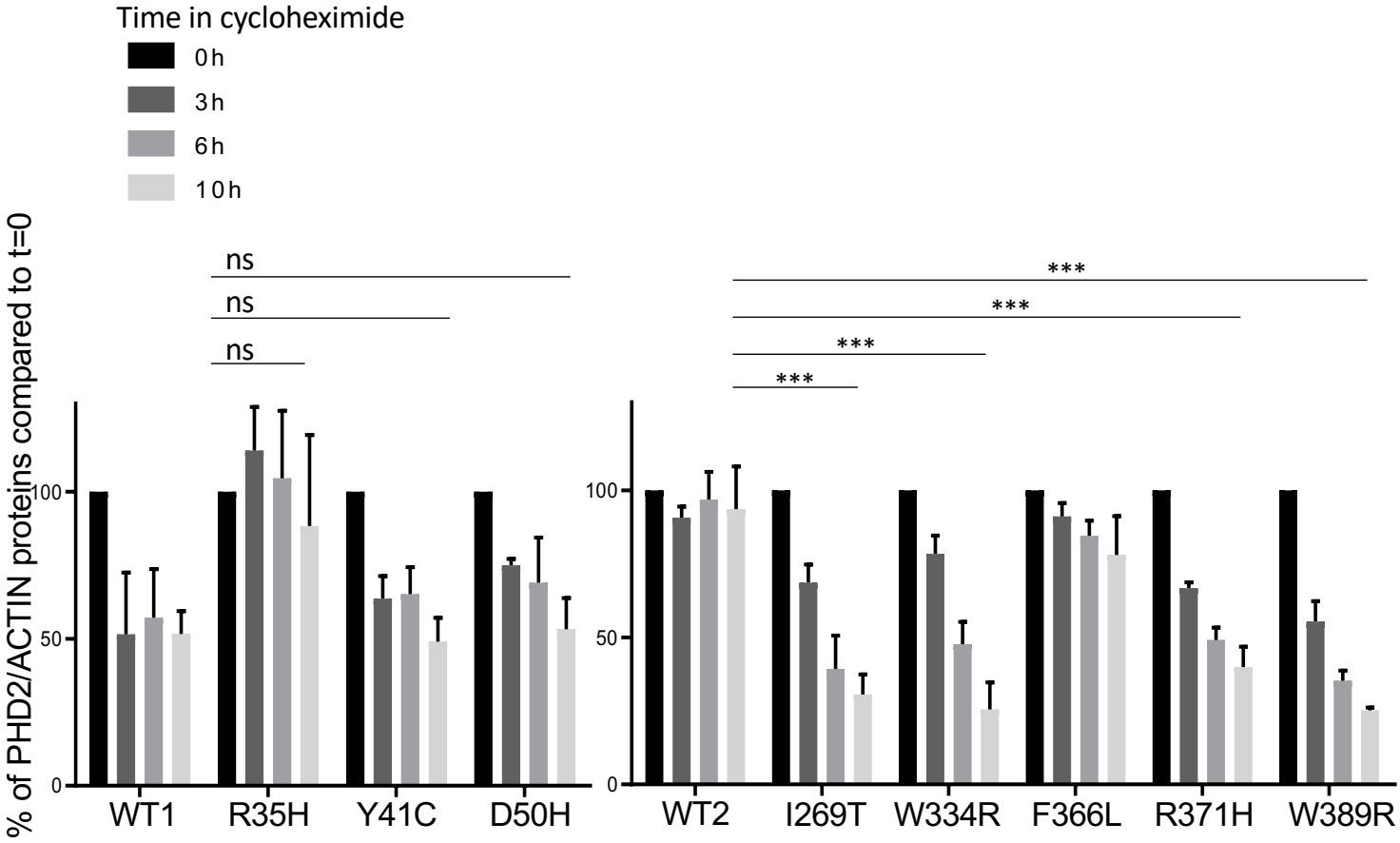


Supplementary Figure 6

A



B

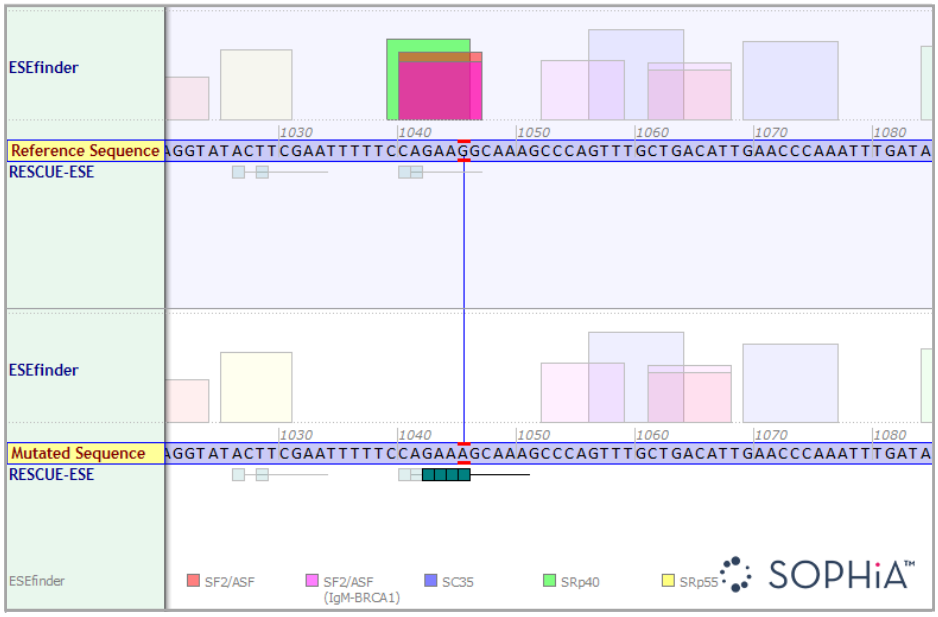


Supplementary Figure 7

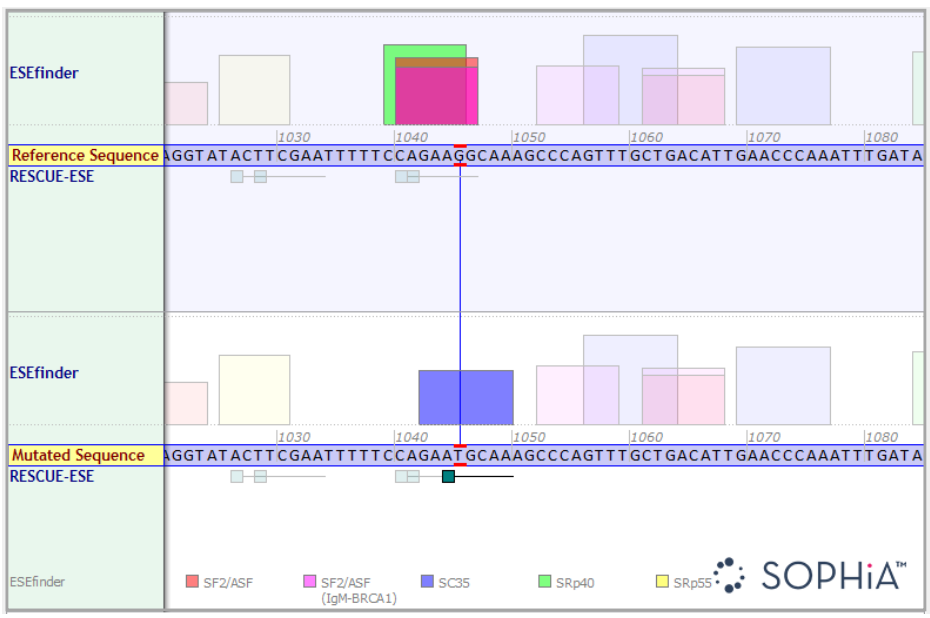
A Alamut analysis: impact of variants on splicing, NM_022051.2 (*EGLN1*)

c.891+1G>A

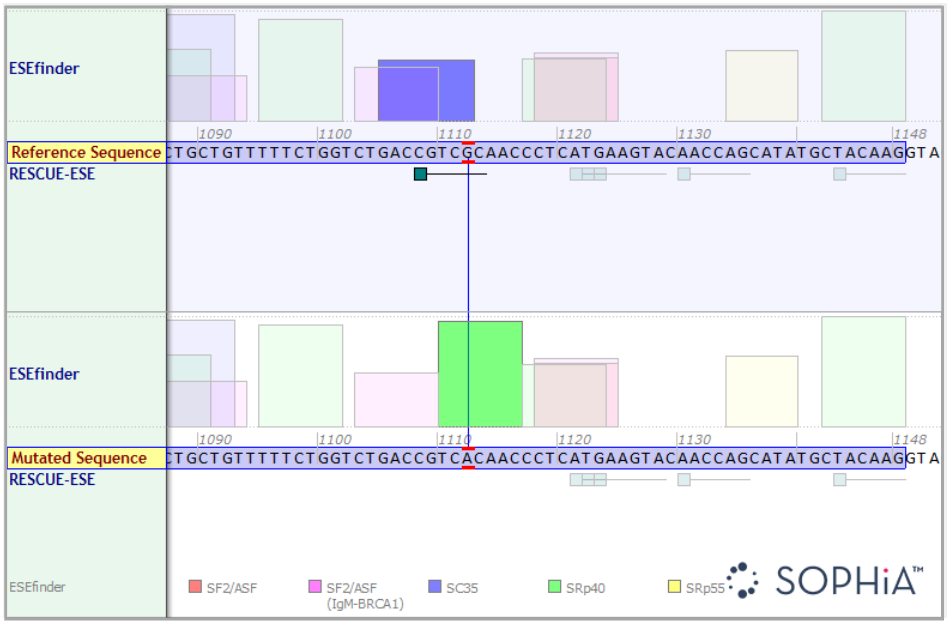
c.1045G>A, p.(Gly349Ser)



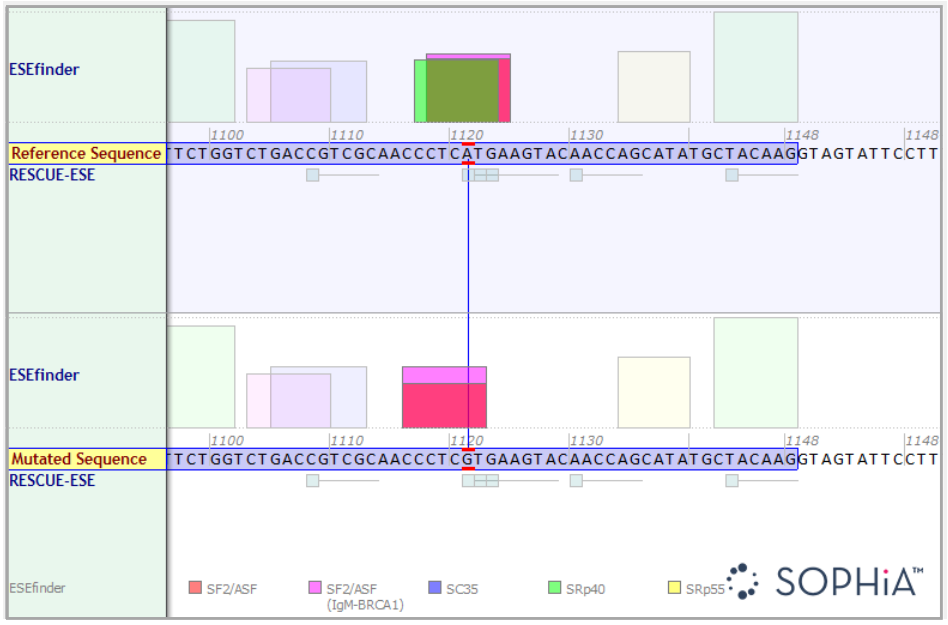
c.1045G>T, p.(Gly349Cys)



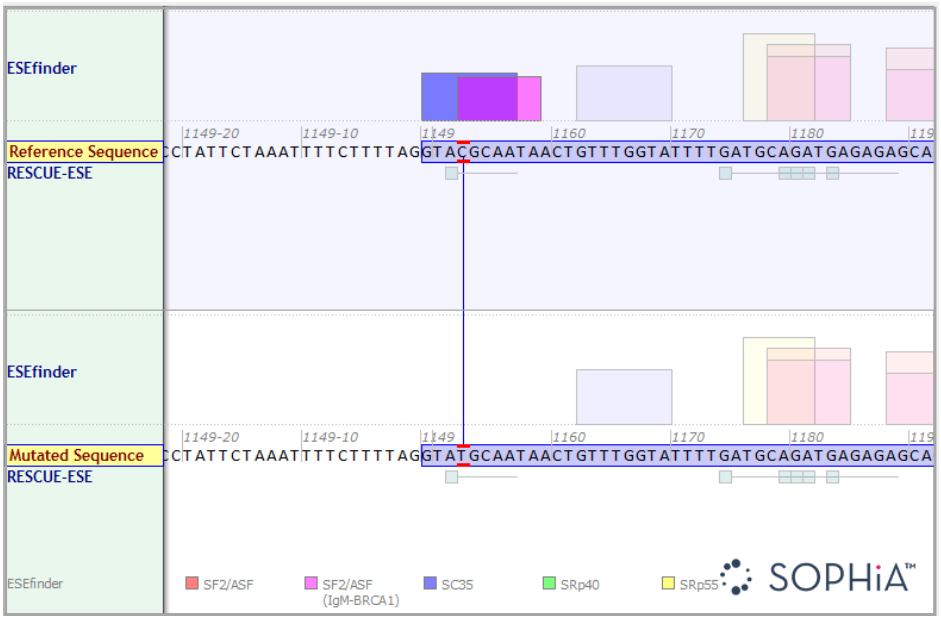
c.1112G>A, p.(Arg371His)



c.1121A>G, p.(His374Arg)

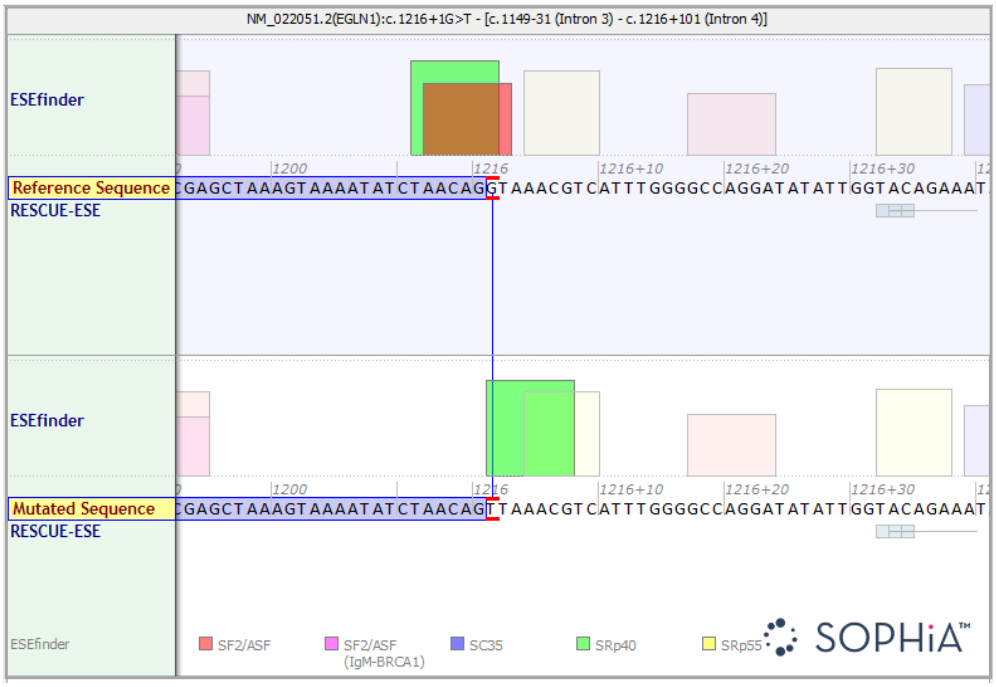


c.1152C>T, p.(Tyr384=)



c.1165T>C, p.(Trp389Arg)

c.1216+1G>T, p.?

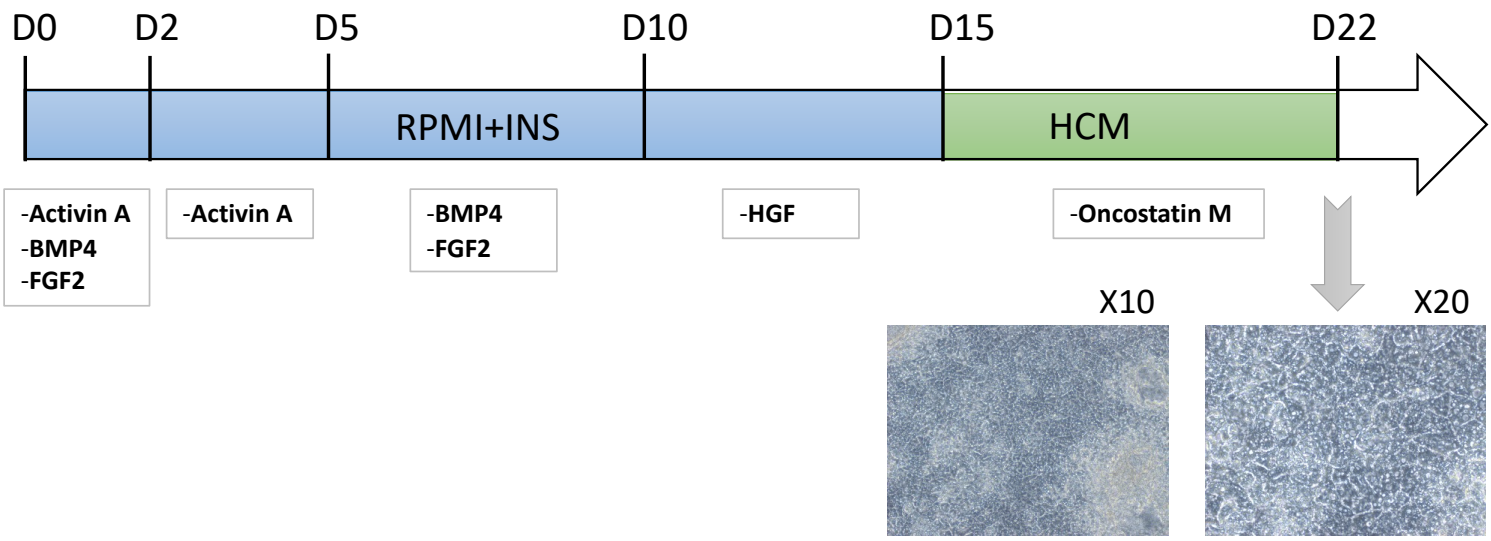


B Mobidetails analysis: impact of variants on splicing using SPiP

		SPiP results (The risk for the variant to alter splicing)	SpliceAI value (5' splice acceptor site)	SpliceAI value (3' splice donor site)	Overall interpretation of SPiP
Exon 1 WT	/	/	/	-0.99559998	/
c.891+1G>A	/	98.41 % [91.47 % - 99.96 %]	/	-0.00043839	Alteration of the consensus splice site
Exon 3 WT	/	/	0.99479997	-0.99739998	/
c.1045G>A	p.(Gly349Ser)	03.59 % [01.56 % - 06.95 %]	0.99415666	-0.99689865	No effect on splicing
c.1045G>T	p.(Gly349Cys)	03.59 % [01.56 % - 06.95 %]	0.99430639	-0.99705541	No effect on splicing
c.1112G>A	p.(Arg371His)	05.03 % [02.44 % - 09.05 %]	0.99334192	-0.99699247	No effect on splicing
c.1121A>G	p.(His374Arg)	07.62 % [04.42 % - 12.08 %]	0.99552345	-0.99782789	No effect on splicing
Exon 4 WT	/	/	0.88029998	-0.88660001	/
c.1152C>T	p.(Tyr384=)	47.89 % [39.44 % - 56.42 %]	0.79292428	-0.76368159	Alteration of an exonic splicing
c.1165T>C	p.(Trp389Arg)	09.76 % [06.06 % - 14.67 %]	0.98129308	-0.98149669	No effect on splicing
c.1216+1G>T	/	98.41 % [91.47 % - 99.96 %]	0.32907361	-8.8e-7	Alteration of the consensus splice site

Supplementary Figure 8

A



B