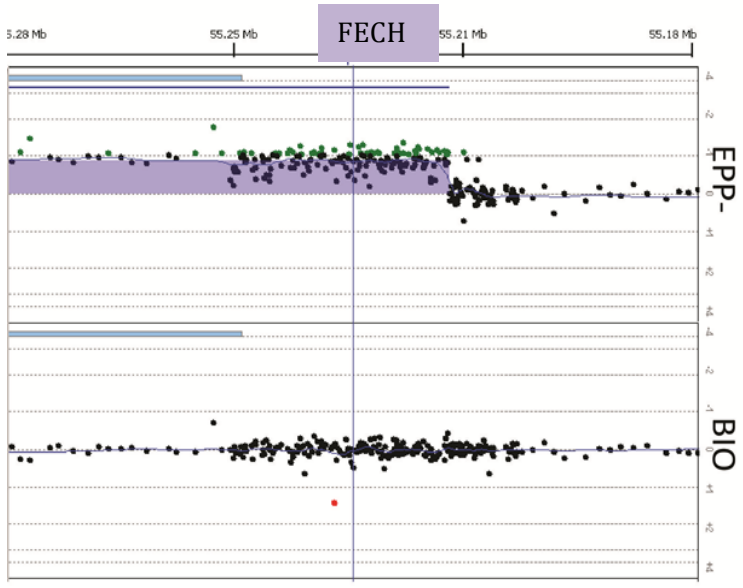
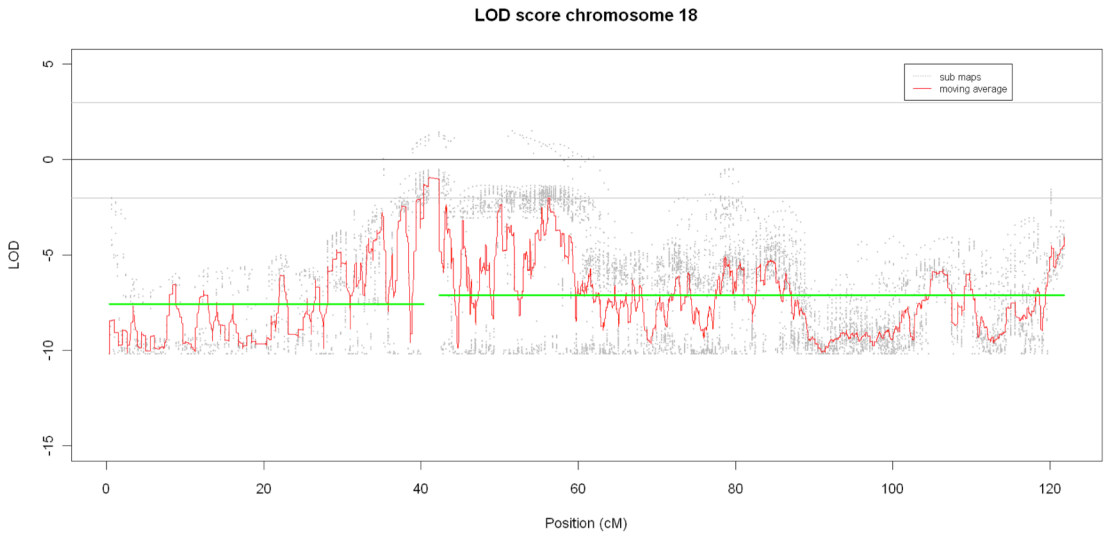
Candidate gene	Reason	Haplotypic exclusion (SNP and/or microsatellite)	Sequencing Exclusion (Exon and flanking sequences)
FECH	EPP	Yes	Yes
ALAS2	XLPP	Yes	Yes[1]
SLC25A37/MFRN1	EPP modifier gene	-	Yes
GATA1	Master transcriptional regulator of erythropoiesis	Yes	Yes
ABCB7	Involved in iron-sulfur cluster transport	Yes	Yes
IRP2	IRE regulation	Yes	-
ABCB10	Erythroid-specific regulator of iron transport; mitochondrial transporter	-	Yes
SLC25A38	Glycine transporter	-	Yes

Supplemental Table 1. Sequencing of candidate porphyria modifier genes. No mutations were identified in the candidate genes in the index case. [1]For *ALAS2* gene, the coding region, including exon 11 and its boundaries responsible for XLPP, were sequenced, and no mutations were identified.

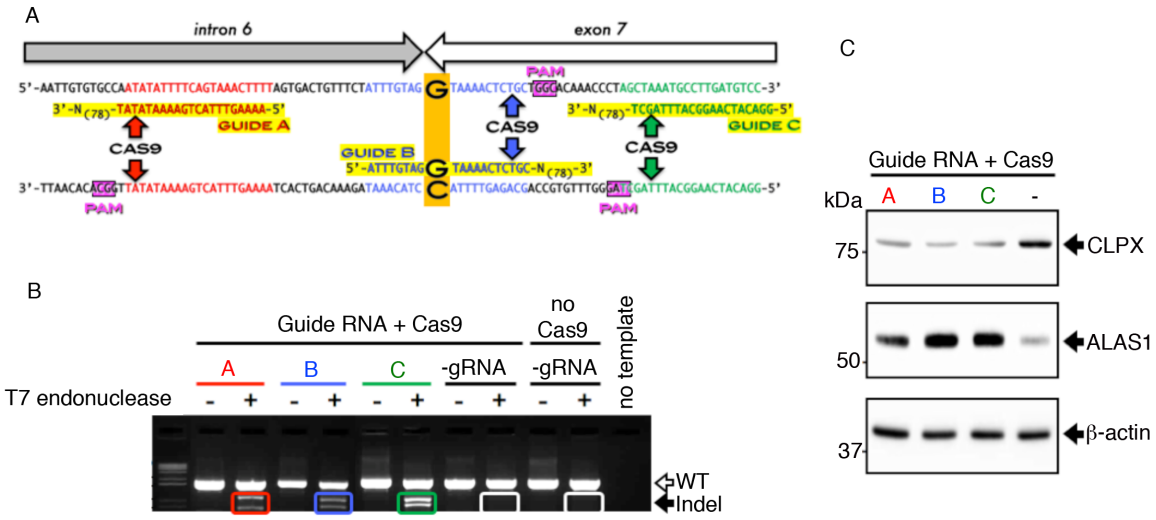
A



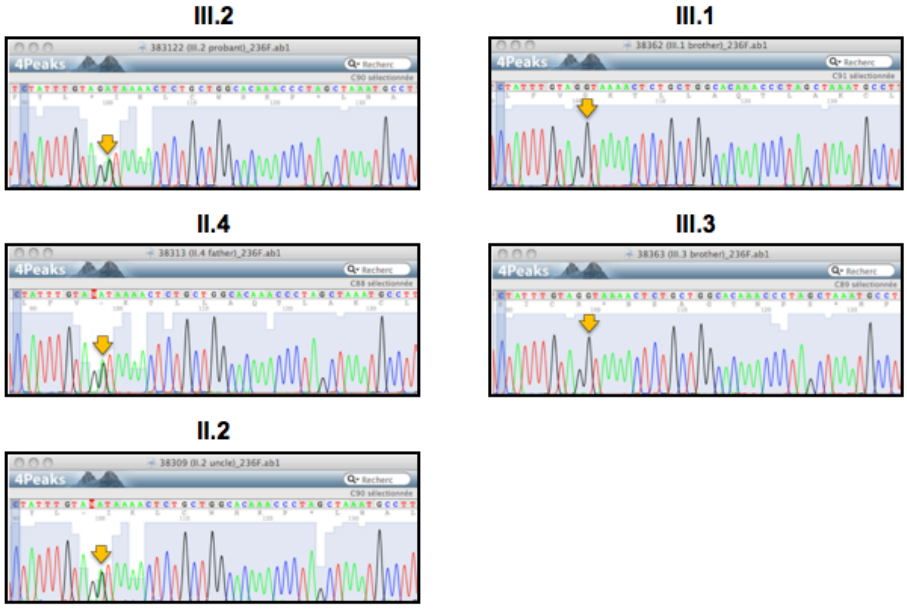
B



Supplemental Figure 1. Comparative genomic hybridization (CGH) and linkage analysis of chromosome 18 in the variant EPP family. (A) Genomic hybridization of a large region on chromosome 18 by CGH array on the Agilent8x15k excludes *FECH* as the genetic cause for EPP in this pedigree. **EPP** sample **(top)** shows analysis of genomic DNA isolated from a classical EPP patient with a large *FECH* deletion as a positive control. **BIO** sample **(bottom)** denotes the proband under study (patient III.2).
(B) Exclusion of linkage between the disease and chromosome 18 mutations (LOD score <-2) (see supplemental methods). The green line represents the mean of the LOD scores. All other chromosomes were subject to similar analyses.



Supplemental Figure 2. ALAS1 protein is upregulated in *CLPX*-deficient K562 cells. **(A)** Guide RNAs allow Cas9 nuclease to cleave either CLPX intron 6 (guide A) or CLPX exon 7 (guides B or C). Cleavage sites are indicated by colored arrows. The site of the CLPX G298D mutation is highlighted in the orange box. **(B)** K562 cells were transfected with plasmid encoding Cas9 in combination with the indicated guide RNAs, or Cas9 alone (-gRNA). 48 hours after transfection, genomic DNA was purified and amplified using primers 240 and 241. Mutagenic (INDEL) events were visualized using a T7 endonuclease I assay (T7E1). The white arrow indicates WT allele (1901 bp) not digested by T7E1. Black arrow (labeled Indel) indicates T7E1-digested DNA fragments. **(C)** Soluble proteins were extracted from frozen pellets of K562 cells and analyzed by western blotting. The CLPX+gRNA transfected cells expressed decreased amounts of CLPX protein and increased amounts of ALAS1 protein compared to the -gRNA control. β-actin was used as a loading control.



Supplemental Figure 3. Sequencing of the *CLPX* mutation in lymphoblastoid cell lines from patients. Genomic *CLPX* from LBCL lines obtained from the proband and her relatives was sequenced. A 901 bp fragment surrounding the mutation was amplified using LA Taq polymerase (TaKaRa) with a forward primer annealing at the beginning of *CLPX* exon 6 and a reverse primer (see supplemental methods) annealing at the end of *CLPX* exon 8. The amplicon was purified and Sanger sequencing was done using internal primer 236.

Supplemental Methods

Linkage analysis

Global linkage analysis was performed with a Human Mapping Nsp 250k array following manufacturer's instructions (Affymetrix). We extracted DNA from all available members of the family (II-1, II-2, II-3, II-4, II-5, III-1, III-2, III-3). Software used to analyze data were as follows: Extraction of .cel files was carried out with Genotyping Console from Affymetrix; identification of genotyping errors, familial relationships, and exclusion of non-informative markers was carried out using Alohomora. The linkage analysis was performed by Merlin. When using PPIX accumulation as the biological parameter, the disease was determined to be dominant (autosomal or X-linked) with full penetrance. The maximal LOD Score was 1.420 and was insufficient to determine linkage regions. However, regions with LOD scores below -2 were identified, allowing their exclusion as candidate linkage regions. Notably, most of the markers of the chromosome 18 (containing the *FECH* gene) and X (containing the *ALAS2* gene) scored below -2.

CGH array procedure and data analysis

CGH labeling and hybridization were performed as recommended in the Agilent 8x15k manual. Briefly, 300 ng of genomic DNA was digested with AluI and RsaI (Promega, Madison, WI) for 2 hours at 37°C. The digested DNAs were labeled by random priming with the Agilent Genomic DNA Labeling Kit Plus (Agilent Technologies) for 2 hours at 37°C, according to the manufacturer's instructions. Patient and pooled normal control DNAs (reference) were labeled with Cy5-dUTP and Cy3-dUTP, respectively. Labeled products were purified with Microcon YM-30 filters (Millipore, Billerica, MA). Patient

and normal control gDNAs were mixed and hybridized with Human Cot I DNA (Invitrogen) at 65°C for 24 hours. DNA hybridization was performed at the University Institute of Hematology (Saint Louis Hospital, Paris) according to the Agilent protocol. A reference versus reference control hybridization step was used to assess array hybridization quality. Arrays were scanned with an Agilent DNA Microarray Scanner (G2565BA). Log2 ratios were determined with Agilent Feature Extraction software (v9.1.3.1) and the global quality of the individual microarrays used in the experiment was validated against the quality metrics of this software. Results were visualized and analyzed with Agilent's CGH Analytics software, and copy number aberrations were detected with the Aberration Detection Method 2 (ADM-2) algorithm using a threshold value of 6.0. This threshold value of the altered copy number was determined from the normal variation in the control hybridization.

Whole exome sequencing (WES)

To identify the underlying genetic cause of the disease in this variant EPP family, WES was performed using the proband's genomic DNA. WES coverage of the captured regions was 96% and 90.67% for 10x and 25x depth of coverage respectively resulting in a mean sequencing depth of 82x per base. Exons of DNA samples were captured with in-solution enrichment methodology (SureSelectXT Human All Exon V.2 Kit; Clinical Research Exome, Agilent, Massy, France). Image analysis and base calling were performed with Real Time Analysis software (Illumina). Genetic variants were identified by an in-house pipeline (IntegraGen). All variants present within intergenic, intronic, and untranslated regions and synonymous substitutions were excluded. Stringent filtering criteria were used to select most likely pathogenic variant(s): only nonsense, missense,

splice site variants or small deletions or insertions (InDels) with a minor allelic frequency ≤0.005 in Agregation exome database were considered to be potentially disease-causing. Erythroid expression ($CD71^+$) of candidate genes was required (BioGPS: http://biogps.org). Variant pathogenicity was predicted with bioinformatic tools: Polymorphism Phenotyping v2 (PolyPhen2, http://genetics.bwh.harvard.edu/pph2/), Sorting Intolerant From Tolerant (SIFT, http://sift.jcvi.org/). Amino acid conservation across species was assessed with UCSC Genome Browser (http://genome.ucsc.edu/index.html; Human GRCh37/hg19 Assembly). The *CLPX* variant selected after WES was validated in the index case and a cosegregation study in the family was performed by bidirectional direct sequencing. PCR products were purified using ExoSAP-IT (USB Products Affymetrics, Cleveland, USA). After purification, both strands were sequenced using a Big Dye Terminator Cycle Sequencing kit (Applied Biosystems, Life Technologies, Carlsbad, CA). Sequencing products were purified (Sephadex G50, GE Healthcare, Piscataway, NJ) and analyzed on a 3130XL Genetic Analyzer using the Seqscape analysis software (v2.6.0) (Applied Biosystems, Life Technologies, Carlsbad, CA) (primers and conditions are available from the authors on request).

cDNA preparation for *CLPX* allele sequencing

Total RNA was extracted from frozen CD34+ cells with RNA PLUS™ (MP-Bio). 2 µg of RNA was diluted in a final volume of 10 µL with water, was denatured for 30 sec at 85°C and for 4 min 30 sec at 70°C. The cDNA synthesis reaction was performed in the presence of 0.5 mM of each dNTP, 10 ng/µL of random hexanucleotide primers, 0.1

mg/mL bovine serum albumin, 1 unit/µL RNase OUT (Invitrogen), 10 mM dithiothreitol, and 10 units/µL SuperScript II™ reverse transcriptase (Invitrogen). The reaction was conducted for 60 min at 42 °C before reverse transcriptase was inactivated for 6 min at 96 °C. After this reaction, the cDNA was diluted with 80 µL of 10 mM Tris-HCl, 0.1 mM EDTA (pH 8.0). *CLPX* PCR amplification was performed with primers (see table below) with FastStart *Taq* polymerase (Roche). The reaction was analyzed on 1.5% agarose gel containing SYBR Safe diluted 1:5 000 (Invitrogen).

NAME	GENE	SEQUENCE (5' > 3')	APPLICATION
240 (Forward)	***CLPX* (exon 6)**	**CTTCAGATTGCTGGAATTAGC**	**- Amplification of gDNA - Amplification of cDNA - Sequencing**
241 (Reverse)	***CLPX* (exon 8)**	**GCTGAACGCCTTCTCCAC**	**- Amplification of gDNA - Amplification of cDNA - Sequencing**
236 (Forward)	***CLPX* (intron 6)**	**GCTGAACGCCTTCTCCAC**	**- Sequencing**

Disruption of *CLPX* in K562 by CRISPR/Cas9

The pSpCas9(BB)-2A-Puro (PX459) expression vector was purchased from Addgene (plasmid #48139) (1). Oligonucleotide sequences of for gRNAs are as follows (F denotes Forward and R denotes Reverse directions.)

NAME	POSITION IN CLPX GENE	SEQUENCE (5' > 3')
guide A-F	intron 6	CACCGAAAGTTTACTGAAAATATAT
guide A-R	intron 6	AAACATATATTTTCAGTAAACTTTC
guide B-F	intron 6 – exon 7	CACCGATTTGTAGGTAAAACTCTGC
guide B-R	intron 6 – exon 7	AAACGCAGAGTTTTACCTACAAATC
guide C-F	exon 7	CACCGGACATCAAGGCATTTAGCT
guide C-R	exon 7	AAACAGCTAAATGCCTTGATGTCC
PX459 plasmid sequencing	not relevant	GAGGGCCTATTTCCCATGATTCC

Guides B and C targeted *CLPX* gene cleavage at exon 7 (between +9 and +10 for B; between +28 and +29 for C); guide A targeted *CLPX* gene cleavage in intron 6 (between -39 and -40). Oligonucleotides were ligated into the *Bbs*I linearized pSpCas9(BB)-2A-Puro (PX459). The integrity of each clone guide sequence was checked by Sanger sequencing with PX459 plasmid sequencing primer (see table above). The resultant plasmids, PX459-CLPX A, PX459-CLPX B and PX459-CLPX C, were purified with EndoFree Plasmid Maxi Kit (Qiagen). 6 µg of plasmid were electroporated with 2 x 10^6 K562 cells the using Nucleofector II (kit V - program T-016 - Lonza). Cells were seeded at a density of 150,000 cells per well in a 6-wells plate 48h before electroporation. Viability was >95% the day of transfection. FACS analysis of cells transfected with the pmaxGFP plasmid revealed that the transfection efficiency was 60-70 % 24h after the nucleofection. Transfected K562 cells were grown with puromycin in the culture medium 24 hours post-transfection (2 µg/µL during 3 days and then 1 µg/µL). Immunoblotting experiments were done on proteins extracted from puromycin selected pooled cells. Generation of INDEL events in the *CLPX* gene was assessed using a T7 endonuclease I assay (New England Biolabs).

Culture of K562 cells

K562 cells were obtained from Lionel Arnaud (2) and were cultured in IMDM, 25 mM HEPES and GlutaMAX (Gibco) supplemented with 10% fetal bovine serum and antibiotics (100 units/mL penicillin and 100 µg/mL streptomycin).

CLPX cRNA preparation for microinjections

Human CLPX-FLAG tagged cDNA (GenBank accession #NM_006660) was used as template for site-directed mutagenesis using the Quikchange kit (Stratagene) as directed. Both the wildtype and p.Gly298Asp CLPX alleles were sub-cloned into the pXT7 expression vector using the *Stbl3 E. coli* strain (Invitrogen) and grown at 30^{o}C. 5'-capped cRNA was generated *in vitro* using the T7 mMessage mMachine kit (Ambion). A small aliquot of the synthesized cRNA was translated *in vitro* with [^{3}H] L-leucine (Specific Activity ~100 Ci/mmol, Amersham) with wheat germ extract (Promega). The radiolabeled translated protein was resolved by SDS-PAGE, fixed, visualized by fluorography (Amplify) at -80^{o}C and revealed efficient translation of the expected sized proteins.

Microinjections and ALAS activity assay

Capped CLPX cRNA mixed with eGFP tracer cRNA was microinjected into zebrafish zygotes at the 1-2 cell stage as previously described (3). At 24 hours post fertilization, zebrafish embryos showing eGFP expression were analyzed for protein expression by western analysis by immunoprecipitaton (below) and measurement of ALAS activity by HPLC as previously described (4). Uninjected zebrafish zygotes from the same clutch were used as control for all comparisons.

Immunoprecipitation and western analysis

Zebrafish embryos at 24 hpf with eGFP expression were lysed in groups of ~50 embryos in RIPA buffer/protease inhibitor cocktail (Roche) with mechanical douncing. The

clarified soluble lysate after high-speed centrifugation was immunoprecipitated with anti-FLAG agarose beads (Sigma) for 12 h at 4^{o}C, then immune complex beads were serially washed in PBST/protease inhibitor cocktail (Roche) and eluted in Laemmli buffer by boiling. The eluted immune complex was resolved on SDS-PAGE and visualized by western analysis.

Supplemental References

1. Ran FA, *et al.* (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8(11):2281–2308.
2. Helias V, *et al.* (2012) ABCB6 is dispensable for erythropoiesis and specifies the new blood group system Langereis. *Nat Genet* 44(2):170–173.
3. Hildick-Smith GJ, *et al*. (2013) Macrocytic anemia and mitochondriopathy resulting from a defect in sideroflexin 4. *Am J Hum Genet* 93(5):906–914.
4. Bergonia HA, Franklin MR, Kushner JP, Phillips JD (2015) A method for determining δ-aminolevulinic acid synthase activity in homogenized cells and tissues. *Clin Biochem* 48(12):788–795.