Supplemental Methods

Whole-exome sequencing and data analysis

Genomic DNA was extracted from leukocytes using the MagnaPure system (Roche). Exome capture was performed with the Sure Select Human All Exon Kit (Agilent Technologies). Agilent Sure Select Human All Exon (58 Mb, V6) libraries were prepared from 3 µg of genomic DNA sheared with an ultrasonicator (Covaris) as recommended by the manufacturer. Barcoded exome libraries were pooled and sequenced with a HiSeq2500 system (Illumina), generating paired-end reads. After demultiplexing, sequences were mapped on the human genome reference (NCBI build 37, hg19 version) with Burrows-Wheeler Aligner (BWA). The mean depth of coverage obtained for the four probands' exome libraries was >120X with > =96% and >=94% of the targeted exonic bases covered at least 15 and 30 independent sequencing reads (>=96% at 15X >=94% at 30X). Variant calling was carried out with the Genome Analysis Toolkit (GATK), SAMtools, and Picard tools. Single-nucleotide variants were called with GATK Unified Genotyper, whereas indel calls were made with the GATK IndelGenotyper_v2. All variants with a read coverage of 23% and a Phred-scaled quality of 20% were filtered out. All the variants were annotated and filtered with PolyWeb, an in-house annotation software program.

Label-free quantification mass spectrometry analysis

Ghosts from 4 healthy donor control samples and three PEL individuals were obtained as described above and then lysed by heating for 5 min at 95°C in the same volume of 2X lysis buffer (100 mM Tris/HCl pH 8.5, 2% SDS). Protein concentration was determined using SDS-PAGE of an aliquot and Imagelab software (BioRad). Peptides were prepared by the Filter-Aided Sample Preparation method (FASP) as described[1,2]. Briefly, 35 µg of protein from each lysate was reduced and alkylated with 10 mM TCEP and 50 mM chloroacetamide for 5 min at

95°C. After cooling to room temperature, extracts were diluted with 300 µL Tris urea buffer (8 M Urea, 50 mM Tris/HCl pH 8.5), transferred onto 30 kDa centrifugal filters and prepared for digestion as described[1]. Proteins were digested overnight at 37°C with 1 µg trypsin (Promega). Peptides were desalted on C18 StageTips and fractionated by strong cationic exchange (SCX) StageTips in 5 fractions as described[3]. After drying, peptides were solubilized in 7 µL of 0.1% TFA containing 10% acetonitrile (ACN), and 2 µL of each sample fraction was analyzed in LC-MS-MS using an Ultimate 3000 Rapid Separation liquid chromatographic system coupled to an Orbitrap Fusion mass spectrometer (both from Thermo Fisher Scientific). Peptides were directly loaded onto a C18 reversed-phase analytical column (2 µm particle size, 100 Å pore size, 75 µm internal diameter, 25 cm length) with a 3 h gradient from 99% solution A (0.1% formic acid in H_2O) to 55% solution B (80% ACN, 0.085% formic acid and 20% H_2O). The Orbitrap Fusion mass spectrometer acquired data throughout the elution process and operated in a data-dependent scheme with full MS scans acquired with the Orbitrap, followed by stepped HCD MS/MS fragmentations acquired with the ion trap on the most abundant ions detected in top speed mode for 3 seconds. Resolution was set to 60,000 for full scans at AGC target $1.0e^6$ within 60 ms MIT. The MS scans spanned from 350 to 1500 m/z. The precursor selection window was set at 1.6 m/z, with AGC target $1.0e^5$ within 60 ms MIT. HCD Collision Energy was set at 30%. Dynamic exclusion was set to 30 s duration.

Mass spectrometry data processing

The mass spectrometry data were analyzed using Maxquant version 1.5.3.30[4]. The database used was a concatenation of human sequences from the UniProt-Swissprot database (UniProt, release 2017-05) and an incremented list of contaminants. The enzyme specificity was trypsin. The precursor mass tolerance was set to 4.5 ppm, and the fragment mass tolerance was set to 20 ppm. Carbamidomethylation of cysteines was set as constant modification, and acetylation of the protein N-terminus and oxidation of methionines were set as variable modifications. A

second peptide search was allowed, and the minimal length of peptides was set at 7 amino acids. The false discovery rate (FDR) was kept below 1% on both peptides and proteins. Label-free protein quantification (LFQ) was performed using both unique and razor peptides. At least 2 ratio counts were required for LFQ[4]. All experiments were analyzed simultaneously with the "match between runs" (MBR) option with a match time window of 0.7 min and an alignment time window of 20 min.

References

1. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods*. 2009;6(5):359-362.

2. Gautier EF, Leduc M, Cochet S, et al. Absolute proteome quantification of highly purified populations of circulating reticulocytes and mature erythrocytes. *Blood Adv*. 2018;2(20):2646-2657.

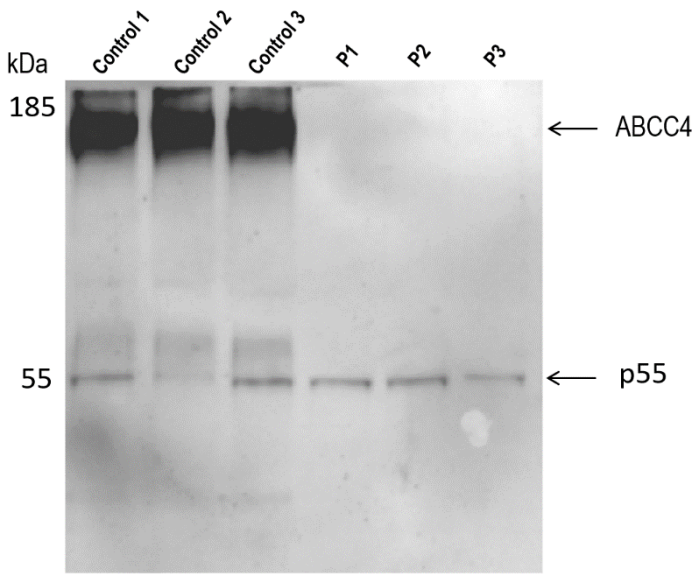
3. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat Methods*. 2014;11(3):319-324.

4. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*. 2014;13(9):2513-2526.

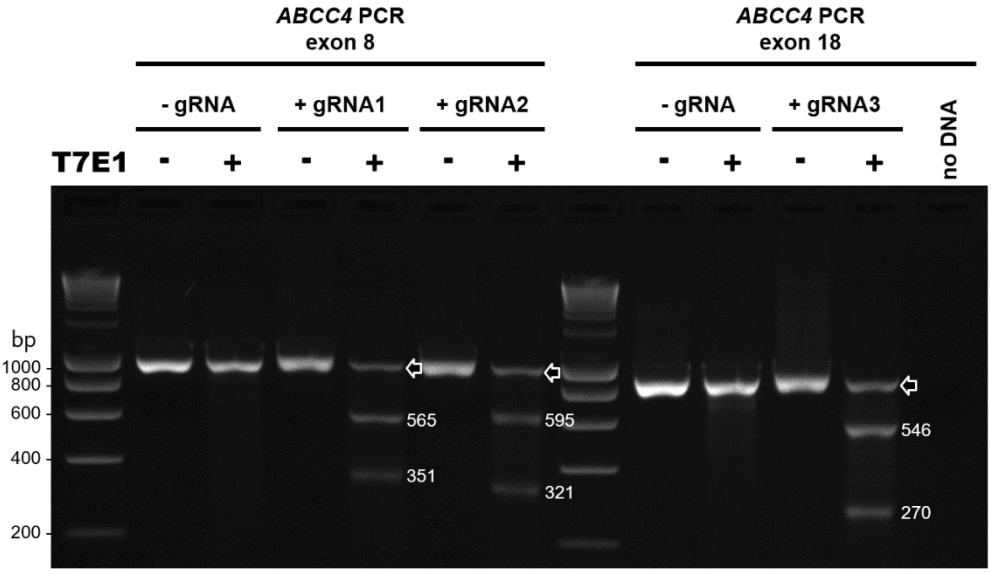
Supplemental Figures

MLPVYQEVKPNPLQDANLCSRVFFWWLNPLFKIGHKRRLEEDDMYSVLPEDRSQHLGEEL
QGFWDKEVLRAENDAQKPSLTRAIIKCYWKSYLVLGIFTLIEESAKVIQPIFLGKIINYF
ENYDPMDSVALNTAYAYATVLTFCTLILAILHHLYFYHVQCAGMRLRVAMCHMIYRKALR
LSNMAMGKTTTGQIVNLLSNDVNKFDQVTVFLHFLWAGPLQAIAVTALLWMEIGISCLAG
MAVLIILLPLQSCFGKLFSSLRSKTATFTDARIRTMNEVITGIRIIKMYAWEKSFSNLIT
NLRKKEISKILRSSCLRGMNLASFFSASKIIVFVTFTTYVLLGSVITASRVFVAVTLYGA
VRLTVTLFFPSAIERVSEAIVSIRRIQTFLLLDEISQRNRQLPSDGKKMVHVQDFTAFWD
KASETPTLQGLSFTVRPGELLAVVGPVGAGKSSLLSAVLGELAPSHGLVSVHGRIAYVSQ
QPWVFSGTLRSNILFGKKYEKERYEKVIKACALKKDLQLLEDGDLTVIGDRGTTLSGGQK
ARVNLARAVYQDADIYLLDDPLSAVDAEVSRHLFELCICQILHEKITILVTHQLQYLKAA
SQILILKDGKMVQKGTYTEFLKSGIDFGSLLKKDNEESEQPPVPGTPTLRNRTFSESSVW
SQQSSRPSLKDGALESQDTENVPVTLSEENRSEGKVGFQAYKNYFRAGAHWIVFIFLILL
NTAAQVAYVLQDWWLSYWANKQSMLNVTVNGGGNVTEKLDLNWYLGIYSGLTVATVLFGI
ARSLLVFYVLVNSSQTLHNKMFESILKAPVLFFDRNPIGRILNRFSKDIGHLDDLLPLTF
LDFIQ**TLLQVVGVVSVAVAVIPWIAIPLVPLGIIFIFLRRYFLETSRDVKRLESTTRSPV**
FSHLSSSLQGLWTIRAYKAEERCQELFDAHQDLHSEAWFLFLTTSRWFAVRLDAICAMFV
IIVAFGSLILAKTLDAGQVGLALSYALTLMGMFQWCVRQSAEVENMMISVERVIEYTDLE
KEAPWEYQKRPPPAWPHEGVIIFDNVNFMYSPGGPLVLKHLTALIKSQEKVGIVGRTGAG
KSSLISALFRLSEPEGKIWIDKILTTEIGLHDLRKKMSIIPQEPVLFTGTMRKNLDPFNE
HTDEELWNALQEVQLKETIEDLPGKMDTELAESGSNFSVGQRQLVCLARAILRKNQILII
DEATANVDPRTDELIQKKIREKFAHCTVLTIAHRLNTIIDSDKIMVLDSGRLKEYDEPYV
LLQNKESLFYKMVQQLGKAEAAALTETAKQVYFKRNYPHIGHTDHMVTNTSNGQPSTLTI
FETAL

Supplemental Figure 1: Primary sequence of ABCC4 protein. Peptides detected by mass spectrometry in PEL+ RBCs were shown in red. Amino acids encoded by the missing exons (21-31) of *ABCC4* gene in PEL-negative RBCs were shown in bold.

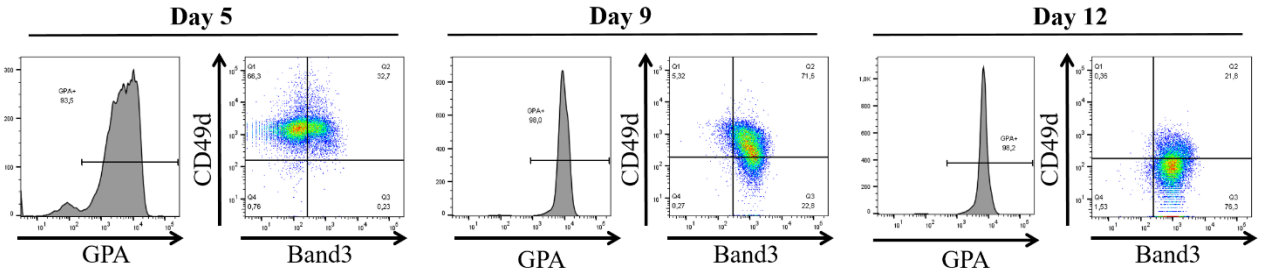


Supplemental Figure 2. Western blot analyses of ABCC4 in the RBC membranes from PEL-positive (controls 1, 2, and 3) and PEL-negative (P1, 2, and 3) individuals. Immunoblotting was performed with a mouse monoclonal antibody directed against ABCC4. No truncated ABCC4 protein could be detected in PEL-negative samples. As a control, the western blot membrane was reprobed with an antibody to p55.

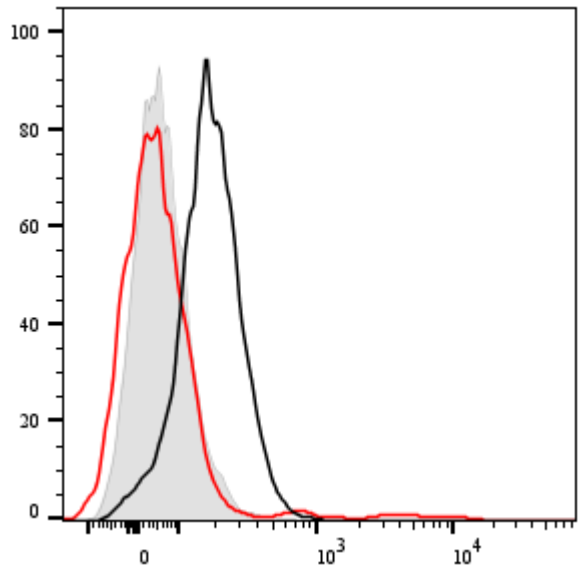


Supplemental Figure 3. K-562 cells were transfected with plasmid encoding Cas9 in combination with the indicated guide RNA (+gRNA), or plasmid encoding Cas9 alone (-gRNA). Fifteen days after transfection, genomic DNA was purified and amplified using primers

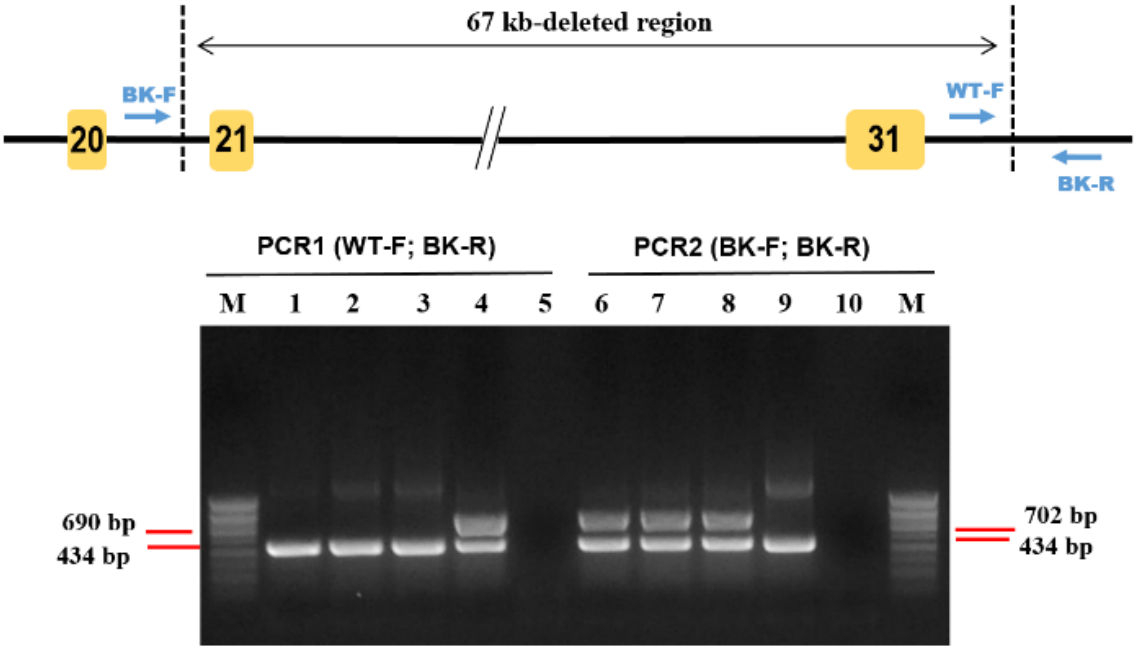
indicated in Supplemental Table 3. Mutagenic (INDEL) events were visualized using a T7 endonuclease I assay (T7E1). The white arrow indicate alleles (916 or 816 bp for exon 8 amplicon or exon 18 amplicon, respectively) not digested by T7E1. Size of T7E1-digested DNA fragments are indicated.



Supplemental Figure 4. S. *In vitro* erythropoiesis. Terminal erythroid differentiation was monitored on days 5, 7, 9, and 12 of culture after EPO addition. Cells were stained with a cocktail of antibodies consisting of an anti-band 3 PE-conjugated (Bristol), anti-GPA Bv421-conjugated (BD Biosciences), and anti-α4-integrin (CD49d) allophycocyanin-conjugated (Miltenyi Biotec) antibodies for 30 min at 4 °C. Erythroblasts were defined as GPA^{pos} cells, and the maturation stage was determined by CD49d and band 3 expression. Immature erythroblasts (i.e., proerythroblasts) were $CD49d^{hi}$/band 3^{lo}, whereas more differentiated erythroblasts, such as orthochromatic erythroblasts, represented $CD49^{lo}$/band 3^{hi} populations.



Supplemental Figure 5. ABCC4 expression at the platelet membrane. Platelet-Rich Plasma (PRP) has been collected from fresh blood of a healthy donor (black profile) and an $ABCC4_{null}$ individual (red profile). FITC-conjugated mouse anti-CD41 was used to identify the platelet cells by flow cytometry and Anti-PEL eluate was used to measure the expression of ABCC4 at platelet surface. Briefly, cells were incubated with Anti-PEL and anti-CD41 antibodies for 1 hour at RT, and then revealed with anti-human IgG-PE (1:100, Beckman Coulter). PEL-labelled platelets were shown inside $CD41^{+}$ cell population.



Supplemental Figure 6. PCR genotyping assay of PEL-negative samples. For detection of *ABCC4* deletion/insertion, two PCRs were performed with different forward primers: WT-F and BK-F for the wild-type (690 bp) and mutated allele (702 bp including the intervening repeat 18-bp sequence), respectively. The PCR products were interpreted on the basis of positive or negative amplification. As an internal control, two additional primers (hGH-F and hGH-R) for *hGH* amplification (434 bp) were used. Lanes 1, 2, 3, 6, 7 and 8 correspond to PEL-negative samples, lanes 4 and 9 correspond to controls (PEL positive), and lanes 5 and 10 correspond to PCR without DNA. M; 100 bp DNA ladder. Primer sequences are available in supplemental table S4. Detailed PCR conditions are available upon request.

Supplemental Tables

Table S1. Primers used to amplify and sequence the breakpoint in *ABCC4*

Name	Sequence	Location	Direction	Position in Chr 13
ABCC4-FWD1	5' GCCATAGTCTCTCACCAAGAAG	Intron 20	Sense	95087062-95087083
ABCC4-AS1	5' GACTTACACATTCATTCAGGTGC	Downstream of 3 'UTR	Antisense	95015360-95015382
ABCC4-AS2	5' CAGGTTGGAGTGTAGTGG	Downstream of 3 'UTR	Antisense	95018199-95018217

Table S2. Primers used for PEL genotyping assay

Name	Sequence	Direction	Position in Chr 13
WT-F	5'CTGCACACTCTGCGATCTGT	Sense	95018863- 95018844
BK-F	TCAACCTTCATTTTGTGTCCTG	Sense	95086397- 95086376
BK-R	TGAGATGGAGTCTCGCTGTG	Antisense	95018174- 95018193

Table S3. Primers used to inactivate *ABCC4* gene by CRISPR/Cas-9

	Location	User name	Sequence (5'->3')
Guide 1	Exon 8	ABCC4-1	CACCGCACTGCCACGAACACGCGGC
		ABCC4-2	AAACGCCGCGTGTTCGTGGCAGTGC
Guide 2		ABCC4-3	CACCGCACACTGCCGAGGAGCACGT
		ABCC4-4	AAACACGTGCTCCTCGGCAGTGTGC
Guide 3	Exon 18	ABCC4-5	CACCGCCAGTTAAGATCTAGCTTCT
		ABCC4-6	AAACAGAAGCTAGATCTTAACTGGC
PCR for T7E1 assay	Exon 8	ABCC4-7	GCATTCCAGCACATGATACT
		ABCC4-10	CAGGGATAGCTAAGTGAAGTTG
	Exon 18	ABCC4-12	ATCCGCGGCATCATGTTCA
		ABCC4-14	TAGTGAAGCCCAGGAGATG
PX459 plasmid sequencing	PX459		GAGGGCCTATTTCCCATGATTCC

Table S4. Hematological parameters of PEL-negative individuals

	Proband 1	Proband 2	Proband 3	Proband 4	Reference values
WBCs (× 10^9/L)	8.60	8.00	6.50	7.20	4.00 - 11.00
RBCs (× 10^{12}/L)	4.81	4.78	4.48	4.36	3.80 - 6.50
HGB (g/dL)	14.40	13.40	12.70	14.00	12.00 – 18.00
HCT (%)	0.44	0.39	0.38	0.41	0.36 - 0.54
MCV (fL)	92	83	86	95	76 - 96
MCV (pg)	29.90	27.90	28.40	32.20	27.00-32.00
MCHC (g/dL)	32.70	33.70	33.10	33.90	31.00 – 35.00
PLT (× 10^9/L)	177	211	358	318	150 - 400
MPV (fL)	9.50	7.90	7.30	9.20	6.00 – 10.00