



**A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia.**

Lionel Arnaud, Carole Saison, Virginie Helias, Nicole Lucien, Dominique Steschenko, Marie-Catherine Giarratana, Claude Prehu, Bernard Foliguet, Lory Montout, Alexandre De Brevern, et al.

► **To cite this version:**

Lionel Arnaud, Carole Saison, Virginie Helias, Nicole Lucien, Dominique Steschenko, et al.. A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia.. American Journal of Human Genetics, Elsevier (Cell Press), 2010, 87 (5), pp.721-7. <10.1016/j.ajhg.2010.10.010>. <inserm-00550723>

**HAL Id: inserm-00550723**

**<http://www.hal.inserm.fr/inserm-00550723>**

Submitted on 29 Dec 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Arnaud *et al.*

## **A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia**

Lionel Arnaud,<sup>1,\*</sup> Carole Saison,<sup>1</sup> Virginie Helias,<sup>1</sup> Nicole Lucien,<sup>1</sup> Dominique Steschenko,<sup>2</sup>  
Marie-Catherine Giarratana,<sup>3</sup> Claude Prehu,<sup>4</sup> Bernard Foliguet,<sup>5</sup> Lory Montout,<sup>1,6,7</sup> Alexandre G.  
de Brevern,<sup>1,6,7</sup> Alain Francina,<sup>8</sup> Pierre Ripoché,<sup>1,7</sup> Odile Fenneteau,<sup>9</sup> Lydie Da Costa,<sup>9,10,19</sup>  
Thierry Peyrard,<sup>1,11</sup> Gail Coghlan,<sup>12</sup> Niels Illum,<sup>13</sup> Henrik Birgens,<sup>14</sup> Hannah Tamary,<sup>15</sup> Achille  
Iolascon,<sup>16,17</sup> Jean Delaunay,<sup>18</sup> Gil Tchernia,<sup>10,20</sup> and Jean-Pierre Cartron<sup>1</sup>

<sup>1</sup>Institut National de la Transfusion Sanguine, 75015 Paris, France ;

<sup>2</sup>Hôpital d'Enfants de Brabois, CHU de Nancy, 54511 Vandoeuvre-les-Nancy, France ;

<sup>3</sup>UMR\_S 938, INSERM, Faculté de Médecine Pierre et Marie Curie, Université Pierre et Marie  
Curie - Paris 6, 75012 Paris, France ;

<sup>4</sup>Laboratoire de Biochimie et Génétique, CHU Hôpital Henri Mondor, 94010 Créteil, France ;

<sup>5</sup>Service commun de microscopie électronique, Faculté de Médecine, Université de Nancy, 54506  
Vandoeuvre-les-Nancy, France ;

<sup>6</sup>Dynamique des Structures et Interactions des Macromolécules Biologiques, DSIMB, 75015  
Paris, France ;

<sup>7</sup>UMR\_S 665, INSERM, Institut National de la Transfusion Sanguine, Université Paris Diderot -  
Paris 7, 75015 Paris, France ;

<sup>8</sup>Hospices civils de Lyon, Hôpital Edouard Herriot, Unité de Pathologie Moléculaire du Globule  
Rouge, 69003 Lyon, France ;

<sup>9</sup>Laboratoire d'Hématologie Biologique, CHU Hôpital Robert Debré, 75019 Paris, France ;

Arnaud *et al.*

<sup>10</sup>Laboratoire d'Hématologie, Hôpital de Bicêtre, 94275 Le Kremlin Bicêtre, France ;

<sup>11</sup>Centre National de Référence pour les Groupes Sanguins, CNRGS, 75011 Paris, France ;

<sup>12</sup>Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, MB R3A 1S1  
Canada ;

<sup>13</sup>Hans Christian Andersen Children's Hospital, Odense University Hospital, 5000 Odense,  
Denmark ;

<sup>14</sup>Herlev Hospital, University of Copenhagen, 2730 Herlev, Denmark ;

<sup>15</sup>Pediatric Hematology Unit, Schneider Children's Medical Center of Israel, 49202 Petach  
Tikvah, Israel ;

<sup>16</sup>Department of Biochemistry and Medical Biotechnologies, University Federico II of Naples,  
80131 Naples, Italy ;

<sup>17</sup>CEINGE - Biotecnologie Avanzate, 80145 Naples, Italy ;

<sup>18</sup>UMR\_S 779, INSERM, Faculté de Médecine Paris-Sud, Université Paris-Sud, 94275 Le  
Kremlin-Bicêtre, France ;

<sup>19</sup>Present address : UMR 1009, INSERM, Institut de Cancérologie Gustave Roussy, Université  
Paris-Sud - Paris 11, 94805 Villejuif, France ;

<sup>20</sup>Present address : Ministère de la Santé, 75007 Paris, France ;

\*Correspondence : larnaud@ints.fr

## Summary

The congenital dyserythropoietic anemias (CDAs) are inherited red blood cell disorders, whose hallmarks are ineffective erythropoiesis, hemolysis and morphological abnormalities of erythroblasts in bone marrow. We have identified a missense mutation in *KLF1* of patients with a hitherto unclassified CDA. *KLF1* is an erythroid transcription factor and extensive studies in mouse models have shown it to play a critical role in the expression of globin genes, but also a wide spectrum of genes potentially essential for erythropoiesis. The unique features of this CDA confirm the key role of *KLF1* during human erythroid differentiation. Furthermore, we show that the mutation has a dominant negative effect on *KLF1* transcriptional activity and unexpectedly abolishes the expression of the water channel *AQP1* and the adhesion molecule *CD44*. Thus, the study of this disease-causing mutation in *KLF1* provides further insights on the roles of this transcription factor during erythropoiesis in human.

## Main text

The CDAs represent a heterogeneous group of rare congenital anemias predominantly caused by dyserythropoiesis in the bone marrow.<sup>1</sup> Three major (types I to III) and several minor subgroups have been differentiated, mainly according to the morphological abnormalities of erythroblast nuclei observed in bone marrow smears (*e.g.* chromatin bridges or double nuclei).<sup>2,3</sup> The gene responsible for CDA I (MIM 224120) was identified by positional cloning in 2002<sup>4</sup> and coined *CDANI* (MIM 607465) but its function remains to be elucidated. The gene responsible for CDA II (MIM 224100) has recently been shown to encode *SEC23B* (MIM 610512), a subunit of the COPII coat protein complex involved in the vesicular transport between the endoplasmic reticulum and Golgi apparatus, whose erythroid-specific role was unsuspected.<sup>5,6</sup> Although CDA I and CDA II represent most cases, the identification of causative genetic defects in other CDA

subgroups or in patients with unclassified CDA may offer further insights into the different pathways underlying erythropoiesis.

The first CDA patient investigated in this study (male patient ME) was born at 28 weeks' gestation to non-consanguineous healthy parents, in a context of acute fetal distress. Hydrops fetalis-associated anemia had been detected at 23 weeks' gestation and treated with two intrauterine transfusions; the karyotype of the fetus was normal. The neonatal examination revealed severe hyperbilirubinemia, hepatomegaly, hypertrophic cardiomyopathy as well as several dysmorphic features (micropenis, hypospadias, large anterior fontanel, hypertelorism). Anemia did not improve after birth and required transfusions. At 4 months of age, the analysis of bone marrow smears showed marked hyperplasia of the erythroid lineage, diagnosing CDA, but the dysplastic changes in the erythroblasts did not clearly fit any classification of CDA (**Figure S1**). Despite treatments with erythropoietin or interferon-alpha, the hemolytic anemia persisted and required recurrent transfusions (2-3 week interval) until a splenectomy was performed at 4 years of age (splenomegaly with no pathologic features). Shortly thereafter, transfusion independence was achieved with hemoglobin levels stabilized at around 8.0 g/dL (**Table S1**). At 13 years of age, patient ME showed short stature (height -3SD, weight -2SD) despite growth hormone therapy and treatment for hypothyroidism, and thalassemic facies.

A striking feature of patient ME's CDA was the very large number of nucleated red blood cells in his peripheral blood, *i.e.* 210% of white blood cells before splenectomy and up to 1,000% thereafter (**Figure 1A** and **Table S1**). Most of these circulating nucleated red blood cells were orthochromatic erythroblasts but only a few of them were enucleating, which suggested a failure of terminal erythroid differentiation. Their analysis by electron microscopy revealed various ultrastructural abnormalities, especially atypical cytoplasmic inclusions and enlarged nuclear pores (**Figure S2**). The *in vitro* study of erythroid differentiation of CD34<sup>+</sup> cells<sup>7</sup> isolated from

patient ME's peripheral blood showed normal proliferation and differentiation but impaired enucleation capacity (**Figure 1B**). Furthermore, when we analyzed by flow cytometry a panel of markers on the surface of his erythrocytes (**Figure S3**) we noticed the absence of CD44, which was confirmed by western blot analysis (**Figure 1C**), as well as reduced expression of two other adhesion molecules, BCAM and ICAM4. CD44 was similarly absent from his mature erythrocytes and circulating erythroblasts, but present on his granulocytes and all the other leukocyte populations (**Figure 1D** and **Figure S4**) suggesting that only the erythroid lineage was affected, consistent with a CDA trait. We also found that patient ME's erythrocytes were deficient in the water channel AQP1 (**Figure 1C**) and consequently, had a reduced water permeability as those of the very rare *AQP1*<sup>-/-</sup> individuals<sup>8</sup> (**Figure S5**). Patient ME's CDA was altogether unique and certainly different from CDA I and CDA II, as suggested by bone marrow analysis, which was later confirmed by the absence of mutations in *CDANI* and *SEC23B* (data not shown).

In order to identify the causative genetic defect, fresh blood samples were taken from patient ME and his relatives after obtaining a signed informed consent for genetic analysis, and genomic DNA was extracted. After exploring without success the possibility of an inherited recessive mutation by performing a SNP-based genome-wide screen in patient ME's family (Affymetrix GeneChip Human Mapping 250K Nsp Array, data not shown), we thought that this unique CDA might be caused by a *de novo* mutation in a transcription factor essential for expression of *CD44* (MIM 107269) and *AQP1* (MIM 107776) amongst others. Since CD44 deficiency was apparently restricted to the erythroid lineage, we first focused our analysis on erythroid transcription factors such as GATA1 or KLF1 (also known as EKLF).<sup>9,10</sup> GATA1 represented an attractive candidate<sup>11,12</sup> but no mutations in *GATA1* (MIM 305371) were detected (data not shown). Sequencing of *KLF1* (MIM 600599) (**Table S2**) in patient ME revealed the presence of two

Arnaud *et al.*

heterozygous mutations: a T to C transition in exon 2 (NM\_006563.3:c.304T>C, NP\_006554.1:p.Ser102Pro) and a G to A transition in exon 3 (NM\_006563.3:c.973G>A, NP\_006554.1:p.Glu325Lys) (**Figure S6**). A comparison with the NCBI dbSNP database (Build 131) showed that c.304T>C mutation corresponded to a previously reported SNP (rs2072597) and was unlikely pathogenic; this was consistent with rs2072597 heterozygosity of his healthy mother (data not shown). In contrast, c.973G>A mutation had never been reported and was not present in 96 regular blood donors nor in patient ME's relatives (**Figure 2A** and data not shown) suggesting it was the disease-causing mutation.

To confirm that *KLF1* mutation c.973G>A was responsible for this unique CDA, we searched for it in other patients. Female patient SF was extensively studied in the 90s and her atypical CDA showed striking similarities with patient's ME: combined deficiency of CD44 and AQP1 on erythrocytes,<sup>13</sup> circulating erythroblasts,<sup>14</sup> unique ultrastructural abnormalities in erythroblasts,<sup>15</sup> as well as increased electrophoretic mobility of the erythrocyte membrane protein Band 3<sup>13</sup> (**Figure S7**). When we sequenced *KLF1* in patient SF, we found the same heterozygous *KLF1* mutation c.973G>A as in patient SF (**Figure 2A**), corroborating that this mutation was responsible for this type of CDA. Of note, analysis of *KLF1* haplotypes of patients ME (c.[304T>C; 973G>A]+[=]) and SF (c.[973G>A]+[=]) does not support the hypothesis of a founder effect for the pathogenic *KLF1* mutation c.973G>A, which is consistent with a *de novo* mutation. In addition to the previously mentioned unique features of this CDA, patient SF was shown to express embryonic  $\zeta$ -globin chain in the majority of her erythrocytes, along with high levels of fetal hemoglobin.<sup>14</sup> When we analyzed patient ME's hemoglobin by isoelectric focusing (**Figure 1E**), we similarly found large amounts of fetal hemoglobin as well as an unusual hemoglobin, migrating like embryonic hemoglobin Portland ( $\zeta_2\gamma_2$  tetramer) which was confirmed

Arnaud *et al.*

by reverse phase liquid chromatography (**Figure S8**). Thus, *KLF1* mutation c.973G>A was associated with a profound dysregulation of globin gene expression and provided *in vivo* evidence of the critical role KLF1 in this complex transcriptional regulation in human, as predicted by thalassemia-associated mutations in the proximal KLF1 binding sites of  $\beta$ -globin gene<sup>16</sup> and extensive studies in mouse models.<sup>17-19</sup> Interestingly, while this manuscript was being finalized, Borg *et al.*<sup>20</sup> reported the identification of a defective *KLF1* allele in a large Maltese family which causes the persistence of high levels of fetal hemoglobin in adults - a benign and usually asymptomatic condition known to alleviate the severity of  $\beta$ -thalassemia and sickle cell disease - and further demonstrated that KLF1 indirectly downregulates fetal globin gene expression by activating *BCL11A* expression.

The pathogenic *KLF1* mutation c.973G>A results in the substitution of the evolutionary conserved glutamate 325 by a lysine (E325K) in the second zinc finger (ZF2) of KLF1 (**Figure 2B**). As arginine 322 and arginine 328, glutamate 325 is predicted to contact DNA<sup>16</sup> (**Figure 2C**). In order to investigate the possible structural effect of E325K mutation, we built a new structural model for the zinc finger domain of KLF1 based on the X-ray structures of Wilms' tumor protein in complex with DNA.<sup>21</sup> The currently used models of KLF1 are based on the structure of Zif268 bound to DNA<sup>22</sup> but alignment of their respective zinc-finger domains requires a two amino-acid gap in KLF1 ZF2, which significantly changes the local topology. According to the new structural model, glutamate 325 is located on a helix and directed toward the DNA backbone (**Figure 2D, upper left panel**). The single amino acid mutation E325K does not alter the overall topology of KLF1 zinc finger domain, but mainly affects the side chain of residue 325 (**Figure 2D, lower panels**). Actually, this charge reversal mutation enhances the electrostatic interaction between KLF1 and the DNA, and potentially creates a novel hydrogen bond (**Figure 2D, upper**

**right panel**). Thus, E325K mutation is predicted to stabilize, rather than disrupt, the binding of KLF1 to its DNA target sequences.

Before testing the potential effect of E325K mutation on KLF1 transcriptional activity, we first wanted to check whether this mutation did not affect the stability or cellular localization of the mutant protein. For this purpose, we constructed vectors encoding either wild type (WT) or mutant (E325K) KLF1 tagged with a Flag epitope at the N-terminus to allow its detection, and transfected them into human erythroid K-562 cells. Flag-tagged KLF1 E325K showed the same expression level as the wild type protein by western blot analysis (**Figure 3A**) and showed the same nuclear localization by immunofluorescence analysis (**Figure 3B**). The combined deficiency of CD44 and AQP1 observed in this unique CDA suggested that *AQP1* and *CD44* genes were direct targets of KLF1, which was consistent with the presence of several potential KLF1 binding sites (CCNCNCCCN)<sup>16</sup> upstream of their respective initiating codons. Therefore, we studied the effect of E235K mutation on KLF1 transcriptional activity with *CD44* and *AQP1* promoter-reporter assays in K-562 cells. Flag-tagged KLF1 WT was able to activate *CD44* and *AQP1* promoters, whereas the mutant E325K showed a markedly reduced transcriptional activity (**Figures 3C and 3D**). Furthermore, when we co-expressed KLF1 E325K with KLF1 WT in order to mimic heterozygosity of patients ME and SF, we observed that KLF1 E325K was able to inhibit the activation of *CD44* promoter induced by KLF1 WT (**Figure 3D**). On the basis of this data, we conclude that E325K mutation has a dominant negative effect on the transcriptional activity of KLF1, in total agreement with the phenotype of heterozygous patients.

This study describes a missense *KLF1* mutation responsible for a human pathology, a hitherto unclassified CDA. Interestingly, a new CDA patient with *KLF1* mutation c.973G>A has just been identified (female patient SE; A.I., unpublished data) suggesting that this type of CDA may be less rare than expected. The previously reported *KLF1* mutations, found in the

heterozygous state too, are responsible for the blood group In(Lu) phenotype<sup>23</sup> (MIM 111150) which is mainly characterized by a reduced expression of Lu (also known as BCAM) on erythrocytes, but not associated with any pathology. Of note, patient SF does not show a reduced expression of Lu (**Figure S9**) indicating that *KLF1* mutation E325K is not directly responsible for the reduced expression of Lu observed in patient ME (**Figure S3**). We assume that distinct *KLF1* mutations may differently affect the gene repertoire of this transcription factor, thus leading to different phenotypes as observed with *GATA1* mutations.<sup>11,12</sup> Fortuitously, while this manuscript was being finalized, Siateckla *et al.*<sup>24</sup> reported that the mouse mutation *Nan*, a dominant ethylnitrosourea-induced mutation causing hemolytic anemia, corresponds to a missense *Klf1* mutation (E339D) affecting the expression of only a subset of its target genes, and further showed that E339D mutation indeed alters Klf1 DNA binding to only a subset of its target sequences. Of note, glutamate 339 in mouse Klf1 is the equivalent of glutamate 325 in human KLF1; however, *Klf1* mutation E339D is intrinsically different from *KLF1* mutation E325K due to the opposite charge of the mutant residue (see structure prediction above), which is consistent with the distinct resulting pathologies in mouse and human. Extensive studies with mouse models suggested that KLF1 played a global role during erythropoiesis by regulating a wide spectrum of genes<sup>19,25,26</sup> but also organizing nuclear hubs for efficient and coordinated transcription of genes co-regulated with  $\beta$ -globin gene.<sup>27,28</sup> This study provides *in vivo* evidence that human KLF1 plays a critical role in the regulation of fetal globin genes, as recently demonstrated by Borg *et al.*<sup>20</sup> (see above), but also other unexpected genes like *AQP1* or *CD44*. While AQP1 deficiency in human doesn't cause dyserythropoiesis,<sup>8</sup> we can not exclude that CD44 deficiency may contribute, even though *CD44* is dispensable for mouse erythropoiesis.<sup>29</sup> Future studies will be required to identify the KLF1 target genes whose expression is essential for human erythropoiesis

Arnaud *et al.*

but deficient in this form of CDA. Finally, we propose that *KLF1* should be systematically sequenced as a novel candidate gene in all CDA cases with unknown genetic cause, which may eventually lead to the discovery of other pathogenic *KLF1* mutations.

### **Supplemental Data**

Supplemental Data include nine figures and two tables.

### **Acknowledgements**

First, we would like to thank the patients and patient ME's relatives for providing blood samples for this study. We would like also to thank D. Sommelet, J. Buisine, L. Douay, M. Goossens, C. Etchebest, Y. Colin, T. Zelinski, P. Gane, T. Cynober, M. Feneant-Thibault, C. Schmitt, R. Russo, M.R. Esposito, C. Menanteau, S. Kappler-Gratias, B.-N. Pham, P.-Y. Le Penec, E. Vera, C. Verheyde, G. Nicolas, B.A. Ballif and M. Le Gall for their contributions to this study. L.A. and J.-P.C. were supported by the National Institute of Blood Transfusion (INTS), the National Institute for Health and Medical Research (INSERM) and Paris Diderot University (Paris 7). G.C. was supported by an operating grant from the Winnipeg Rh Institute Foundation to Dr T. Zelinski. H.T. was supported by an Israel Science Foundation grant (No. 699/\_03-18.4), by an Israeli Ministry of Science, Culture and Sport grant, in the framework of the Israel-France Program, and by an Israeli Ministry of Science-Eshkol Fellowship. A.I. was supported by the Italian Ministry of University and Research (grant MUR-PS 35-126/Ind), the Italian Telethon Foundation (project GGP09044) and Regione Campania (DGRC 1901/2009).

### **Web resources**

The URLs for data presented herein are as follows:

Arnaud *et al.*

Online Mendelian Inheritance in Man (OMIM) browser, <http://www.ncbi.nlm.nih.gov/omim/>

Entrez Gene browser, <http://www.ncbi.nlm.nih.gov/gene/>

Entrez Single Nucleotide Polymorphism (SNP) browser, <http://www.ncbi.nlm.nih.gov/snp/>

Structure modeling of KLF1, <http://www.dsimb.inserm.fr/~debrevern/KLF1>

## Figure Titles and Legends

### **Figure 1. Analysis of the peripheral blood of CDA patient ME featuring unique abnormalities.**

(A) Peripheral blood smear from the patient (right panels, sample taken on 2008/10/28) or a control (left panel, sample taken at the same time) stained with May-Grünwald Giemsa, showing the large number of circulating erythroblasts (purple nuclei) as well as poikilocytosis, anisocytosis and fragmented erythrocytes in the patient; the scale bars represent 40  $\mu\text{m}$ .

(B) Study of the number of enucleated cells during *in vitro* erythroid differentiation of CD34<sup>+</sup> cells from the patient (purple, sample taken on 2005/10/25) or a control donor (blue), showing markedly reduced enucleation capacity of the patient's cell culture.

(C) Western blot analysis of CD44, CD55, AQP1 and p55 in erythrocyte membrane lysates from the patient (sample taken on 2004/06/07) his healthy mother (C2) or a random control (C1), showing the combined deficiency of CD44 and AQP1 in the patient's erythrocytes.

(D) Flow cytometry analysis of CD44 on mature erythrocytes (left panels) and granulocytes (right panels) from the patient (bottom panels, sample taken on 2010/01/05) and a control (top panels, sample taken at the same time) showing the erythroid-specific deficiency of CD44 in the patient. For the analysis of erythrocytes, whole blood samples were co-stained with fluorochrome-conjugated anti-CD44 (black histogram) or isotype control antibody (white histogram) and anti-CD71, and the mature erythrocytes were gated on FSS, SCC and CD71<sup>-</sup> to eliminate the reticulocytes and erythroblasts; for the analysis of granulocytes, nucleated blood cells were first isolated by hypotonic erythrocyte lysis, co-stained as above but directly gated on FSC and SSC.

Arnaud *et al.*

(E) Isoelectric focusing analysis of the different hemoglobin (Hb) variants in the patient (sample taken on 2008/10/28) and two controls (C1 and C2) showing atypical globin expression in the patient. The patient exhibited very high levels of fetal Hb ( $\alpha_2\gamma_2$  tetramer, 37.3%, normal range: less than 1%) as well as embryonic Hb Portland ( $\zeta_2\gamma_2$  tetramer, 2.9%, normal range: absent) as ascertained by reverse phase liquid chromatography; adult HbA and HbA2 were at 55.5% ( $\alpha_2\beta_2$  tetramer, normal range: 90-100%) and 1.2% ( $\alpha_2\delta_2$  tetramer, normal range: 2-3%) respectively. Extensive sequencing of patient ME's globin loci detected no gross abnormalities but a heterozygous mutation in  $\alpha 1$ -globin gene (c.62\_63insT, p.His21fsX36) which could not be responsible for the profound  $\beta$ -globin locus dysregulation and was indeed present in his healthy paternal aunt. Of note, patient SF was carrier of a 4bp deletion in the promoter of  $\Lambda\gamma$ -globin gene, as her healthy father<sup>14</sup>.

All studied blood samples were taken after splenectomy and at least 6 months post transfusion.

**Figure 2. Identification of the *KLF1* mutation associated with this CDA and structural analysis of the mutant transcription factor.**

(A) Detail of *KLF1* sequencing in patient ME, his unaffected mother and sister, and unrelated patient SF, showing the same heterozygous *KLF1* mutation in both CDA patients. The experimental sequences were aligned with the NCBI reference sequence of *KLF1* (NG\_013087.1). Genomic DNA samples from patient ME's father and patient SF's parents were not available.

(B) Diagram showing the structure of *KLF1* (based on NG\_013087.1) and its products, with the localization of the mutation c.973G>A, p.Glu325Lys (E325K) found in both CDA patients. *KLF1* consists of 3 exons (boxes; black color represents coding regions and grey represents untranslated

Arnaud *et al.*

regions) and encodes a peptide of 362 amino acids with a proline-rich domain at the amino-terminus (transactivation domain in brown) and three C<sub>2</sub>H<sub>2</sub>-type zinc fingers (ZF) at the carboxy-terminus (DNA-binding domain in green). The pathogenic mutation is located in the third exon and encodes a single amino acid change in the second zinc finger of the transcription factor.

(C) Sequence alignment of the second zinc finger of KLF1 from various mammalian species, showing the conservation of a glutamate at amino acid position 325 (red box). The two cysteines and the two histidines contacting Zn<sup>2+</sup> are indicated in bold, and the three conserved residues contacting the DNA are indicated by stars.

(D) Modeling structure of wild type (left) and variant E325K (right) zinc-finger domain of KLF1 showing the enhanced electrostatic interaction between the variant E325K and the DNA backbone. The change of glutamate to lysine at position 325 has a double effect by reversing the side-chain charge from negative to positive, and extending the side-chain length toward the negatively charged DNA backbone; the putative hydrogen bond created by E325K mutation is shown as a yellow dashed line. The bottom panels show the overall structure of the protein-DNA complexes, as well as the electrostatic potential surfaces (negative in red, positive in blue) of wild type and variant proteins; the top panels show a close-up view of a cartoon representation of the region around residue 325, highlighting as sticks the side chain of residue 325 (orange) and the closest nucleotide (light blue); of note, residue 325 is not oriented toward the nucleotide base but the phosphate.

**Figure 3. Characterization of the effect of E325K mutation on KLF1 function.**

(A) Western blot analysis of transfected human erythroid K-562 cells showing that E325K mutation does not affect the stability of KLF1. Constructs encoding Flag-tagged KLF1 wild type (WT) or mutant (E325K), or empty vector, were transfected in K-562 cells, and protein

Arnaud *et al.*

expression was analyzed after 24h by western blot with an anti-Flag and an anti-GAPDH (loading control).

(B) Immunofluorescence analysis of transfected K-562 cells showing that E325K mutation does not affect the nuclear localization of KLF1. K-562 cells transfected as in (a) were analyzed after 24h by immunofluorescence with an anti-Flag (green) along with propidium iodide for DNA staining (red). No anti-Flag staining was detected in K-562 cells transfected with the empty vector (data not shown).

(C) *AQPI* promoter-reporter assay in K-562 cells showing that E325K mutation affects the transcriptional activity of KLF1. K-562 cells were co-transfected with an *AQPI* promoter-Photinus luciferase (Pluc) construct and a *HSV-TK* promoter-Renilla luciferase (Rluc) construct for normalization, along with 2  $\mu$ g of the indicated *KLF1* constructs, and the luciferase activities were analyzed after 24h; the results are shown as means  $\pm$  s.d. (n=3) of Pluc activity normalized by Rluc activity.

(D) *CD44* promoter-reporter assay in K-562 cells showing that E325K mutation has a dominant negative effect on the transcription activity of KLF1. Not only has KLF1 E325K a markedly reduced transcriptional activity, but also it is able to impede the transcriptional activity of co-expressed KLF1 WT. K-562 cells were co-transfected with a *CD44* promoter-Pluc construct and a *HSV-TK* promoter-Rluc construct for normalization, along with 2 or 1  $\mu$ g of the *KLF1* constructs as indicated, and the luciferase activities were analyzed after 24h; the results are shown as means  $\pm$  s.d. (n=3) of Pluc activity normalized by Rluc activity.

## References

- 1 R. Renella, W.G. Wood (2009) The congenital dyserythropoietic anemias. *Hematol Oncol Clin North Am* 23, 283-306
- 2 H. Heimpel, F. Wendt, D. Klemm, H. Schubotho, L. Heilmeyer (1968) [Congenital dyserythropoietic anemia]. *Arch Klin Med* 215, 174-94
- 3 H. Heimpel, K. Kellermann, N. Neuschwander, J. Hogel, K. Schwarz (2010) The morphological diagnosis of congenital dyserythropoietic anemia: results of a quantitative analysis of peripheral blood and bone marrow cells. *Haematologica* 95, 1034-6
- 4 O. Dgany, N. Avidan, J. Delaunay, T. Krasnov, L. Shalmon, H. Shalev, T. Eidelitz-Markus, J. Kapelushnik, D. Cattani, A. Pariente *et al.* (2002) Congenital dyserythropoietic anemia type I is caused by mutations in codanin-1. *Am J Hum Genet* 71, 1467-74
- 5 K. Schwarz, A. Iolascon, F. Verissimo, N.S. Trede, W. Horsley, W. Chen, B.H. Paw, K.P. Hopfner, K. Holzmann, R. Russo *et al.* (2009) Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II. *Nat Genet* 41, 936-40
- 6 P. Bianchi, E. Fermo, C. Vercellati, C. Boschetti, W. Barcellini, A. Iurlo, A.P. Marcello, P.G. Righetti, A. Zanella (2009) Congenital dyserythropoietic anemia type II (CDAII) is caused by mutations in the SEC23B gene. *Hum Mutat* 30, 1292-8
- 7 M.C. Giarratana, L. Kobari, H. Lapillonne, D. Chalmers, L. Kiger, T. Cynober, M.C. Marden, H. Wajcman, L. Douay (2005) Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells. *Nat Biotechnol* 23, 69-74
- 8 G.M. Preston, B.L. Smith, M.L. Zeidel, J.J. Moulds, P. Agre (1994) Mutations in aquaporin-1 in phenotypically normal humans without functional CHIP water channels. *Science* 265, 1585-7
- 9 S.I. Kim, E.H. Bresnick (2007) Transcriptional control of erythropoiesis: emerging mechanisms and principles. *Oncogene* 26, 6777-94
- 10 V.G. Sankaran, J. Xu, S.H. Orkin (2010) Advances in the understanding of haemoglobin switching. *Br J Haematol* 149, 181-94
- 11 L.M. Hollanda, C.S. Lima, A.F. Cunha, D.M. Albuquerque, J. Vassallo, M.C. Ozelo, P.P. Joazeiro, S.T. Saad, F.F. Costa (2006) An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis. *Nat Genet* 38, 807-12
- 12 K.E. Nichols, J.D. Crispino, M. Poncz, J.G. White, S.H. Orkin, J.M. Maris, M.J. Weiss (2000) Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet* 24, 266-70

Arnaud *et al.*

- 13 S.F. Parsons, J. Jones, D.J. Anstee, P.A. Judson, B. Gardner, E. Wiener, J. Poole, N. Illum, S.N. Wickramasinghe (1994) A novel form of congenital dyserythropoietic anemia associated with deficiency of erythroid CD44 and a unique blood group phenotype [In(a-b-), Co(a-b-)]. *Blood* 83, 860-8
- 14 W. Tang, S.P. Cai, B. Eng, M.C. Poon, J.S. Waye, N. Illum, D.H. Chui (1993) Expression of embryonic zeta-globin and epsilon-globin chains in a 10-year-old girl with congenital anemia. *Blood* 81, 1636-40
- 15 S.N. Wickramasinghe, N. Illum, P.D. Wimberley (1991) Congenital dyserythropoietic anaemia with novel intra-erythroblastic and intra-erythrocytic inclusions. *Br J Haematol* 79, 322-30
- 16 W.C. Feng, C.M. Southwood, J.J. Bieker (1994) Analyses of beta-thalassemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem* 269, 1493-500
- 17 B. Nuez, D. Michalovich, A. Bygrave, R. Ploemacher, F. Grosveld (1995) Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature* 375, 316-8
- 18 A.C. Perkins, A.H. Sharpe, S.H. Orkin (1995) Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature* 375, 318-22
- 19 R. Drissen, M. von Lindern, A. Kolbus, S. Driegen, P. Steinlein, H. Beug, F. Grosveld, S. Philipsen (2005) The erythroid phenotype of EKLF-null mice: defects in hemoglobin metabolism and membrane stability. *Mol Cell Biol* 25, 5205-14
- 20 J. Borg, P. Papadopoulos, M. Georgitsi, L. Gutierrez, G. Grech, P. Fanis, M. Phylactides, A.J. Verkerk, P.J. van der Spek, C.A. Scerri *et al.* (2010) Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat Genet* 42, 801-5
- 21 R. Stoll, B.M. Lee, E.W. Debler, J.H. Laity, I.A. Wilson, H.J. Dyson, P.E. Wright (2007) Structure of the Wilms tumor suppressor protein zinc finger domain bound to DNA. *J Mol Biol* 372, 1227-45
- 22 M. Elrod-Erickson, T.E. Benson, C.O. Pabo (1998) High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* 6, 451-64
- 23 B.K. Singleton, N.M. Burton, C. Green, R.L. Brady, D.J. Anstee (2008) Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype. *Blood* 112, 2081-8
- 24 M. Siatecka, K.E. Sahr, S.G. Andersen, M. Mezei, J.J. Bieker, L.L. Peters (2010) Severe anemia in the Nan mutant mouse caused by sequence-selective disruption of erythroid Kruppel-like factor. *Proc Natl Acad Sci U S A* 107, 15151-6

Arnaud *et al.*

25 D. Hodge, E. Coghill, J. Keys, T. Maguire, B. Hartmann, A. McDowall, M. Weiss, S. Grimmond, A. Perkins (2006) A global role for EKLF in definitive and primitive erythropoiesis. *Blood* 107, 3359-70

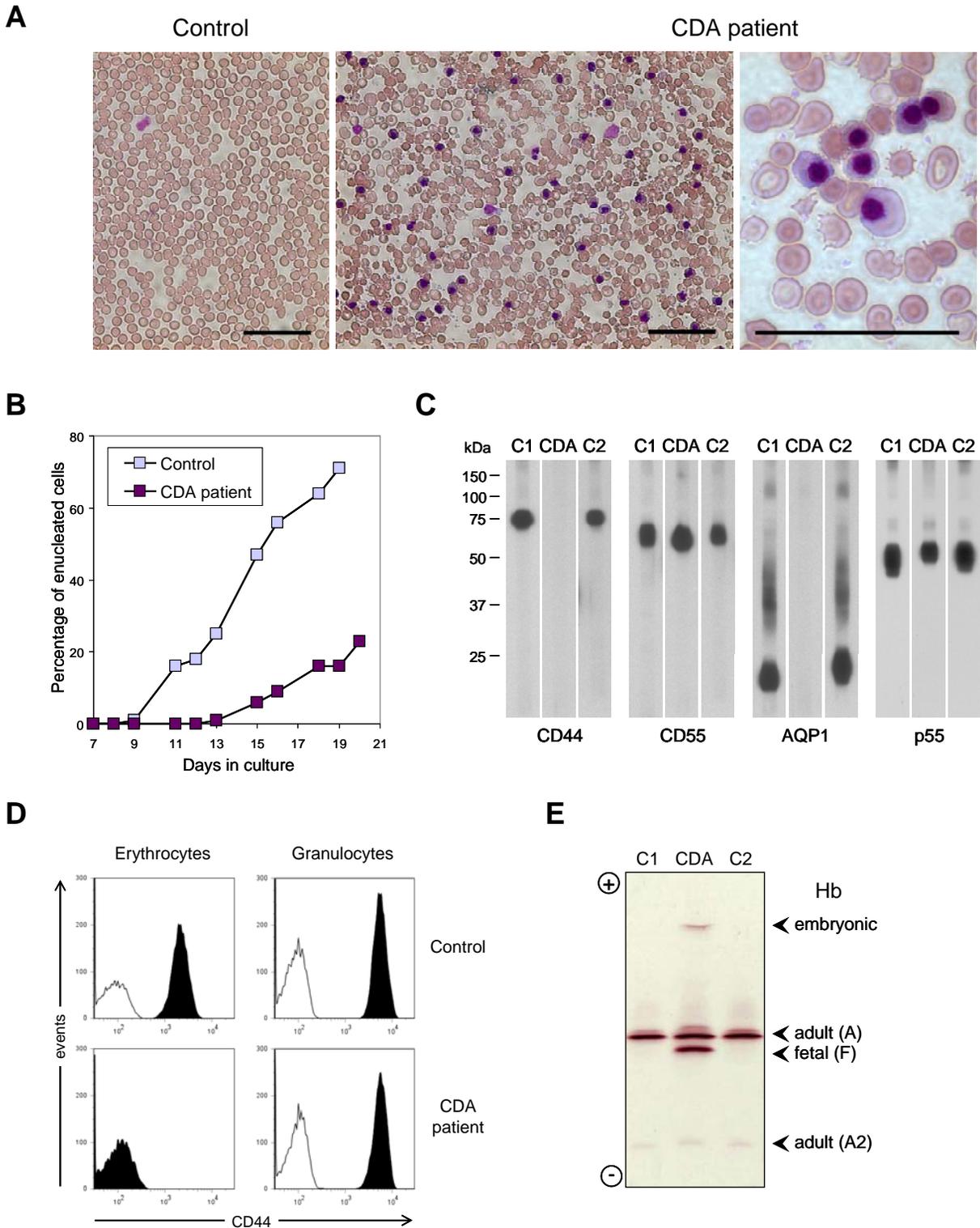
26 M.R. Tallack, T. Whittington, W.S. Yuen, E.N. Wainwright, J.R. Keys, B.B. Gardiner, E. Nourbakhsh, N. Cloonan, S.M. Grimmond, T.L. Bailey *et al.* (2010) A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. *Genome Res* 20, 1052-63

27 G.P. Patrinos, M. de Krom, E. de Boer, A. Langeveld, A.M. Imam, J. Strouboulis, W. de Laat, F.G. Grosveld (2004) Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev* 18, 1495-509

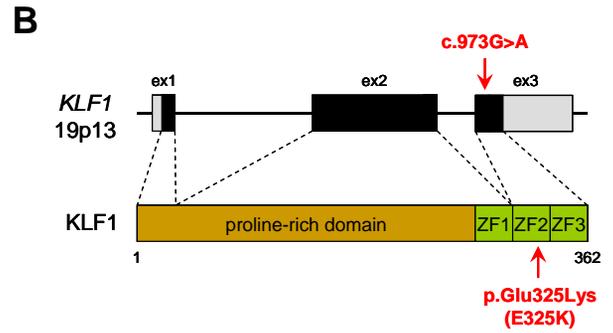
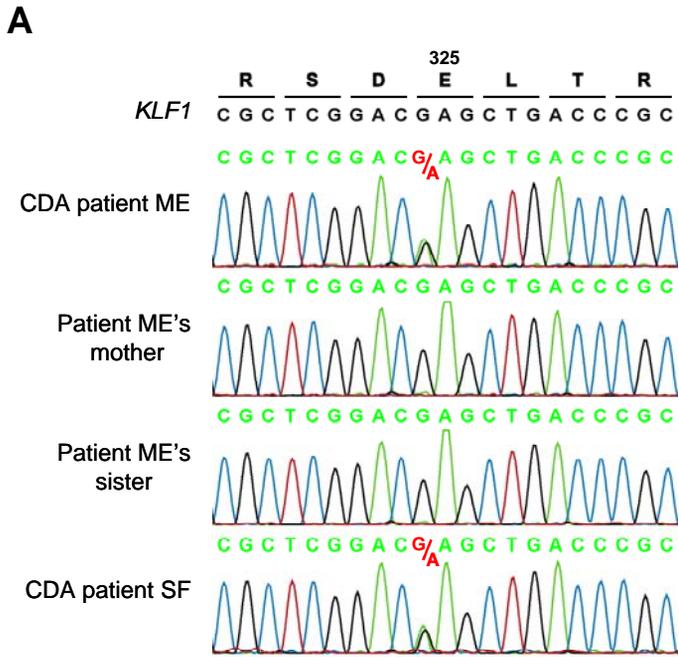
28 S. Schoenfelder, T. Sexton, L. Chakalova, N.F. Cope, A. Horton, S. Andrews, S. Kurukuti, J.A. Mitchell, D. Umlauf, D.S. Dimitrova *et al.* (2010) Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* 42, 53-61

29 R. Schmits, J. Filmus, N. Gerwin, G. Senaldi, F. Kiefer, T. Kundig, A. Wakeham, A. Shahinian, C. Catzavelos, J. Rak *et al.* (1997) CD44 regulates hematopoietic progenitor distribution, granuloma formation, and tumorigenicity. *Blood* 90, 2217-33

# Arnaud *et al.*, Figure 1



# Arnaud *et al.*, Figure 2

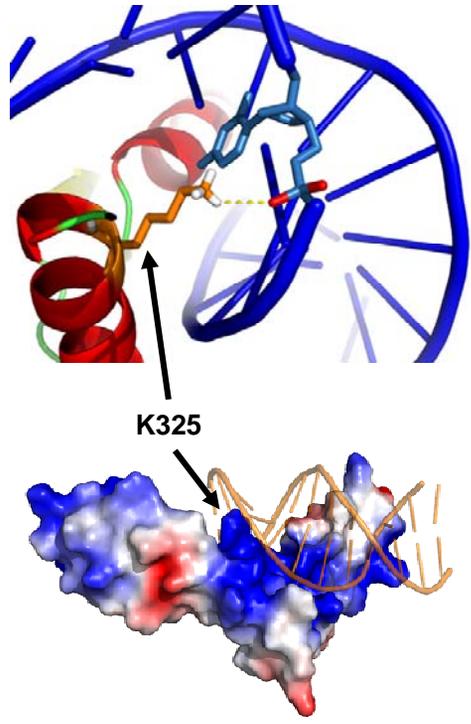
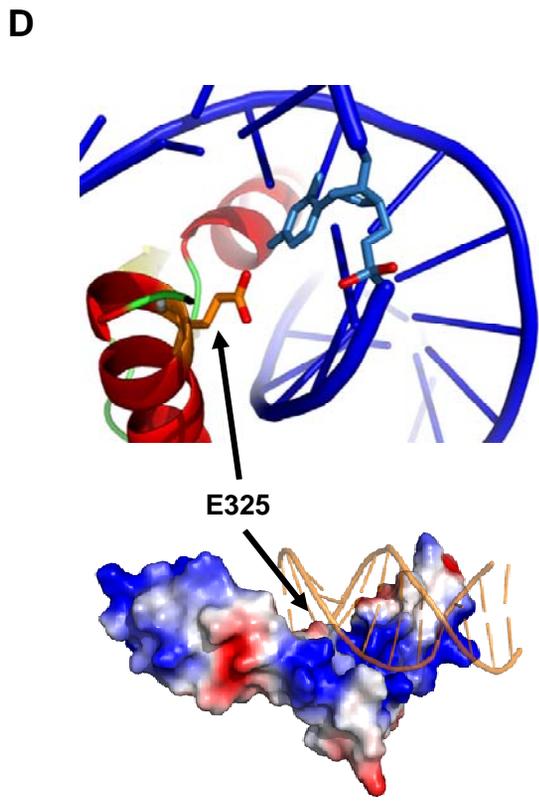


**C**

KLF1 ZF2

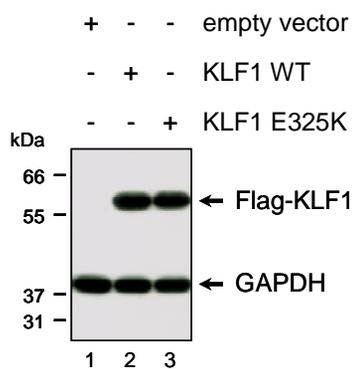
human	GEKPYA	<del>C</del>	TWEG	C	G	W	R	F	A	R	S	<del>D</del>	E	L	T	R	H	Y	R	K	H	T	G			
macaque	GEKPYA	<del>C</del>	TWEG	C	G	W	R	F	A	R	S	<del>D</del>	E	L	T	R	H	Y	R	K	H	T	G			
cow	GEKPYA	<del>C</del>	TW	D	G	C	G	W	R	F	A	R	S	<del>D</del>	E	L	T	R	H	Y	R	K	H	T	G	
pig	GEKPYA	<del>C</del>	TW	D	G	C	G	W	R	F	A	R	S	<del>D</del>	E	L	T	R	H	Y	R	K	H	T	G	
dog	GEKPYA	<del>C</del>	TW	D	G	C	G	W	R	F	A	R	S	<del>D</del>	E	L	T	R	H	Y	R	K	H	T	G	
rabbit	GEKPYA	<del>C</del>	TW	D	G	C	G	W	K	F	A	R	S	<del>D</del>	E	L	T	R	H	Y	R	K	H	T	G	
mouse	GEKPYA	<del>C</del>	S	W	D	G	C	D	W	R	F	A	R	S	<del>D</del>	E	L	T	R	H	Y	R	K	H	T	G
rat	GEKPYA	<del>C</del>	S	W	D	G	C	N	W	R	F	A	R	S	<del>D</del>	E	L	T	R	H	Y	R	K	H	T	G

\* \* \*

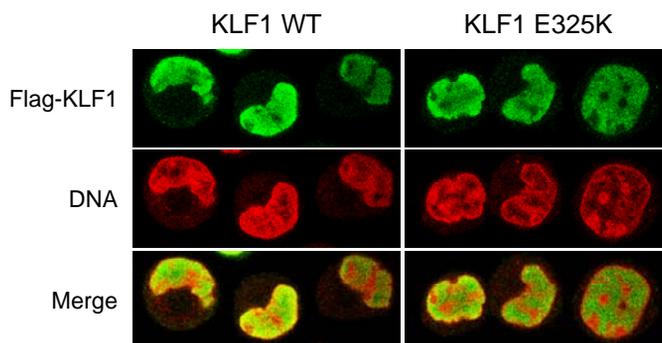


# Arnaud *et al.*, Figure 3

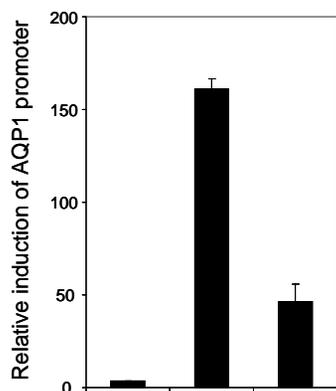
**A**



**B**

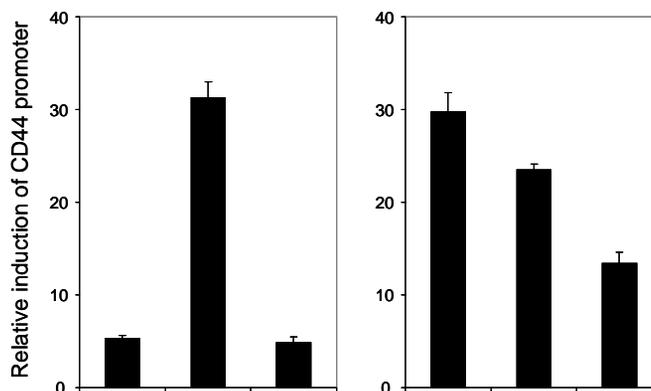


**C**



empty vector	2	-	-
KLF1 WT	-	2	-
KLF1 E325K	-	-	2

**D**



empty vector	2	-	-
KLF1 WT	-	2	-
KLF1 E325K	-	-	2

empty vector	-	1	-
KLF1 WT	2	1	1
KLF1 E325K	-	-	1

Arnaud *et al.*

**The American Journal of Human Genetics**

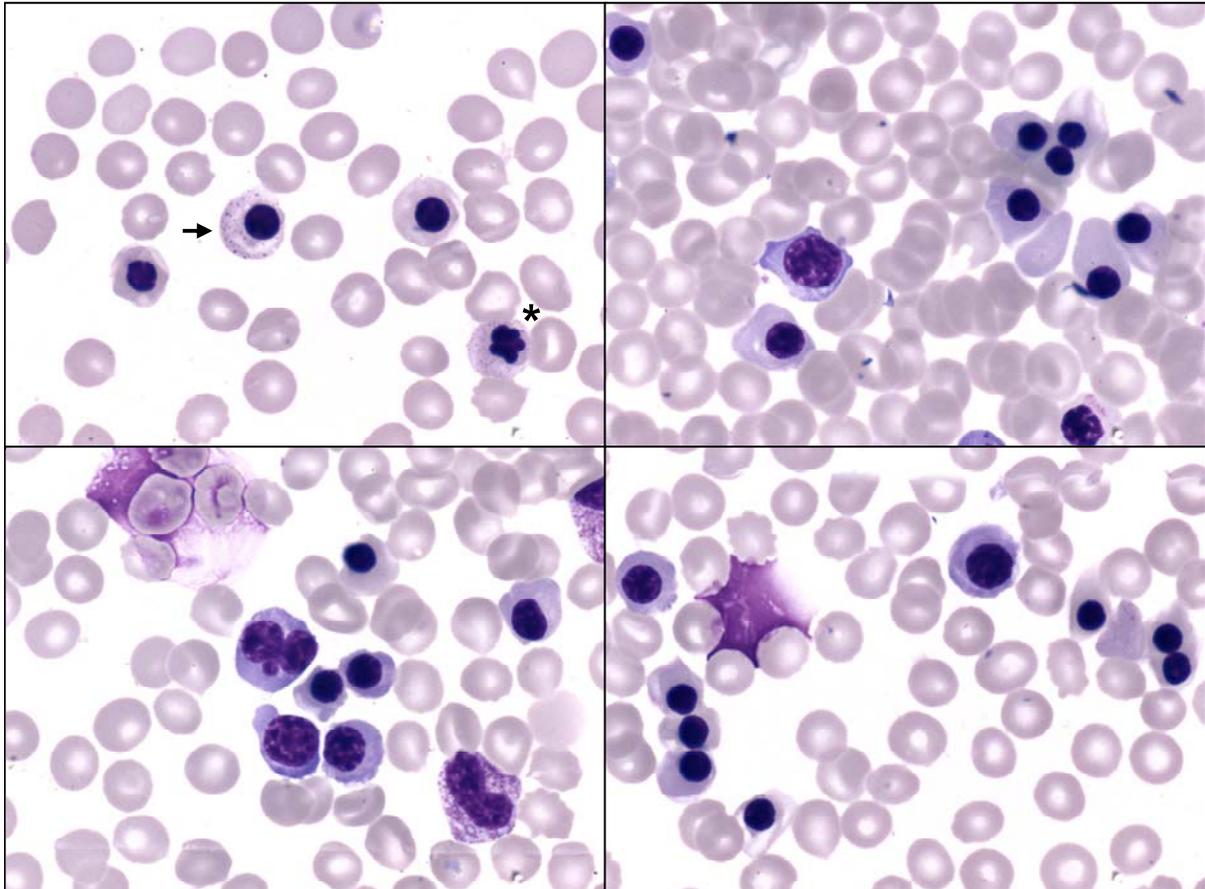
## **Supplemental Data**

### **A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia**

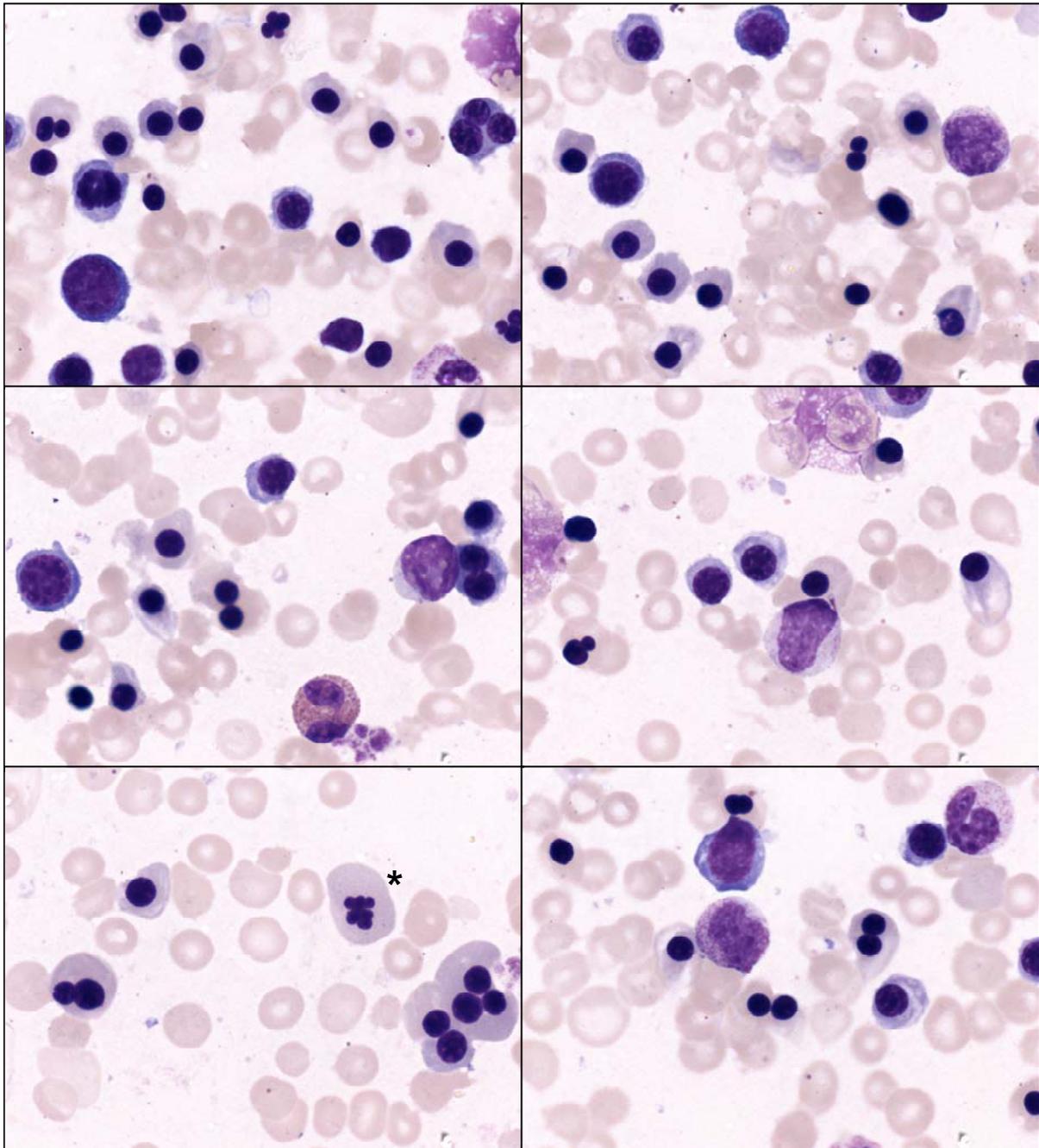
Lionel Arnaud, Carole Saison, Virginie Helias, Nicole Lucien, Dominique Steschenko, Marie-Catherine Giarratana, Claude Prehu, Bernard Foliguet, Lory Montout, Alexandre G. de Brevern, Alain Francina, Pierre Ripoche, Odile Fenneteau, Lydie Da Costa, Thierry Peyrard, Gail Coghlan, Niels Illum, Henrik Birgens, Hannah Tamary, Achille Iolascon, Jean Delaunay, Gil Tchernia and Jean-Pierre Cartron.

**Nine Figures and Two Tables**

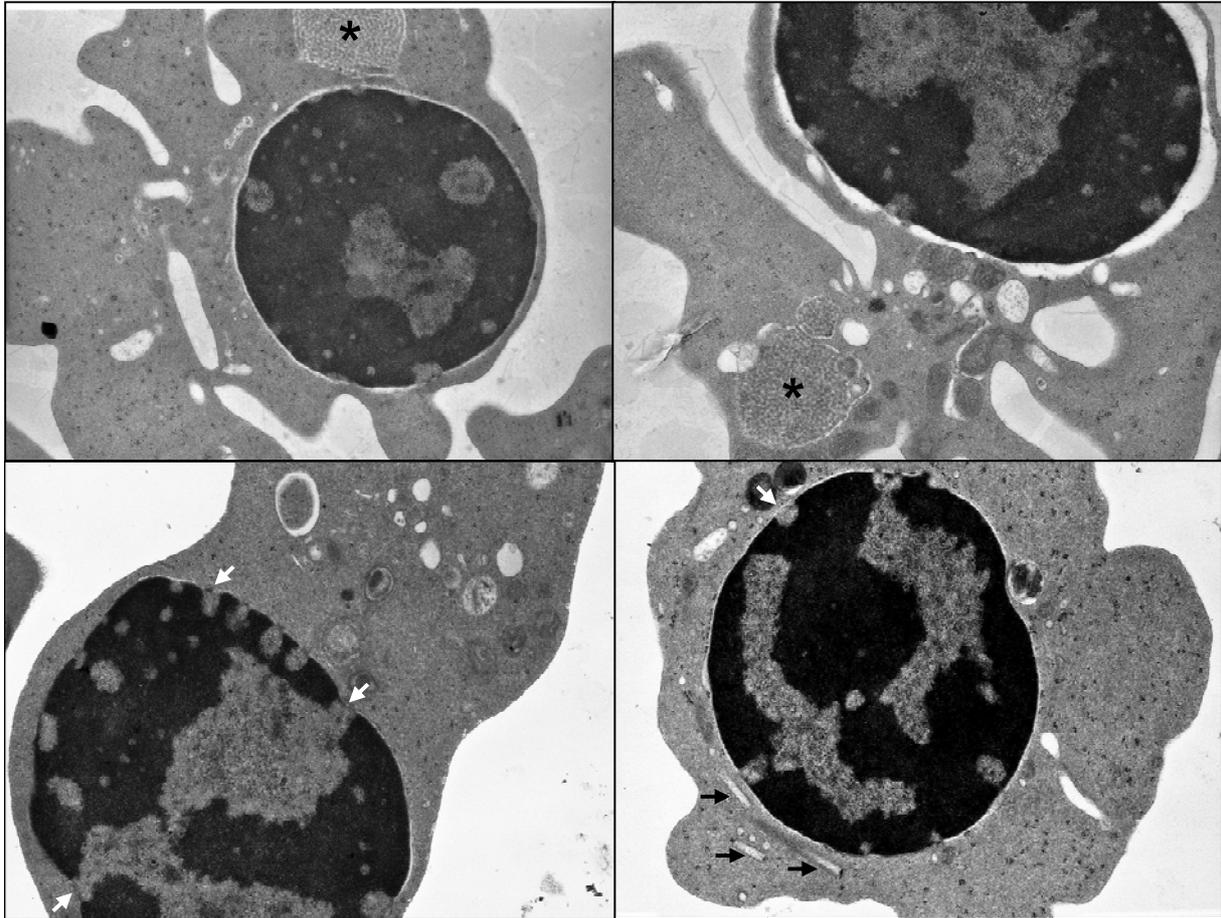
**Figure S1. (A) Bone marrow smear from the CDA patient ME before splenectomy** (sample taken on 1997/07/14) showing marked erythroid hyperplasia (77%) with a majority of acidophilic erythroblasts (58%). Binucleated erythroblasts (1.0%), multinucleated erythroblasts (1.5%) and few karyorrhexis (star) are observed. In some erythroblasts, the cytoplasm contains fine or coarse basophilic stippling of pale grayish-blue (arrow).



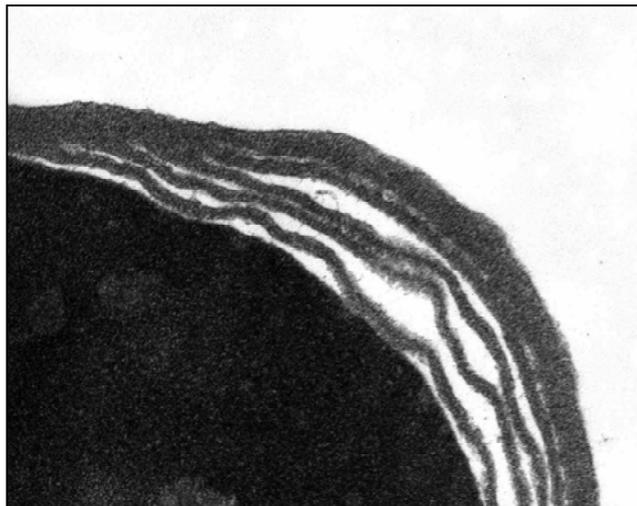
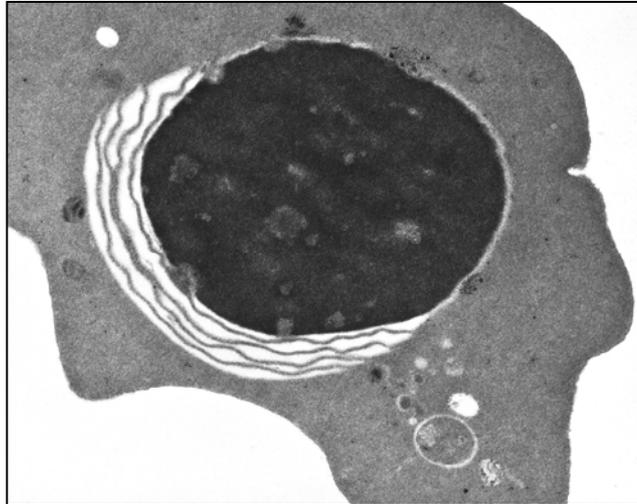
**Figure S1. (B) Bone marrow smear from the CDA patient ME after splenectomy** (sample taken on 2001/02/07) showing marked erythroid hyperplasia (80%) with a majority of acidophilic erythroblasts (55%). Binucleated erythroblasts (4.5%), multinucleated erythroblasts (2.5%) and few karyorrhexis (star) are observed.



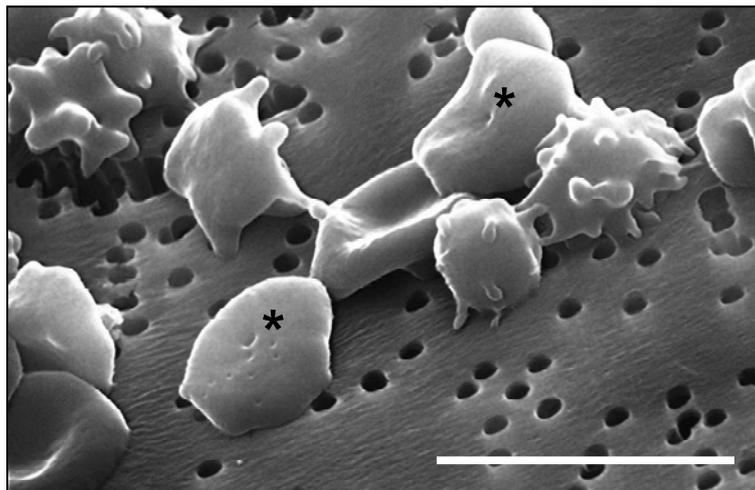
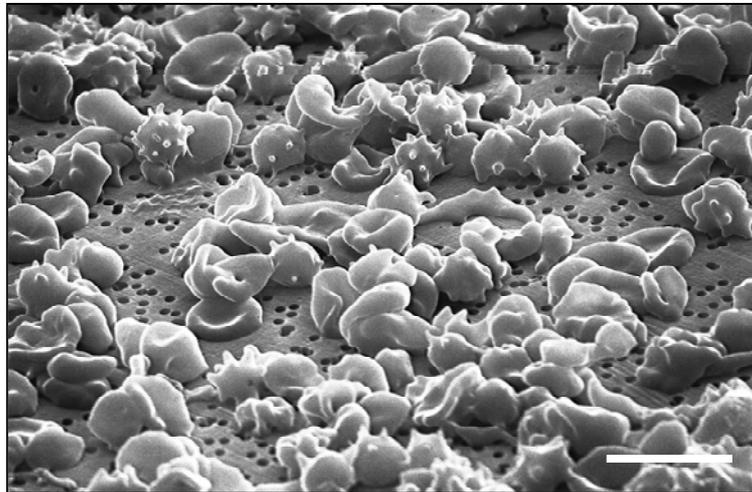
**Figure S2. (A) Transmission electron microscopy of circulating erythroblasts of the CDA patient ME showing tubulosaccular inclusions (stars) and elongated cylindrical structures (black arrows) in the cytoplasm, and abnormally large nuclear pores (white arrows). Original magnification  $\times 10,000$ .**



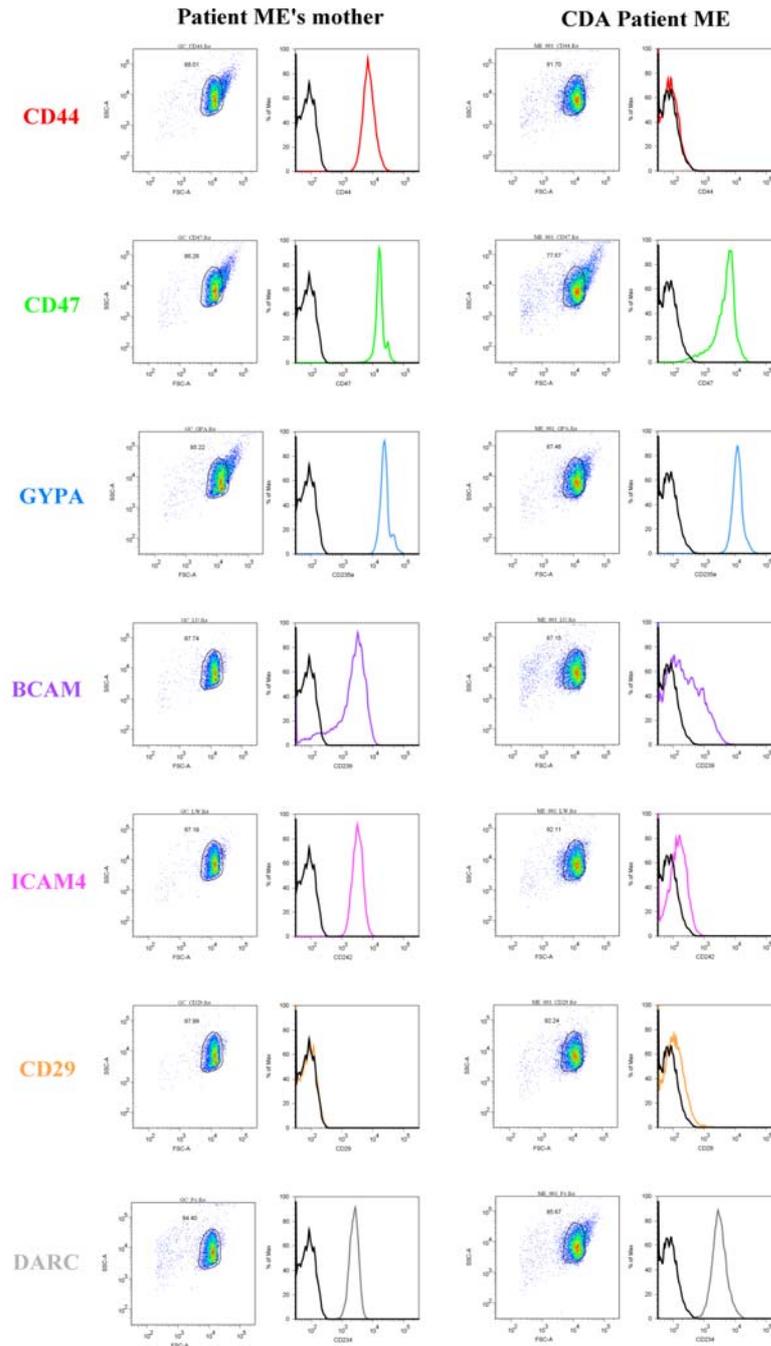
**Figure S2. (B) Transmission electron microscopy of circulating erythroblasts of the CDA patient ME showing a wide area of endoreplication of the nuclear membrane. Original magnification  $\times 8,000$  (top panel) and  $\times 35,000$  (bottom panel).**



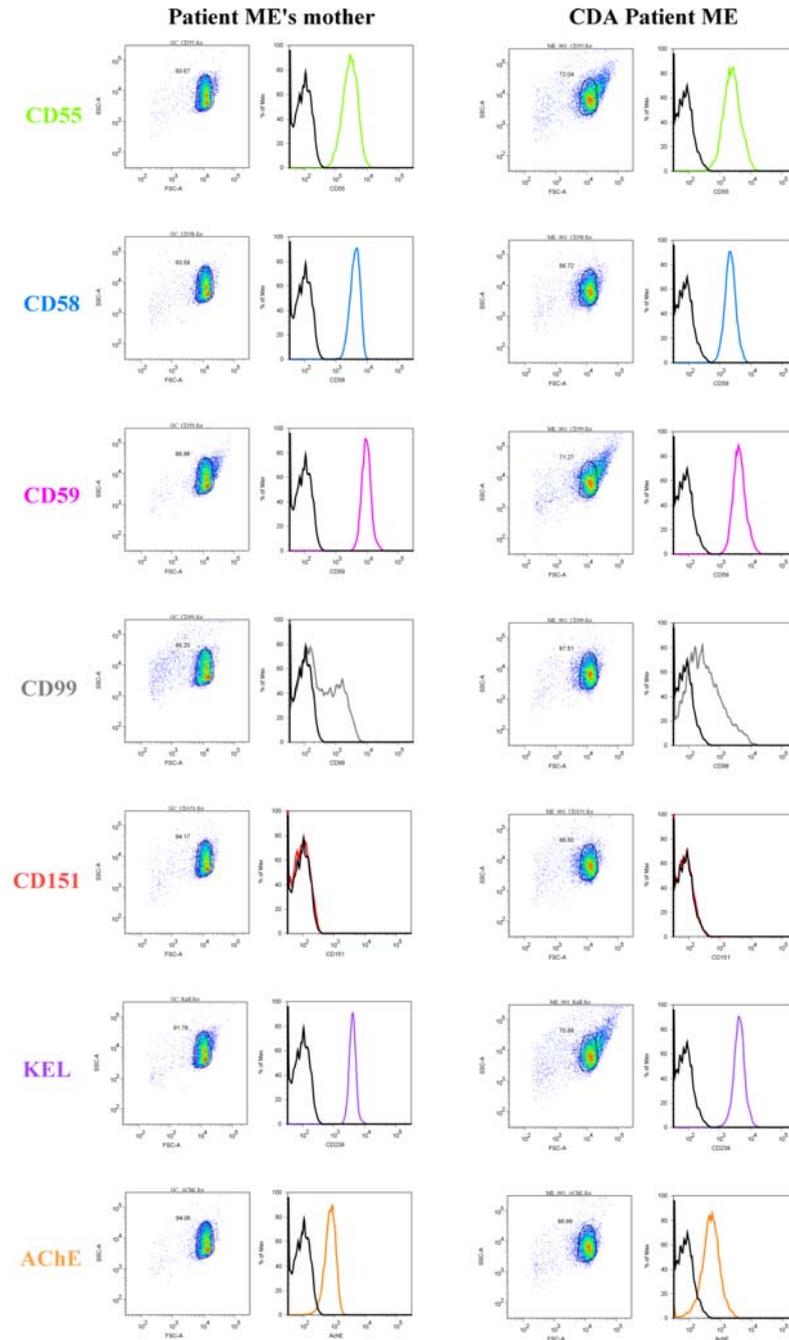
**Figure S2. (C) Scanning electron microscopy of blood of the CDA patient ME showing severe anisopoikilocytosis (top panel, original magnification  $\times 1,930$ ) and small invaginations in erythrocyte membrane (stars, bottom panel, original magnification  $\times 4,590$ ). The scale bars represent 10  $\mu\text{m}$ .**



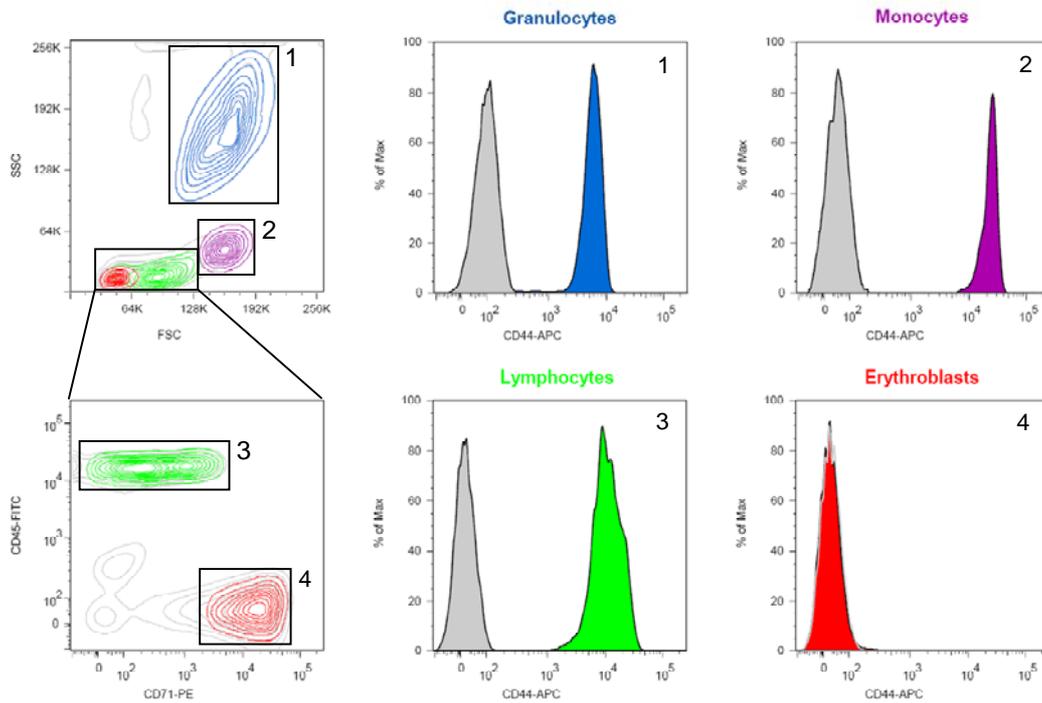
**Figure S3. (A) Flow cytometry analysis of cell markers CD44, CD47, GYPA, BCAM/Lu, ICAM4/LW, CD29 and DARC on erythrocytes of the CDA patient ME (right histograms, sample taken on 2010/01/05) or his healthy mother (left histograms, sample taken at the same time). Erythrocytes were stained with mouse monoclonal antibodies (profiles in color) or an isotype control antibody (profile in black) and phycoerythrin-conjugated goat F(ab')<sub>2</sub> anti-mouse IgG(H+L), and were gated on FSC and SSC (density dot plots). This blood sample from patient ME was taken 4 years after his most recent transfusion, and thus was free of exogenous erythrocytes. Patient ME's erythrocytes abnormally showed no expression of CD44, reduced expression of BCAM and ICAM4, and slightly increased expression of CD29. Similar results have consistently been observed on samples taken since 2003.**



**Figure S3. (B) Flow cytometry analysis of cell markers CD55, CD58, CD59, CD99, CD151, KEL and AChE on erythrocytes of the CDA patient ME (right histograms, sample taken on 2010/01/05) or his healthy mother (left histograms, sample taken at the same time).** Erythrocytes were stained with mouse monoclonal antibodies (profiles in color) or an isotype control antibody (profile in black) and phycoerythrin-conjugated goat F(ab')<sub>2</sub> anti-mouse IgG(H+L), and were gated on FSC and SSC (density dot plots). This blood sample from patient ME was taken 4 years after his most recent transfusion, and thus was free of exogenous erythrocytes. The expression of CD99 is highly variable and patient ME's profile is not abnormal.



**Figure S4. Flow cytometry analysis of CD44 expression in the nucleated cells present in the CDA patient ME's peripheral blood**, showing absence of CD44 on his erythroblasts but presence on his leukocyte populations. Nucleated blood cells were isolated by hypotonic erythrocyte lysis and triple stained with anti-CD44-APC or isotype antibody-APC, anti-CD71-PE and anti-CD45-FITC; granulocytes (panel 1) and monocytes (panel 2) were directly gated on FSC and SSC (top left contour plot), while lymphocytes (CD71<sup>-</sup>/CD45<sup>+</sup>, panel 3) and erythroblasts (CD71<sup>+</sup>/CD45<sup>-</sup>, panel 4) were further gated on CD71 and CD45 expression (bottom left contour plot); the labeling corresponding to anti-CD44-APC (colored histograms) or isotype antibody-APC (grey histograms) was analyzed with the same voltage for all cell populations.

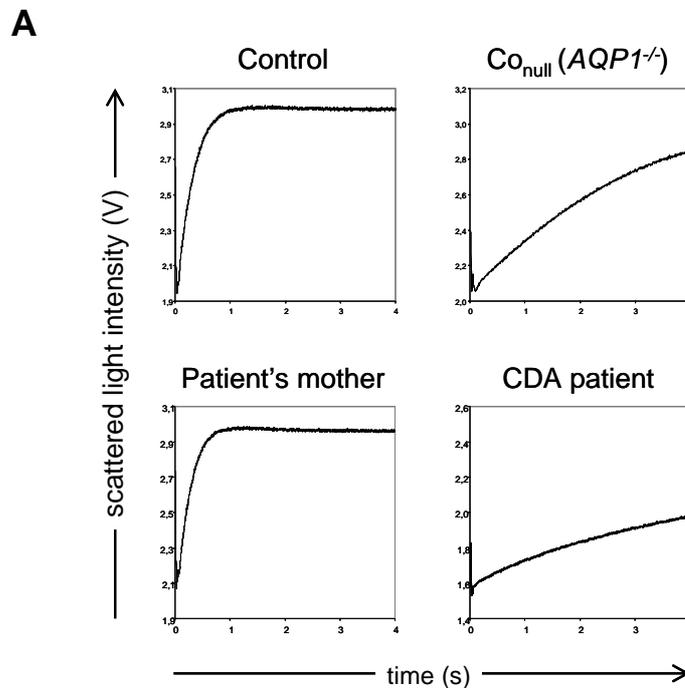


**Figure S5. Analysis of the water permeability of the CDA patient ME's erythrocytes confirming his AQP1 deficiency.**

(A) Stopped-flow light scattering measurements of erythrocytes from a control donor (top left panel), an *AQP1*<sup>-/-</sup> person (also known as Co<sub>null</sub> blood group phenotype, top right panel), the CDA patient ME (bottom right panel) and his healthy mother (bottom left panel) in response to a 150 mOsm inwardly directed gradient of mannitol at 15°C; the rapid increase in light scattering results from the shrinkage of erythrocytes due to water efflux, allowing to determine their osmotic water permeability.

(B) Osmotic water permeability coefficients (*Pf*) of indicated erythrocytes. As the extremely rare *AQP1*<sup>-/-</sup> individuals, patient ME's erythrocytes showed severely reduced water permeability. Of note, AQP1 is a major erythrocyte membrane protein that carries the Colton blood group antigens as well as ABO blood group antigens.

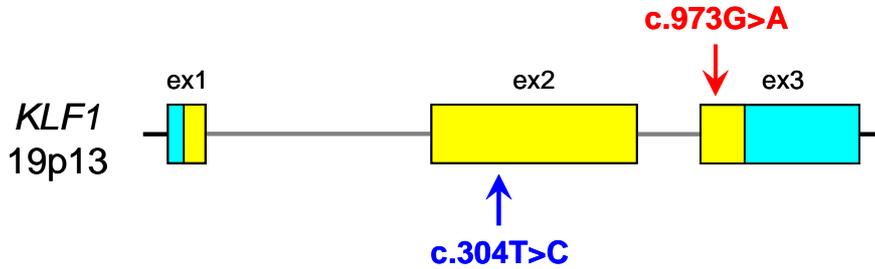
Due to AQP1 deficiency of patient ME's erythrocytes, their Colton phenotype is expected to be Co(a-b-) while blood group genotyping (BioArraySolutions HEA BeadChip Kit) predicts the following extended phenotype: C+E-c+e+, K-k+, Kp(a-b+), Js(a-b+), Fy(a+b+), Jk(a+b-), M+N-S-s+, Lu(a-b+), Di(a-b+), Co(a+b-), Do(a-b+), Hy+Jo(a+), LW(a+b-), Sc:1,-2.



**B**

Samples	ABO type	<i>Pf</i> (cm.s <sup>-1</sup> ) at 15°C	
		+ HgCl <sub>2</sub>	
CDA patient	A	0.002485	0.001635
Patient's mother	A	0.036400	0.003250
Control	A	0.031300	0.003760
Co <sub>null</sub> ( <i>AQP1</i> <sup>-/-</sup> )	O	0.004185	0.003930
Controls	O	0.030025	0.003441
(n=4)		± 0.004470	± 0.000343

**Figure S6. (A) Experimental sequence data of *KLF1* from the CDA patient ME compared with genomic NCBI Reference Sequence NC\_000019.8.**



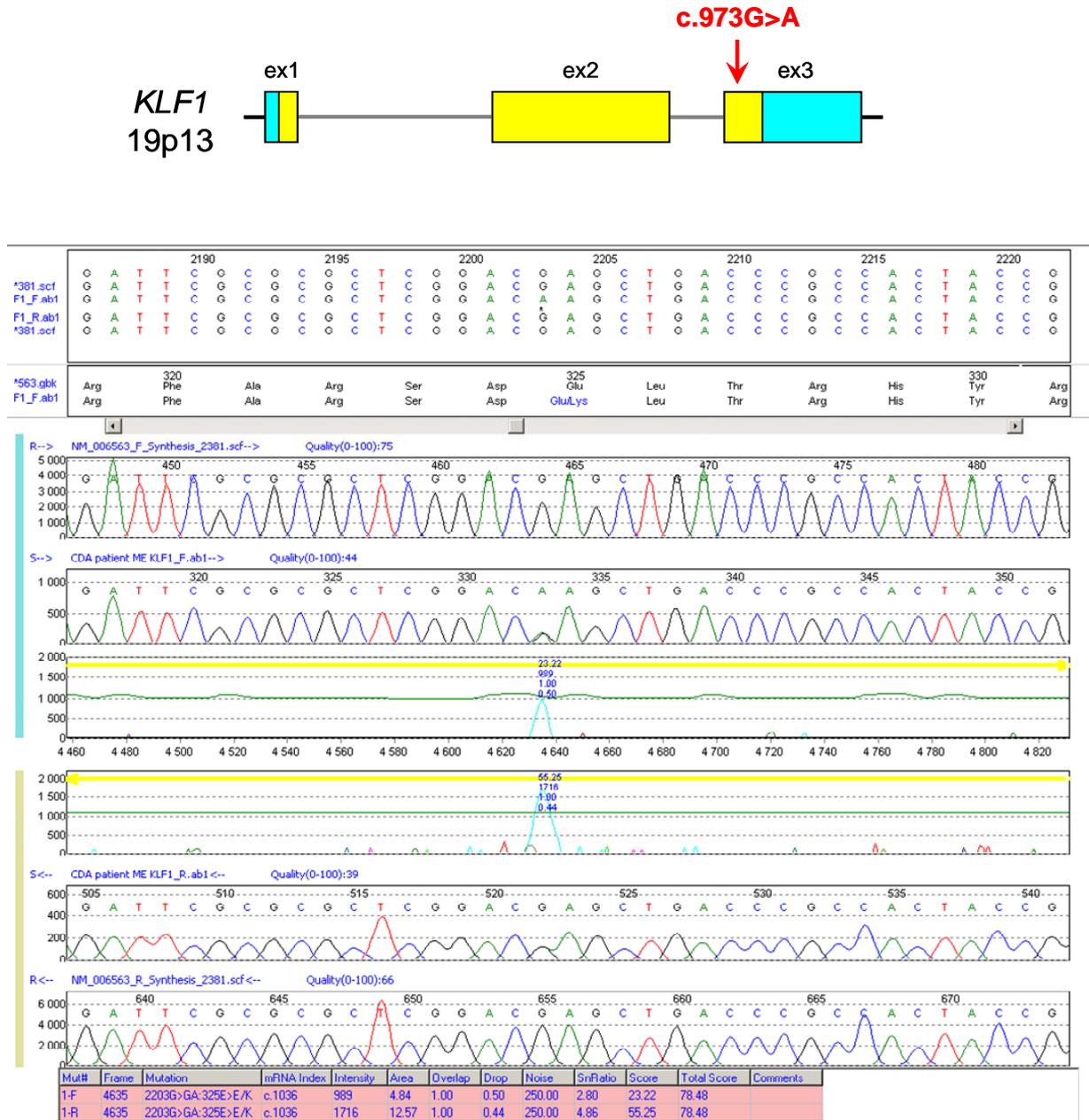
NNN coding sequence  
NNN untranslated sequence  
NNN intronic sequence

```

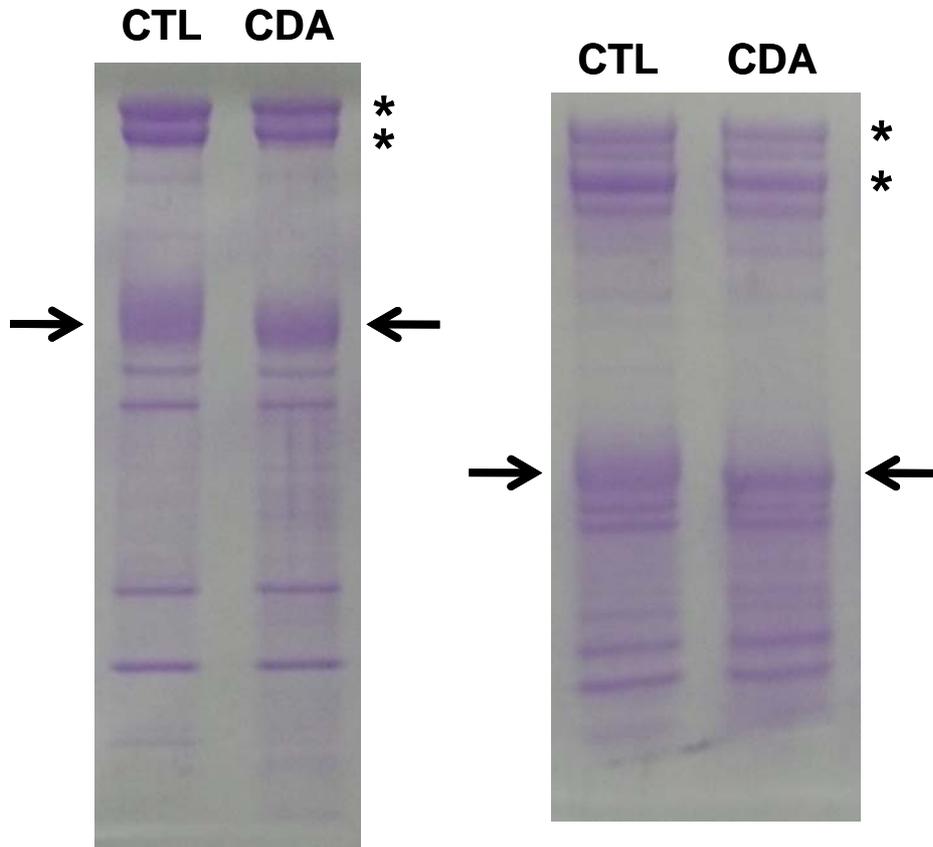
GCCGCCTTGCTTGGCTTATCAGAGGCTGCAGCCAATCAGCTAAGGACAGAGAGGAGCCCTCGAAGGGGCTATCACAG
CCTCAGAGTTCACGAGGCAGCCGAGGAAGAGGAGGCTTGAGGCCAGGGTGGGCACCAGCCAGCCATGGCCACAGCCGAGACC
GCCTTGCCCTCCATCAGCACACTGACCGCCCTGGGCCCTTCCCGGACACACAGGATGACTTCCTCAAGGTGGGGCCTAGAAG
GTGGGGTCTAGGTGGGCTGGCTGGAATCCAGGGCCACAGTCACAGATCTTGGGGTCCAGACCTGCATCTTGACCTGAAATCAA
GAGACTTAACCAGGACTGAGGTACGCTCAGTCCAGGAGAGGAGATCTCAGCTTAGTCTGGCAGGGGGTGGAGGGTGGTCTA
GGGGTTGAGGTTCTAAGTGTGATCTATTTTCGGTAATAGAAAACGAAGGTAGCCTGGGCAACATGGTGAACCCCTATCTCTAC
AAAAAATACCAAAAACATTAGGCCAGGCATGGGGGCGTGTGCCGTGTAGTCCCAGTACTCCGTAGGCTGATGCAGGAGGATCA
TTAGAGCCCAGGAGATTAAGGATACAGTGCAGCTGCACACTGCACCTCCAGCCTGGGCAAAAGAGTAAGACCCATCTCAAGAA
AAAAAAAAAAAAAAAAAGGAACGAGATCTAGGCTCACAGACAATCTTCCAGATATCAGCGGGAAATGATGAGTGTGTCTGGGGGA
CATCCAAAATTTTCGGAATTAATCTTGTTTTGGGAGACAGGGAAGGAGAGGGATGTTCTGGGGGAAAACATAAGTCAAGGCTGGC
ATCCTCTCCCCGTCCCTGCCAGTTTCCATCTCCAGCAGCTGTGCTACCCCTTCCCCATCCCCGAGTGTGGTTCAGATA
GTGGAAGTCTTATCTCCTGTCTCCAGCCAGACCTGATCGGTTTCTGTCCCTGGAGCTGGGGGGGAGCGGGGAGAGGGCGGT
TAGAGGGGCAGTGTGGGGAAGTGGGACAGACAGACAGGCAACAAGACCCCTTCCAAAGCCTCTGCGTCAGAGTGTCCAGC
CCGCGATGTCCCTGGGCAGGGCACCCAGTGTCCACCAACCTCGAGCTGCCTGTCCCTCCCGCAGTGGTGGCGCTCCGAAG
AGGCGCAGGACATGGGCCCGGGTCCCTCGACCCACGGAGCCGCCCTCCACGTGAAGTCTGAGGACCAGCCCGGGGAGGAA
GAGGACGATGAGAGGGGCGCGGACGCCACCTGGGACCTGGATCTCCTCCTCACCACCTTCTCGGGCCCGGAGCCCGGTGGCCG
GCCCCAGACCTGCGCTCTGGCGCCACGAGGGCCYCCGGGGCGCAATATCCGCCGCCCGCCGAGACTCTGGGCGCATATGCTG
GCGGCCCGGGGCTGGTGGCTGGGCTTTTGGGTTTCGGAGGATCACTCGGGTGGGTGGCCCTGCCTGCGAGCCCGGGCTCC
GACGCTCTCGTGGGCCAGCCCTGGCTCCAGCCCGGCCCGGACGCCAAGGCGCTGGCGCTGCAACCGGTGTACCCGGGGCC
CGGCGCCGCTCCTCGGTTGGCTACTTCCCGCGGACCGGGCTTTCAGTGCCTGGCGCGTCCGGCGCCCTACGGGCTACTGT
CCGGGTACCCCGCATGTACCCGGCGCCTCAGTACCAAGGGCACCTCCAGCTCTTCCCGGGCTCCAGGGACCCGCGCCCGGT
CCCGCCACGTCCCTCCTTCTGAGTTGTTTGGGACCCGGGACGGTGGGCACTGGACTCGGGGGGACTGCAGAGGATCCAGG
TGTGATAGCCGAGACCCGCGCCATCCAAGCGAGGCCAGCTTCTGTGGCGCGCAAGAGGCAGGCAGCGCACACGTGCGCGCAC
CGGGTTGCGGCAAGAGCTACACCAAGAGCTCCACCTGAAGGCGCATCTGCGCACGCACACAGGTGAGGGGGCGGGGCCCCGG
ACATGAGAAAGGGCGCGCGCCCGCTGTAGTTACAGGGGAAGAAGGTTGCAGAGGGCGGGACTTGGACTTGGCTGGCCTCTG
AGAGTGAGTGCCTCCTTAAATTTTGTGCCCTAGGGGCTCACTTGTTCATCCTAGTCCCAGCCAGGCTGAGTAAAGGGGTG
TGGCCAGATGCAGGGGACCCGGGACATGACTGGGACAGACAGTGGCGCTTATGGCTTCCTTGTCCCTAGGGGAGAAGCCATA
CGCCTGCACGTGGGAAGGCTGCGGCTGGAGATTGCGCGCTCGGACRAGCTGACCCGCACTACCGGAAACACACGGGGCAGC
GCCCTTCCGCTGCCAGCTCTGCCACGTGCTTTTTCGCGCTCTGACCACCTGGCCTTGACATGAAGCGCCACCTTTGAGCC
CTGCCCTGGCACTTGGACTCTCCTAGTACTGGGGATGGGACAAGAAGCCTGTTTGGTGGTCTCTTACACGGACGCGCGTGA
CACAATGCTGGGTGGTTTTCCACGAATGGACCTCTCCTGGACTCGCGTTCCTCCAAAGATCCACCAAATATCAAACACGGAC
CCATAGACAGCCCTGGGGGAGCCTCTTACGGAAAATCCGACAAGCCTTCAGCCACAGGGAGCCACACAGAGATGTCCAAACTG
TCGTGCAAAACCCAGTGAGACAGACCGCCAAATAAACGGACTCAGTGGACACTCAGACCAGCTCCAGATGGCCCTGGACAGCA
GGAGAGGGTGTGGGATGAGGCTTCCAGAGACCTGGGTCTAGAAAAGCGGCTCCTGAAGGTCCCTTATTGTGGCTGATATTA
CTGTCAATGGTTATGGGTCCTATAAAAAATGCCCTCCAGATAAAGCTTCAGCGTTGGCTGAATTTTGGGTTTGGAGGCTG
GAAGATGGCAGGAGGCTGTAAACCTAGAGGTCAATTTAGAGGG
    
```

**Y (heterozygous T/C) for NM\_006563.3:c.304T>C, NP\_006554.1:p.Ser102Pro (rs2072597)**  
**R (heterozygous G/A) for NM\_006563.3:c.973G>A, NP\_006554.1:p.Glu325Lys**

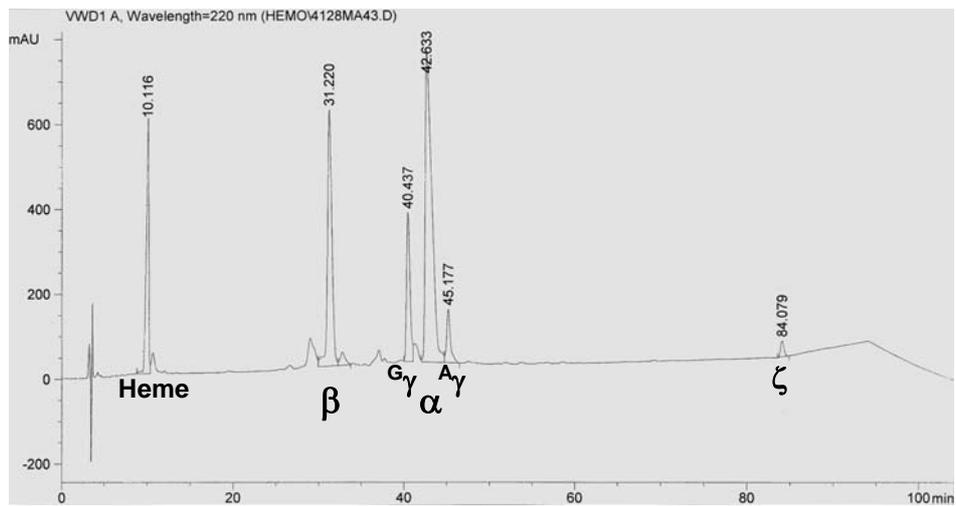
**Figure S6. (B) Analysis of heterozygous KLF1 mutation c.973G>A (E325K) from the CDA patient ME in both directions with Mutation Surveyor software (SoftGenetics).**



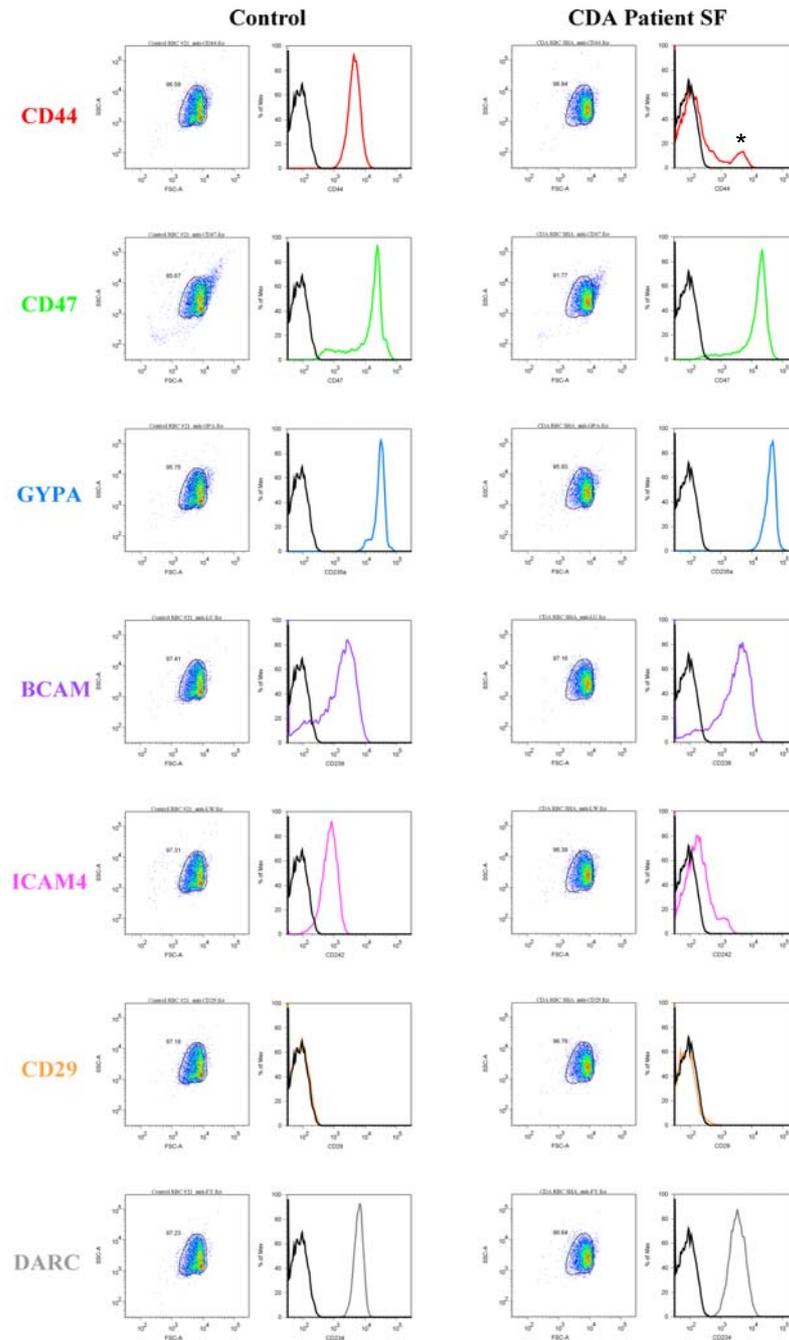
**Figure S7. Analysis of the erythrocyte membrane proteins from the CDA patient ME** (CDA, sample taken on 2010/01/05) or a control (CTL) showing abnormal migration of Band 3 (arrows) as usually observed in CDA II but also slight reduction of spectrins (stars). Proteins were resolved either by Laemmli's (left gel) or Fairbanks' (right gel) electrophoretic method, and stained with Coomassie Blue.



**Figure S8. Reverse phase liquid chromatography (RPLC) elution pattern of the hemolysate from the CDA patient ME** showing the presence of the hydrophobic  $\zeta$ -globin chain (retention time 84 min). This confirmed that the fast migrating hemoglobin observed in isoelectric focusing (IEF) was indeed embryonic hemoglobin Portland ( $\zeta_2\gamma_2$  tetramer). The peaks and the corresponding retention times of heme,  $\beta$ -,  $G\gamma$ -,  $A\gamma$ - and  $\alpha$ -globin chains are indicated.



**Figure S9. Flow cytometry analysis of cell markers CD44, CD47, GYPA, BCAM/Lu, ICAM4/LW, CD29 and DARC on erythrocytes of the CDA patient SF (right histograms, sample taken on 2009/06/29) or a healthy control (left histograms).** Erythrocytes were stained with mouse monoclonal antibodies (profiles in color) or an isotype control antibody (profile in black) and phycoerythrin-conjugated goat F(ab')<sub>2</sub> anti-mouse IgG(H+L), and were gated on FSC and SSC (density dot plots). Patient SF's erythrocytes abnormally showed no expression of CD44 and reduced expression of ICAM4; the small CD44<sup>+</sup> population (9.5%, star) likely corresponded to the residual erythrocytes of the most recent transfusion (2009/04/20) and should be taken into account for the analysis of the other markers, especially ICAM4.



**Table S1. Essential hematologic parameters of CDA patient ME from birth, before and after splenectomy.**

Red blood cell count (RBC), hemoglobin (HGB), hematocrit (HCT), mean cellular volume (MCV), mean cellular hemoglobin (MCH), mean cellular hemoglobin concentration (MCHC), red-blood-cell distribution width (RDW), platelet count (PLT), mean platelet volume (MPV), white blood cell count (WBC). Macrocytosis, severe thrombocytosis and extremely high erythroblastosis appeared after splenectomy. The three last RBC transfusions were on January 2004 (PT-Hb 7.2 g/dL), February 2005 (PT-Hb 7.2 g/dL) and October 2006 (PT-Hb 7.3 g/dL).

Parameters	Date Age (Hospital) Units	12/12/1996 BIRTH (28 weeks' gestation)												06/09/2000 SPLENECTOMY (4 years)				
		12/12/1996 1 day (Strasbourg)	14/02/1997 2 months (Strasbourg)	14/03/1997 3 months (Strasbourg)	05/05/1997 5 months (Nancy)	16/09/1997 9 months (Nancy)	29/01/1998 1 year (Nancy)	20/07/1998 1 year (Nancy)	20/05/1999 2 years (Nancy)	17/08/1999 2 years (Nancy)	29/03/2000 3 years (Nancy)	03/07/2000 3 years (Nancy)	20/04/2001 4 years (Paris)	14/05/2002 5 years (Paris)	25/10/2005 8 years (Nancy)	28/10/2008 11 years (Nancy)	10/02/2009 12 years (Nancy)	05/01/2010 13 years (Nancy)
RBC	10 <sup>12</sup> /L	4.0	2.7	3.8	3.3	2.2	2.7	2.7	2.4	2.7	1.7	1.8	2.5	2.5	2.3	2.7	2.9	2.6
HGB	g/dL	11.5	7.5	11.4	9.4	6.5	7.4	7.6	6.7	7.6	4.7	5.4	8.0	8.4	7.5	8.4	9.1	8.0
HCT	%	34.1	22.5	33.9	27.9	19.4	21.9	22.8	20.0	23.5	14.2	16.7	26.0	27.0	23.5	27.6	30.1	26.7
MCV	fL	86.3	85.0	89.0	83.6	87.7	82.3	84.9	83.8	87.4	85.3	85.9	106.9	108.9	101.0	102.4	102.9	103.4
MCH	pg	29.1	28.1	29.9	2.8	29.4	27.6	28.2	28.0	28.2	28.3	27.8	32.2	33.8	32.2	31.1	31.1	31.9
MCHC	g/dL	33.7	33.0	33.6	33.7	33.5	33.8	33.3	33.5	32.3	33.1	32.3	30.1	31.0	31.9	30.4	30.2	30.0
RDW	%	18.4	16.7	14.6	14.1	14.6	17.5	13.9	20.9	18.2	20.6	24.0						
Reticulocytes	10 <sup>9</sup> /L	52.5	13.9	62.9	9.0	80.7	134.0	21.5					838	362	422	351	469	304
PLT	10 <sup>9</sup> /L	406.0	393.0	229.0	380.0	447.0	339.0	223.0	230.0	211.0	173.0	83.0	526.0	612.0	1020.0	1101.0	1036.0	1264.0
MPV	fL	10.4	12.0	10.3	8.9	7.9	8.4	8.6	8.1	8.5	7.8	7.3	10.6	9.3	7.5	9.2	8.1	8.3
WBC	10 <sup>9</sup> /L	9.8	20.4	8.2	8.1	13.0	24.2	11.9	15.9	16.0	19.2	27.8	22.7	24.9	15.9	13.9	16.3	16.4
Neutrophils	10 <sup>9</sup> /L	3.2	13.5	2.1	1.5	3.7	5.6	3.3	5.6	3.5	11.9	18.6	7.1	7.7	7.8	5.5	9.8	9.2
Eosinophils	10 <sup>9</sup> /L	0.4	0.6	0.2	0.3	0.2	0.2	0.1	0.3	0.1	0.2	0.2	1.4	0.8	0.3	0.3	0.3	0.2
Basophils	10 <sup>9</sup> /L	0.0	0.0	0.0	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0
Lymphocytes	10 <sup>9</sup> /L	3.6	3.3	4.8	4.8	7.4	16.5	6.3	8.4	9.8	6.3	7.2	10.9	14.7	6.4	6.1	4.7	4.8
Monocytes	10 <sup>9</sup> /L	1.1	2.0	1.2	1.2	1.0	1.9	2.1	1.6	2.6	0.8	1.7	3.4	1.5	1.1	1.8	1.3	1.8
Neutrophils	%	33.0	66.0	25.0	18.5	28.4	23.0	28.0	35.0	22.0	62.0	67.0	31.0	34.0	34.3	24.4	42.9	40.5
Eosinophils	%	4.0	3.0	2.0	3.8	1.7	1.0	1.0	2.0	0.4	0.8	0.6	6.0	3.3	1.4	1.2	1.5	0.7
Basophils	%	0.0	0.0	0.0	0.7	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.7	0.0
Lymphocytes	%	37.0	16.0	58.0	59.1	56.9	68.0	53.0	53.0	61.0	33.0	26.0	48.0	64.7	28.0	26.8	20.7	20.9
Monocytes	%	11.0	10.0	14.0	14.4	7.7	8.0	18.0	10.0	16.0	4.0	6.0	15.0	6.6	4.8	7.9	5.7	8.0
Erythroblasts	/100 WBCs	470	12	128	few	12	30	few	100	110	80	210	920	594	1000	1000	500	500

**Table S2. PCR and sequencing primers used to sequence *KLF1*.**

Location and direction of the primers are based on NCBI *KLF1* mRNA Reference Sequence NM\_006563.3 while position is based on NCBI human chromosome 19 genomic contig NT\_011295.11. Complete *KLF1* was sequenced after PCR amplification of two overlapping fragments using high fidelity DNA polymerase as follows. A 5' *KLF1* fragment was amplified by semi-nested PCR with primers KLF1-1, KLF1-2 and KLF1-11, and sequenced with primers KLF1-2, KLF1-6 and KLF1-15. A 3' *KLF1* fragment was amplified by semi-nested PCR with primers KLF1-7, KLF1-3 and KLF1-4, and sequenced with primers KLF1-7, KLF1-8 and KLF1-4. PCR products were sequenced with ABI BigDye terminator chemistry (GATC Biotech) after gel purification (Macherey-Nagel). Sequencing data were analyzed with DNA Workbench software (CLC bio) and Mutation Surveyor software (SoftGenetics).

Name	Sequence	Location	Direction	Position
KLF1-1	TAGTCTTTAACCCCAGCCCCAG	5' side of exon 1	Sense	4260969-4260948
KLF1-2	ACGTGAAGTTTGTGCCCCAG	5' side of exon 1	Sense	4260939-4260920
KLF1-3	AGGGGGCTTGTGGAGTTGAG	3' side of exon 3	Antisense	4257897-4257917
KLF1-4	AGGAGATGAGGGTGTGTAAGG	3' side of exon 3	Antisense	4257935-4257955
KLF1-6	TGATCGGTTTCTGTCCCTGG	Intron 1	Sense	4259958-4259939
KLF1-7	AGGATCACTCGGGTTGGGT	Exon 2	Sense	4259457-4259439
KLF1-8	TACACCAAGAGCTCCCACC	Exon 2	Sense	4258978-4258960
KLF1-11	CGGGTCCCAAACAACCTCAG	Exon 2	Antisense	4259122-4259140
KLF1-15	GCTTTGGAAAGGGTCTTG	Intron 1	Antisense	4259847-4259865