



Ferrata Storti Foundation

Genetic disarray follows mutant KLF1-E325K expression in a congenital dyserythropoietic anemia patient

Lilian Varricchio,¹ Antanas Planutis,² Deepa Manwani,³ Julie Jaffray,⁴ W. Beau Mitchell,⁵ Anna Rita Migliaccio^{4,6,*} and James J. Bieker^{1,2,7,8,*}

¹Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA;

²Department of Cell, Developmental, and Regenerative Biology, Icahn School of

Medicine at Mount Sinai, New York, NY, USA; ³Division of Hematology/Oncology, The Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, NY, USA;

⁴Children's Hospital Los Angeles, University of Southern California Keck School of

Medicine, Los Angeles, CA, USA; ⁵Department of Pediatrics, Icahn School of Medicine at

Mount Sinai, New York, NY, USA; ⁶Dipartimento di Scienze Biomediche e NeuroMotorie,

Alma Mater Studiorum, Università di Bologna, Bologna, Italy; ⁷Black Family Stem Cell

Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA and ⁸Mindich

Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New

York, NY, USA

*ARM and JJB contributed equally as senior authors on this work.

Haematologica 2019
Volume 104(12):2372-2381

ABSTRACT

Congenital dyserythropoietic anemia type IV is caused by a heterozygous mutation, Glu325Lys (E325K), in the KLF1 transcription factor. Molecular characteristics of this disease have not been clarified, partly due to its rarity. We expanded erythroid cells from a patient's peripheral blood and analyzed its global expression pattern. We find that a large number of erythroid pathways are disrupted, particularly those related to membrane transport, globin regulation, and iron utilization. The altered genetics lead to significant deficits in differentiation. Glu325 is within the KLF1 zinc finger domain at an amino acid critical for site specific DNA binding. The change to Lys is predicted to significantly alter the target site recognition sequence, both by subverting normal recognition and by enabling interaction with novel sites. Consistent with this, we find high level ectopic expression of genes not normally present in the red cell. These altered properties explain patients' clinical and phenotypic features, and elucidate the dominant character of the mutation.

Correspondence:

JAMES J. BIEKER
james.bieker@mssm.edu

Received: October 22, 2018.

Accepted: March 12, 2019.

Pre-published: March 14, 2019.

doi:10.3324/haematol.2018.209858

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/104/12/2372

©2019 Ferrata Storti Foundation

Material published in *Haematologica* is covered by copyright. All rights are reserved to the Ferrata Storti Foundation. Use of published material is allowed under the following terms and conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>.

Copies of published material are allowed for personal or internal use. Sharing published material for non-commercial purposes is subject to the following conditions:

<https://creativecommons.org/licenses/by-nc/4.0/legalcode>,

sect. 3. Reproducing and sharing published material for commercial purposes is not allowed without permission in writing from the publisher.



Introduction

Congenital dyserythropoietic anemias (CDA) encompass a set of rare, chronic anemias that are heterogeneous but commonly exhibit morphologically abnormal bone marrow erythropoiesis. There are four subtypes that are differentiated by their causative genetic mutations: type I in *CDA1*, type II in *SEC23B*, and type III in *KIF23*.¹⁻³ The genetic cause for the rare CDA type IV (OMIM 613673) is unique in that it derives from a transcription factor mutation. A G-to-A transition in one allele of exon 3 of *KLF1* (erythroid Krüppel-like factor; EKLF) results in the substitution of a glutamate 325 by a lysine (E325K). CDA type IV patients present with severe hemolytic anemia, splenomegaly, elevated fetal hemoglobin (HbF), iron overload, red cell osmotic fragility, and dyserythropoiesis in the bone marrow.⁴⁻¹³ These clinical properties do not fit into the typical categories for hereditary persistence of fetal hemoglobin, β -thalassemia, or hereditary spherocytosis.^{5,12,13} The high levels of nucleated and bi-nucleated erythroid cells in the bone marrow and in the peripheral blood are quite striking.

The E325K mutation, within the second zinc finger of the KLF1 protein, is at a universally conserved residue/amino acid that is critical for proper DNA target site recognition. KLF1¹⁴ is a zinc finger hematopoietic transcription factor that plays a global role in activation of genes critical for genetic control within the erythroid lineage.¹⁵⁻¹⁸ It performs this essential function by binding to its cognate DNA 5'CCM-CRCCCN3' element, interacting with basal transcription factors, and recruiting

chromatin remodeling proteins and histone modifiers. This results in accurate and precise local and global chromatin organization. Chromatin immunoprecipitation sequencing (ChIP-seq) analyses together with RNA expression profiling underscore the importance of KLF1 in red cell biology, demonstrating its multifunctional roles in tissue-specific gene expression, lineage determination, and terminal maturation.

Over 70 human mutations in KLF1 have been identified over recent years.^{19,20} Monoallelic mutation of *KLF1* in humans usually leads to a benign outcome that is nonetheless phenotypically important.²⁰⁻²⁴ Expression of certain cell surface markers, such as CD44 and Lutheran antigen, are affected but do not exert any physiological effect.²⁵ Expansion of these cells in culture is also not affected.^{10,25} Of clinical importance, however, haploinsufficient levels of KLF1 lead to altered β -globin switching and elevated γ -globin expression, which can be advantageous, particularly in areas with endemic β -thalassaemia.²⁶ Related to this, some single and compound mutations in human KLF1 lead to anemias in addition to CDA such as spherocytosis, microcytic hypochromic anemia, pyruvate kinase deficiency,^{10,27,28} or in the most extreme case, *hydrops fetalis*.²⁹ Expression of about 700 genes is dependent on KLF1 in humans. As a result of its central importance, one might predict an extensive cascade of changes would follow from the KLF1-E325K mutation, particularly likely in this case given the broad role of KLF1 in the control of erythropoiesis.

Two sets of observations in the mouse are particularly informative for the present study. One is from studies of the monoallelic mouse neonatal anemia (*Nan*) mutation that resides at the same amino acid of KLF1, albeit with an aspartate substitution (E339D).^{30,31} These mice exhibit a lifelong anemia due to a distorted erythroid transcriptional output. The E339D mutation not only yields a variant with a more circumscribed binding specificity compared to wild type (WT),³¹ but also one that recognizes a novel, more degenerate target sequence uniquely recognized by Nan-KLF1.^{32,33} The second observation is that, unlike the *Nan* mutant, mouse erythroid cells totally ablated in KLF1 do not enucleate, but instead stall at the orthochromatic erythroblast stage.³⁴ Many of these cells are also bi-nucleated. As a result, mutant expression or insufficient levels of KLF1 can separately contribute to defective erythroid expression and phenotypic properties.

Although the CDA type IV red cell cellular and phenotypic properties have been described, the molecular mechanism/details by which the KLF1-E325K mutation exerts its effect and causes these significant changes has not been previously addressed. A limitation of studying this disease has been the paucity of starting material due to its rarity. As a result, we directed our efforts towards analysis of derived erythroid cells from the peripheral blood of our published patient.

Methods

Cell sources

Analysis used RNA from patient peripheral blood cells leftover from our previously published study⁵ that had received Institutional Review Board approval; no new patient material was obtained for the present study. Mononuclear cells from the peripheral blood (PBMC) had been isolated and cryopreserved as

described.³⁵ Non-patient PBMC were purchased from AllCells (PB003F).

Human erythroid massive amplification protocol

Peripheral blood mononuclear cells underwent a two-step culture;^{36,37} one for proliferation, with harvests at day (d)11 and d15; the second for differentiation that was harvested after the d11 culture was differentiated for an additional five days. Under these conditions, the normal sample attained CD235a⁺ levels of over 70% (*data not shown*). As a result, we used these prescreened reagents and conditions for the proliferation/differentiation experiment of normal cells in parallel with the patient PBMC sample.

Peripheral blood mononuclear cells (10E6 cells/ml) were cultured in IMDM plus 20% FBS, SCF (100 ng/mL), IL-3 (1 ng/mL), EPO (5 U/mL), dexamethasone and estradiol (both 1 μ M). Also included were deionized human serum albumin (5%), human iron saturated transferrin, liposomes plus cholesterol (400 μ g/mL), and lecithin (1.2 mg/mL).^{36,37} Differentiation was enabled by increasing erythropoietin (EPO) to 10 U/mL, removing dexamethasone, and including recombinant human insulin (40 ng/mL) and T3 (1 μ M). Both cell sources were successfully cultured and expanded in this way, giving us confidence that the human erythroid massive amplification (HEMA) approach enables direct analysis of CDA patient erythroid samples.

Morphological analysis of d11 expanding cells was derived from analysis of three sets of cytopins from two experiments, each analyzed and quantified independently by two of the authors.

RNA isolation and analysis

Total RNA from all samples was isolated with Trizol (Sigma). Bioanalyzer (Agilent) analysis showed that RIN values were all between 7.8-8.7 except for the normal differentiated d5 sample, which had a value of 3.4. These samples were used for polyA⁺ library preparation using the Bioo Scientific (NEXTflex) Rapid Directional kit (NOVA-5138-07). Next generation sequencing was performed on an Illumina NextSeq 500. Sequencing yielded 75 nt single end reads, >30 million per sample.

RNA-seq data have been submitted to the Gene Expression Omnibus GSE128718.

Real-time qualitative polymerase chain reaction (RT-qPCR) was performed on cDNA generated with a mix of oligo-dT and random hexamers (Quantabio 95048-025). Primers were as previously described.³⁸ Errors after combining quantities with their own uncertainties were calculated as in <http://lectureonline.cl.msu.edu/~mmp/labs/error/e2.htm>.

Bioinformatics

Expression data from human *in vitro* expanded primary erythroid cells analyzed at 5 stages (Pimentel, 2014 #2477) was obtained from GEO (GSE53635). Sequenced reads were mapped to the human genome (hg38) using Tophat2. Accepted hits were tested for differential expression analysis using Cuffdiff2 with the blind dispersion method. Heatmap plots for gene clusters were created in R package pHeatmap.

Gene set enrichment analysis (GSEA) (<http://software.broadinstitute.org/gsea/index.jsp>) was performed {Subramanian, 2005 #2464; Mootha, 2003 #2465} using gene set lists from selected expression clusters (described by Li *et al.*).³⁹ Venn diagrams were generated with Venny (<http://bioinfogp.cnb.csic.es/tools/venny/> or <https://www.stefanijol.nl/venny>) or BioVenn (<http://www.biovenn.nl/index.php>) and were used to identify non-overlapping genes within sets. The DAVID analysis tool v6.8 (<https://david.ncifcrf.gov/>) was used as described.⁴⁰

Results

Establishment of expansion protocol

Our patient had been analyzed with respect to hematologic parameters such as red cell surface expression, peripheral blood smear, and globin expression pattern.⁵ For the present study, erythroid cells were expanded from PBMC using an *ex vivo* culture system. This culture system contains stimulatory cytokines along with the critical inclusion of dexamethasone to enable efficient expansion of the small number of erythroid progenitors present in a typical mononuclear cell preparation.^{36,37} Using this protocol we established and expanded erythroid cells from the CDA type IV patient in parallel with a normal control. Morphological examination (Figure 1) reveals that the expanded patient cells exhibit bi- and multi-nucleated cells with abnormal nuclei (approx. 40%) as seen in the original bone marrow and blood smears of the patient.⁵ These are observed in approximately 5% of cells from the normal control.

Similar to the limitation in studying erythropoiesis in murine KLF1-null cells,³⁴ expression of many of the cell surface markers for differentiation are quite low and not informative for staging purposes, as they are KLF1 targets. As a result, we assessed and compared the range of cellular morphologies in the d11 proliferating samples, and find these are not significantly different (Figure 1C).

Global dysregulation of erythroid genes in the CDA cell

Given the availability of cells from only a single patient, we aimed to analyze cells at three time points to increase

the robustness of our data. As a result, the PBMC underwent a 2-step culture: one for proliferation, with harvests at d11 and d15; the second for differentiation started at d11 and harvested after an additional five days. RNA was isolated from all samples and analyzed for gene expression *via* deep RNA sequencing. Focusing our analysis on selected targets, we find a radical alteration of gene expression that covers cell cycle, membrane protein, and globin switching deficits, all congruent with the phenotypic properties of CDA type IV cells (Figure 2). For example, although expression of some cell surface molecules (*CD47*, *CD55*, *CD58*) are minimally affected as noted before by FACS analysis of primary patient cells⁴ (and thus serve as controls), *GYPA* (*CD235a*) expression is dramatically lower. This is mimicked by radically low levels of KLF1-regulated structural proteins such as *ICAM4*, tropomodulin (*TMOD1*), and band 4.2 (*EPB42*), likely contributing to the fragility of patients' red cells and their membrane abnormalities. Transport proteins such as anion transporter band 3 (*SLC4A1*), ion channel *PIEZO1*, potassium-calcium channel *KCNN4*, ABC transporter *ABCB6*, and aquaporin water channel *AQP3* exhibit extremely low levels of expression (Figure 2A).

Cells from the CDA patient express high levels of γ -globin, reaching approximately 90% of total globin (*Online Supplementary Figure S1A*). With respect to regulation of globin switching, two KLF1 targets that act as γ -globin repressors were checked.⁴¹⁻⁴³ *BCL11A* levels are decreased but not that of *ZRF/Pokemon* (*ZBTB7A*) (Figure 2B), suggesting that the increase in γ -globin follows derepression as a result of the drop in *BCL11A* protein. Contributing to

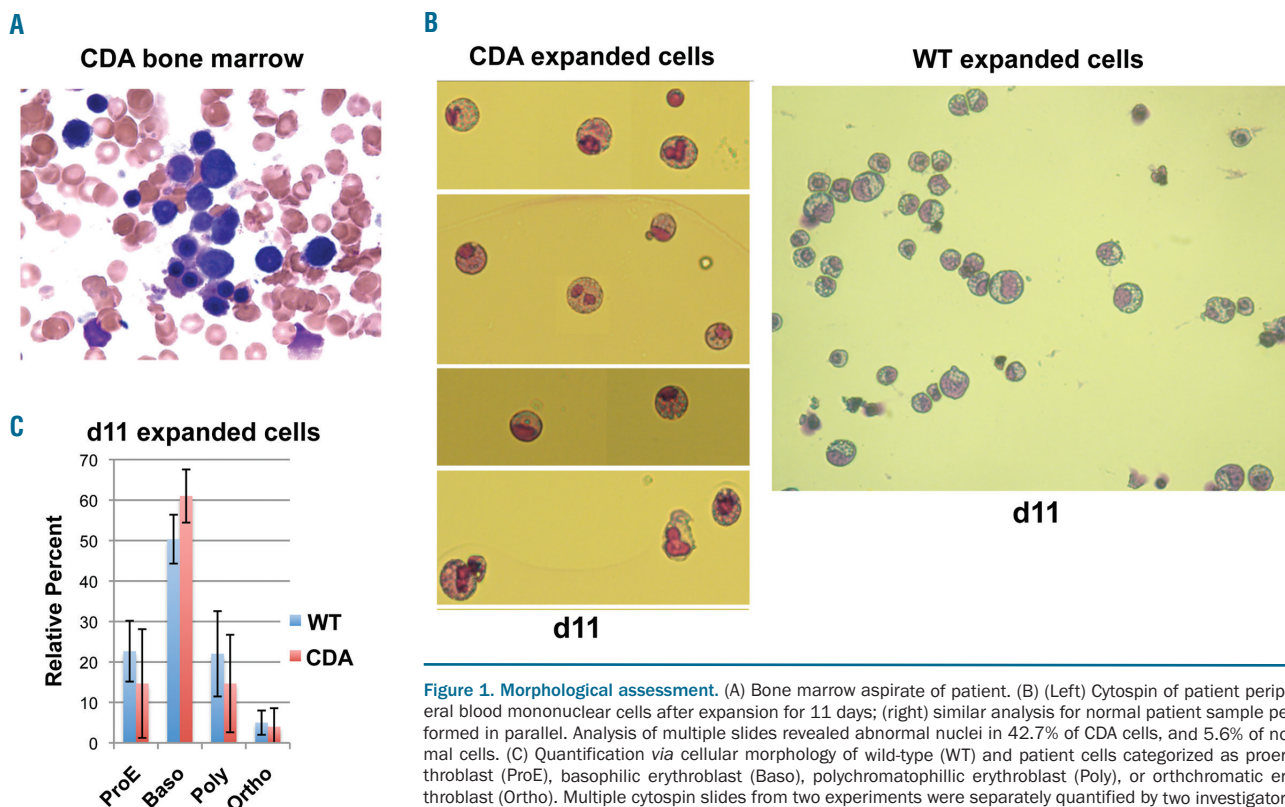


Figure 1. Morphological assessment. (A) Bone marrow aspirate of patient. (B) (Left) Cytospin of patient peripheral blood mononuclear cells after expansion for 11 days; (right) similar analysis for normal patient sample performed in parallel. Analysis of multiple slides revealed abnormal nuclei in 42.7% of CDA cells, and 5.6% of normal cells. (C) Quantification *via* cellular morphology of wild-type (WT) and patient cells categorized as proerythroblast (ProE), basophilic erythroblast (Baso), polychromatophilic erythroblast (Poly), or orthochromatic erythroblast (Ortho). Multiple cytopsin slides from two experiments were separately quantified by two investigators.

this phenotypic effect, even though both alleles are expressed (Online Supplementary Figure S1B), total *KLF1* expression is lower in the CDA erythroid cell (Figure 2B) verified by direct RT-qPCR (*data not shown*).

Cell cycle effects are less straightforward to interpret. Both cell cycle stimulators as well as inhibitors play important roles at different stages of erythroid terminal differentiation.¹⁸ In the present case (Figure 2C), *E2F2* and *E2F4* levels are overall lower in the patient cells, while p21 (*CDKN1A*) is higher. Surprisingly in this context, *TP53* levels are lower in the patient cells, suggesting that the increase in p21 occurs independently of p53.⁴⁴ Expression of pro- or anti-apoptosis protein-coding genes are minimally changed, with the exception of *BCL2* and possibly *PIM1*, whose levels are higher in the CDA cell.

Given the phenotypic problems with nuclear extrusion, bi-nuclei, and bridges, it is of interest that a number of cytokinesis/mitosis proteins implicated in these cellular processes such as pleckstrin (*PLEK2*) and *TRIM58* are lower in expression (Figure 2D). In a link to CDA type III patients that express mutant *KIF23*, levels of this gene are decreased (as also noted in the *KLF1* hydrops patient).²⁹ Other CDA-associated genes such as *CDAN1* (type I) and *SEC23B* (type II) are not significantly altered.

Given the extent of disruption of iron utilization in the CDA type IV patients, we also queried this subset of genes as well, and found varied effects (Figure 2E). For example, erythroferrone (*FAM132B*), a regulator of hepcidin, is dramatically down-regulated in the CDA patient (and thus different from CDA type II⁴⁵), as are the transferrin receptors (*TFR2* and *TFRC*). Although levels of the iron exporter ferroportin (*SCL40A1*) and the heme regulator

SLC48A1 are not significantly altered, this combination likely leads to a net inability to mobilize and incorporate iron and heme in the CDA patient, particularly when coupled to the low expression of the *ABCB6* gene (noted earlier) that would decrease activity of this important porphyrin importer. These conditions could account for the anisopoikilocytosis seen in spite of normal MCV. Ferritin levels (*FTL* and *FTH1*) are increased, which may contribute to the high levels of stored iron observed in the patient. By the same token, increased heme oxygenase (*HMOX1*) could account for the hyperbilirubinemia seen in all patients with type IV CDA. Alternatively, increased bilirubin may be secondary to hemolysis, with *HMOX1* expression increased in response. These effects are reminiscent of those resulting from the mouse *Nan* mutation, which contains a change in the same amino acid (albeit to D) and leads to extensive iron and heme regulation disruption.³³

Differentiation deficits

To identify expression changes in the CDA cell during the process of proliferation and differentiation, we compared genes expressed ≥ 5 FPKM (Fragments Per Kilobase of transcript per Million) in all samples. Individual Venn comparisons reveal that the d11 and d15 proliferating samples each overlap $>72\%$, but the differentiation d5 set does not continue this pattern, only overlapping approximately 55% (Figure 3A). This suggests that terminal differentiation does not proceed normally in the CDA erythroid cell. To more directly address this, we analyzed RNA expression of our samples and compared it to that from cells undergoing normal human erythropoiesis.^{39,46,47} At

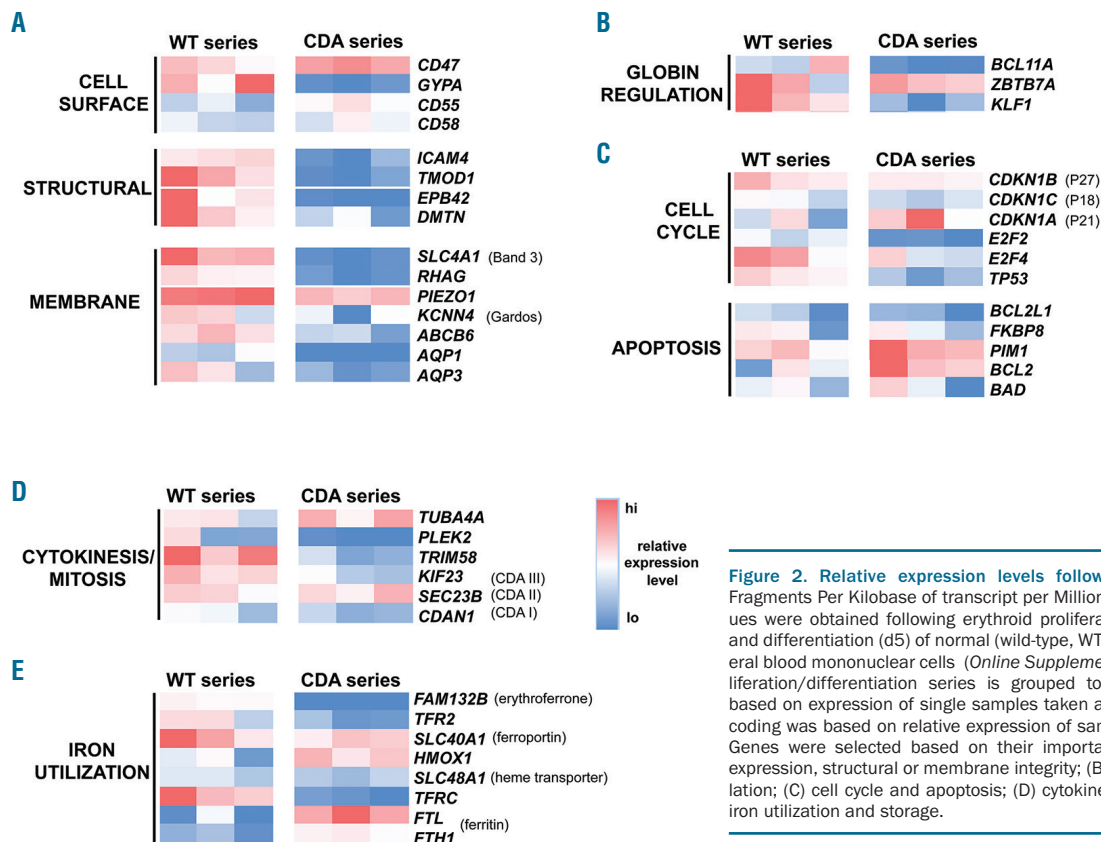


Figure 2. Relative expression levels following RNA-seq analysis. Fragments Per Kilobase of transcript per Million (FPKM) expression values were obtained following erythroid proliferation [day d11 and d15] and differentiation (d5) of normal (wild-type, WT) or patient (CDA) peripheral blood mononuclear cells (Online Supplementary Table S1). The proliferation/differentiation series is grouped together and color-coded based on expression of single samples taken at each time point. Color-coding was based on relative expression of samples within each group. Genes were selected based on their importance in: (A) cell surface expression, structural or membrane integrity; (B) β -like globin gene regulation; (C) cell cycle and apoptosis; (D) cytokinesis and mitosis; and (E) iron utilization and storage.

stages of differentiation from CD34⁺ to the proerythroblast stage ('cluster 16'),³⁹ gene enrichment analysis of d11 proliferating WT cells unsurprisingly shows a strong overlap with genes that are highest in the proerythroblast cell (Figure 3B). However, this is not the case with the CDA cells, which show a negative correlation (Figure 3B). In fact, the CDA cells show a positive correlation with genes

that are highest at the CD34⁺ stage ('cluster 1')³⁹ (Figure 3B). Once CDA cells are differentiated at d5, they show a slight increase in overlap with genes expressed late (bracketed region in Figure 3B).

As an aid to visualize subsequent erythroid maturation changes in expression, a heat map analysis was performed with genes increasing in expression between the proery-

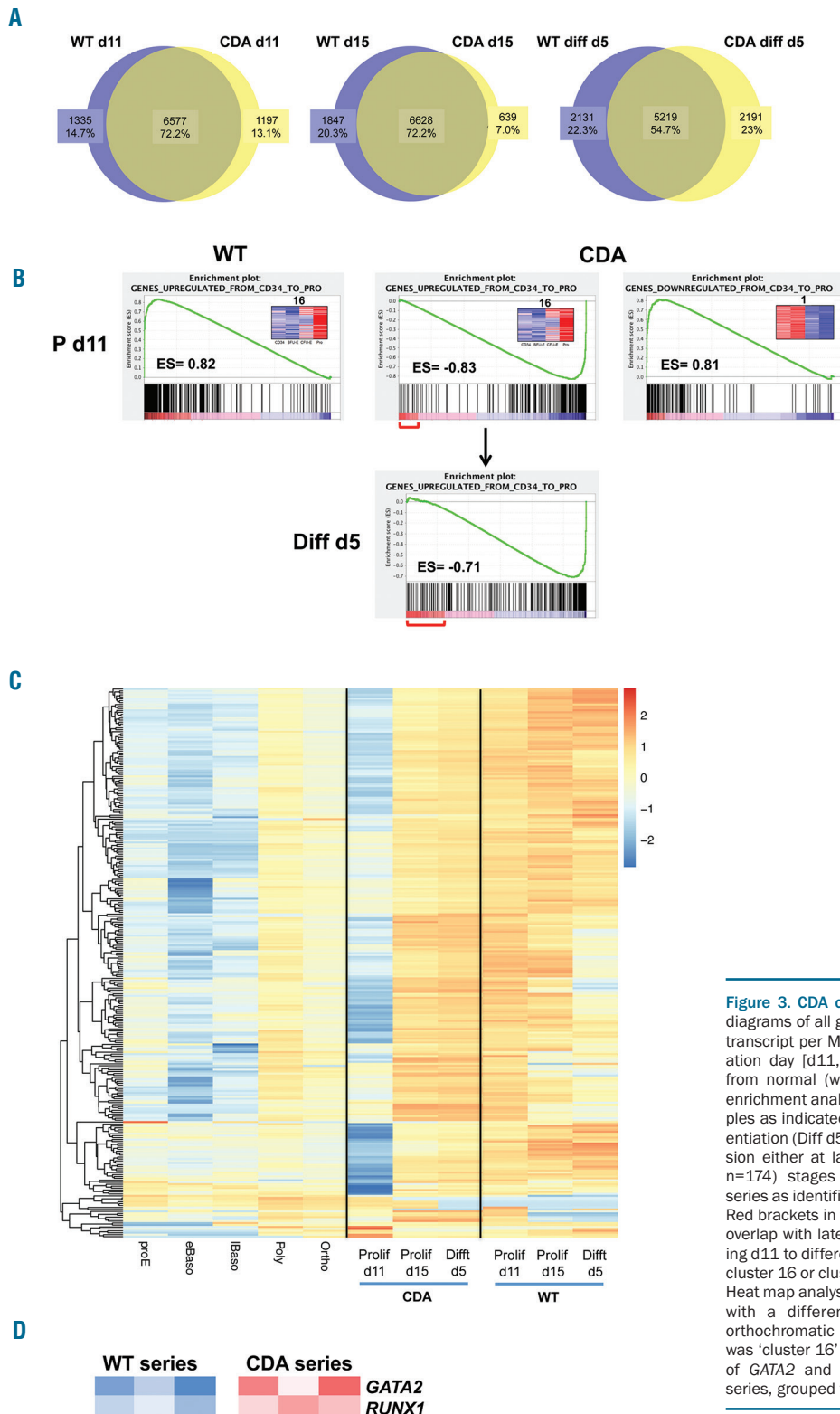


Figure 3. CDA cells are defective in differentiation. (A) Venn diagrams of all genes expressed ≥ 5 Fragments Per Kilobase of transcript per Million (FPKM) and compared between proliferation day [d11, d15, or differentiation d5 samples, derived from normal (wild-type, WT) or patient (CDA)]. (B) Gene set enrichment analysis (GSEA) analysis of WT or CDA patient samples as indicated from proliferating d11 (P d11) or after differentiation (Diff d5) were compared to genes enriched for expression either at late ('cluster 16', $n=268$) or early ('cluster 1', $n=174$) stages of the CD34⁺/BFU-E/CFU-E/proerythroblast series as identified by Li *et al.*³⁹ (shown as insets in each graph). Red brackets in the CDA samples show the increase in positive overlap with late expressing genes when comparing proliferating d11 to differentiation d5 samples. Inserts are heat maps of cluster 16 or cluster 1 from Li *et al.*³⁹ ES: enrichment score. (C) Heat map analysis of WT or CDA patient samples in conjunction with a differentiation series (proerythroblast through to orthochromatic erythroblast as indicated).⁴⁶ The starting set was 'cluster 16' from (B). (D) Relative FPKM expression values of *GATA2* and *RUNX1* from the proliferation/differentiation series, grouped together and color-coded as in Figure 2.

throblast and orthochromatic stages.^{46,47} Starting from the dataset of genes that are highest in the proerythroblast ('cluster 16' set in Figure 3B), we see that many of these genes continue to increase in expression during the normal erythroid lineage differentiation process (Figure 3C). Focusing first on the WT series, we see that the d11 set is most closely aligned with the polychromatic erythroblast expression pattern (Figure 3C). Many of these genes then decrease in expression upon further proliferation (d15), but most obviously after differentiation is established. The pattern for the CDA series is different in two significant ways (Figure 3C). 1) The expression of this gene set at d11 is quite low, consistent with its GSEA expression analysis. 2) Although the d15 set is converging on the polychromatic/orthochromatic erythroblast pattern, it remains minimally changed even after differentiation.

These data suggest that the CDA erythroid cell retains a significant residual early proliferation expression pattern, consistent with the erythroid hyperplasia seen in the CDA type IV patients' bone marrow.^{4,5,12,13} Further supporting this, expression of early hematopoietic transcriptional markers such as *GATA2* and *RUNX1* remain high in the CDA samples (Figure 3D). We conclude that terminal differentiation is aberrant and does not proceed properly in the CDA erythroid cell; importantly, this follows from a combination of both the presence of KLF1-E325K and a hypomorphic level of total KLF1 RNA expression.

Ectopic expression of non-erythroid genes

The E325K change in KLF1-CDA is at a critical amino acid within the DNA recognition sequence. Based on structural arguments^{10,14} as well as the precedent from the Nan-KLF1 mutation at the same site (E339D),^{31,32} it is likely that the change in KLF1-CDA confers recognition of atypical sites in the genome. This would lead to ectopic expression of genes that are normally not expressed in the erythroid cell. Substitution of a lysine for glutamate at amino acid (aa) 325 alters the middle residue of the critical "X,Y,Z" amino acids, which play determining roles in DNA target site recognition.^{48,49} Based on the most parsimonious model, one may predict the K325 residue would now recognize guanine on the G-rich strand and alter the recognition sequence to 3'GGKGGGGN5'. This would be a significant change, as a pyrimidine (T or C) is normally present at the underlined site.

To identify these potential ectopic targets, we overlapped data sets from a Venn analysis of all expressed genes (≥ 5 FPKM) that are exclusive to CDA (i.e. not expressed in WT) in proliferating d11, d15, and differentiating d5 samples. This yields a unique set of 184 genes (Figure 4A). Many of these are membrane proteins that are normally enriched in lymphoid, dendritic, myeloid, or monocyte cells. Perusal of the top 35 differentially expressed of these show that red cell character and identity have been altered.

One of the most far reaching results from analysis of the Nan-KLF1 mutant was that the neomorphic expression pattern³² led to systemic effects that altered the hematologic properties of the mouse, including changes in levels of specific proteins and cytokines in the serum and feedback inhibition of erythropoiesis.³³ In the present case, expression of the unique CDA genes is truly ectopic, as they do not overlap genes up-regulated in the KLF1-defective *hydrops* erythroid cell²⁹ (Figure 4B). The extent of the level of misexpression of specific, normally non-erythroid

targets in the CDA cell is shown in Figure 4C. *CCL13* is a chemokine implicated in inflammation, with potential respiratory issues. This connects it with *LTC4S*, which codes for leukotriene synthase, a gene normally expressed only in the lung whose product is also implicated in inflammation and respiration. *PDPN* codes for podoplasmin, also expressed in the lung but in addition implicated in aberrant platelet aggregation. *IL17RB* is a protein that binds to the IL17 receptor. These increases are not due to a global dysregulation, as expression of genes adjacent to those affected are not significantly changed (*data not shown*).

Of mechanistic interest, each of these genomic regions (*CCL13*, *LTC4S*, *PDPN*, *IL17RB*) contain multiple copies of the predicted novel recognition sequence that could potentially be recognized by KLF1-E325K (Figure 4D). In support of this idea, a recent study in the Siatecka lab, based on a binding site-selection strategy for CDA zinc fingers, demonstrates that each of these putative sites bind *in vitro* to CDA-KLF1, but not WT KLF1, as judged by gel shift assays (K Kulczynska *et al.*, 2019, submitted manuscript).

IL17RB is critical for expression of IL8 (*CXCL8*), a molecule that, if mis-expressed, could have systemic effects beyond the erythroid cell, particularly with respect to neutrophil activation and respiratory inflammation^{50,51} (analogous to the misexpression of IFN β in the Nan-KLF1 erythroid cell).³³ We find that IL8 RNA expression levels are quite high in the CDA samples, but not detectable in the WT (Figure 4C). Consistent with the model, the *CXCL8* genomic region does not contain potential ectopic binding sites for CDA-KLF1 (*data not shown*) and thus is likely indirectly activated by KLF1 through *IL17RB*.

We conclude that these and other ectopic targets are mis-expressed in the CDA erythroid cell by virtue of expression of KLF1-E325K and its action *via* its novel recognition site.

Discussion

Altered erythroid biology in the CDA patient cell

A mutation in KLF1, at E325K, is responsible for CDA type IV. Seven patients have been described so far, and they all share some similar phenotypic and clinical characteristics.⁴⁻¹³ However, changes of RNA expression within these patients' erythroid cells has not been previously described. This is of interest as type IV is unique among CDA in that it is a transcription factor, rather than a structural protein, that is mutated.^{1,3} We find a major alteration in the normal patterns of red cell gene expression, such that globin regulation, cell surface protein expression, membrane transport, cytokinesis, and iron utilization are all dramatically affected. These transcriptional effects go far in explaining the common patient phenotypes described in the literature.

However, a cautionary note must be acknowledged, stemming from the limitation that we could only analyze samples from one patient due to the rarity of the disease. The question is whether our documented changes could be due to inter-individual variation and/or whether we are comparing equivalent stages. This is a general question in the field that needs to be considered (e.g. as extensively discussed),^{52,54} particularly with respect to difficulties in finding the best way to compare patient/normal popula-

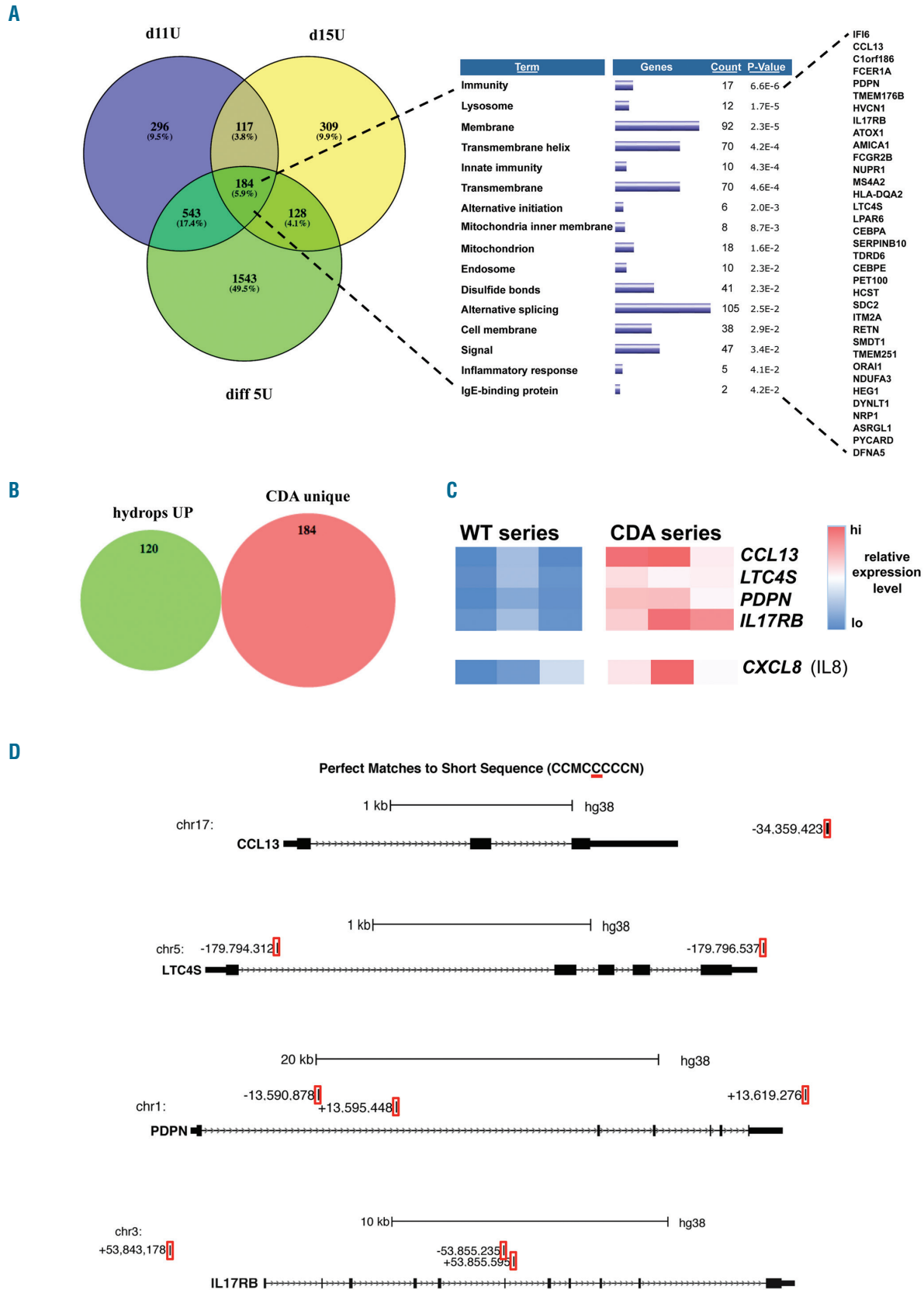


Figure 4. Ectopic expression in CDA samples. (A) Venn diagram showing overlap between genes uniquely expressed in CDA proliferation day (d)11 (d11U), d15 (d15U), and differentiation d5 (diff 5U) samples (*Online Supplementary Table S2*). DAVID analysis of terms enriched in the triple overlap between these samples (184 genes total) is shown, as are the top 35 genes most highly expressed in this enriched set. (B) Venn diagram of genes whose expression increases in the absence of KLF1²⁹ compared to the CDA-unique gene list from (A). (C) Relative Fragments Per Kilobase of transcript per Million (FPKM) expression values of four of the top CDA-unique genes of the proliferation/differentiation series grouped together and color-coded as in Figure 2. Also included are data for CXCL8 (IL8), which is downstream target of IL17RB. (D) Genome browser layouts of the CDA-unique genes showing the location of perfect nucleotide matches to the novel consensus sequence potentially bound by KLF1-E325K (5'CCMCCCCN3' on the C-rich strand).

tions of diseased cells, whether derived from individual patients or from a more extended grouping. To surmount this, we took care in our reliance on morphological assessment and global expression evaluation and comparisons, particularly given the limitations with respect to flow analysis of surface differentiation markers. Importantly, the observation that our patient's cells have difficulty in establishing differentiation in culture has been noticed in another patient.¹³ The only solution to this dilemma is to analyze additional patients, and/or to establish a ready source of cells (e.g. induced pluripotent stem cells from the patient) so that, at least, technical replicates can be more easily generated and analyzed.

As only *BCL11A*, but not *ZBTB7A*, levels were affected in the patient samples, it may be surprising that γ -globin accounted for up to 90% of total β -like globin in our analysis, given that the present patient's HbF levels were 42%.⁵ However, erythroid cells in culture may not exactly mimic the *in vivo* situation; for example, shRNA knock-down of *ZBTB7* in differentiating human erythroblasts led to near 90% γ -globin levels.⁵⁵

KLF1 ablation in the mouse leads to increased megakaryocyte colony-forming potential and gene expression (at the expense of erythropoiesis) as part of its role in regulating bipotential lineage decisions in the MEP.⁵⁶⁻⁶⁰ We do not find a similar increase in expression of megakaryocyte-restricted genes⁶¹ (Online Supplementary Figure S2). Although the CDA patient cells have higher levels of *FLI1* and *PECAM* expression, these likely follow from differentiation deficits rather than lineage divergence.

Patients who are compound heterozygous for KLF1 mutations present with non-spherocytic hemolytic anemia.²⁷ In the present case, the extensive disruption of structural and transport membrane proteins explain the membrane fragility and may also account for the low number of cells obtained during differentiation, when the acquisition of the erythroid-specific cell membrane structure is essential for survival.^{62,65} This is a suboptimal situation that not only explains the hemolysis but also the apparent lack of differentiation seen in the CDA cultures and in the patient, and could also explain the lowered red cell survival.^{6,8} In other words, rather than resulting solely from a differentiation block, the anemia and apparent hyperplasia may follow the physical survival and preferential enrichment of immature cells in the CDA patient.

Mechanistic implications of the E325K substitution

The dominant effect of KLF1-E325K expression follows from mutation of only one allele that is sufficient to produce the altered genetic and cellular properties of the CDA red cell. Our data suggest this follows from two different causes. First, recognition by KLF1-E325K of its normal cognate site is impaired;^{4,10} indeed, a quantitative reduction at all tested promoters by KLF1-E325 has been observed¹⁰ (K Kulczynska, 2019, submitted manuscript). KLF1 is known to interact with transcriptional regulators, histone modification proteins, and chromatin remodelers,¹⁷ and is critical for formation of the proper 3D chromatin complex and transcription factories at a number of erythroid target sites.^{64,65} Accordingly, altered downstream

consequences in the CDA cell may well follow from destabilized or incorrect protein complex formation that interferes with optimal WT activity.¹⁰

This effect is likely augmented/intensified by our second major observation that total KLF1 RNA levels are low. Hypomorphic levels of KLF1 are known to negatively affect expression of only a minority of selected targets, and these effects appear phenotypically benign.^{19,23-25} However, in the present scenario low expression has been compounded by co-expression of a mutant allele. The sum of these changes is a dramatically dysregulated erythroid cell with altered physical and expression parameters.

It is instructive to compare our data with that from the *Nan* mouse.^{31,66} Although the amino acid substitution is different, this mouse is anemic by virtue of intrinsic red cell parameters that are changed, many in a similar way to that of the CDA type IV patient. However, there are two fundamental differences. 1) The most obvious difference is that a substitution of lysine for glutamic acid (in our patient) is not expected to yield the same effect as substitution of aspartic acid (in the *Nan* mouse), as the location of this change is at a critical DNA recognition amino acid in the zinc finger structure. 2) The more subtle corollary of this is that the *Nan* mutation only affects a subset of its normal 5'CCMC**R**CCCN3' target sites (because *Nan*-KLF1 still recognizes 5'CCMC**G**CCCN3'); as a result, many KLF1 targets are not affected in the *Nan* erythroid cell.³¹ This is less the case in the present situation, as the lysine substitution in KLF1-E325K would not favor recognition of a sequence with "R" in the middle position.^{10,32,48,49}

However, one important concept derived from the *Nan* mouse is directly relevant to the CDA erythroid cell: that any amino acid change at this critical glutamic acid residue (to D or to K) leads to recognition of an abnormal target sequence, and thus ectopic expression of genes normally not present in the red cell.³² Such misexpression in the *Nan* mouse led to measurable and physiologically effective levels of secreted proteins in the serum that contributed to its splenomegaly and anemia.³³ We suggest a similar occurrence here that could contribute to some of the non-erythroid characteristics (short stature, gonadal dysgenesis) observed in many of the CDA type IV patients. Some of the most highly dysregulated genes described here predict that it might be useful to monitor respiratory and/or autoimmune/inflammatory issues in CDA patients.^{30,51,67} In this context, it will be of interest to compare RNA expression profiles of as many of the other patients as possible, particularly with respect to gender differences and identification of potential modifier loci that affect the other, non-shared phenotypes of the CDA type IV patients.¹³

Acknowledgments

This work was supported by National Institutes of Health grant R01 DK046865 to JJB, by the Myeloproliferative Neoplasms Research Consortium to ARM, and by R01 HL134684 to JJB and ARM. We thank Nithya Gnanapragasam, Kaustav Mukherjee, Li Xue, and Giovanni Miglicaccio for discussion throughout the study; Sunita D'Souza for PBMC purification; and Ravi Sachidanandam and Saboor Hekmaty for library preparation and deep sequencing.

References

- Iolascon A, Esposito MR, Russo R. Clinical aspects and pathogenesis of congenital dyserythropoietic anemias: from morphology to molecular approach. *Haematologica*. 2012;97(12):1786-1794.
- Iolascon A, Heimpel H, Wahlin A, Tamary H. Congenital dyserythropoietic anemias: molecular insights and diagnostic approach. *Blood*. 2013;122(13):2162-2166.
- Moreno-Carralero MI, Horta-Herrera S, Morado-Arias M, et al. Clinical and genetic features of congenital dyserythropoietic anemia (CDA). *Eur J Haematol*. 2018;101(3):368-378.
- Arnaud L, Saison C, Helias V, et al. A dominant mutation in the gene encoding the erythroid transcription factor KLF1 causes a congenital dyserythropoietic anemia. *Am J Hum Genet*. 2010;87(5):721-727.
- Jaffray JA, Mitchell WB, Gnanapragasam MN, et al. Erythroid transcription factor EKLF/KLF1 mutation causing congenital dyserythropoietic anemia type IV in a patient of Taiwanese origin: Review of all reported cases and development of a clinical diagnostic paradigm. *Blood Cells Mol Dis*. 2013;51(2):71-75.
- Agre P, Smith BL, Baumgarten R, et al. Human red cell Aquaporin CHIP. II. Expression during normal fetal development and in a novel form of congenital dyserythropoietic anemia. *J Clin Invest*. 1994;94(3):1050-1058.
- Parsons SF, Jones J, Anstee DJ, et al. 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*. 1994;83(3):860-868.
- Tang W, Cai SP, Eng B, et al. Expression of embryonic zeta-globin and epsilon-globin chains in a 10-year-old girl with congenital anemia. *Blood*. 1993;81(6):1636-1640.
- Wickramasinghe SN, Illum N, Wimberley PD. Congenital dyserythropoietic anaemia with novel intra-erythroblastic and intra-erythrocytic inclusions. *Br J Haematol*. 1991;79(2):322-330.
- Singleton BK, Lau W, Fairweather VS, et al. Mutations in the second zinc finger of human EKLF reduce promoter affinity but give rise to benign and disease phenotypes. *Blood*. 2011;118(11):3137-3145.
- de-la-Iglesia-Inigo S, Moreno-Carralero MI, Lemes-Castellano A, Molero-Labarta T, Mendez M, Moran-Jimenez MJ. A case of congenital dyserythropoietic anemia type IV. *Clin Case Rep*. 2017;5(3):248-252.
- Ortolano R, Forouhar M, Warwick A, Harper D. A Case of Congenital Dyserythropoietic Anemia Type IV Caused by E325K Mutation in Erythroid Transcription Factor KLF1. *J Pediatr Hematol Oncol*. 2018;40(6):e389-e391.
- Ravindranath Y, Johnson RM, Goyette G, Buck S, Gadgeel M, Gallagher PG. KLF1 E325K-associated Congenital Dyserythropoietic Anemia Type IV: Insights Into the Variable Clinical Severity. *J Pediatr Hematol Oncol*. 2018;40(6):e405-e409.
- Miller IJ, Bieker JJ. A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. *Mol Cell Biol*. 1993;13:2776-2786.
- Siatecka M, Bieker JJ. The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood*. 2011;118(8):2044-2054.
- Tallack MR, Perkins AC. KLF1 directly coordinates almost all aspects of terminal erythroid differentiation. *IUBMB Life*. 2010;62(12):886-890.
- Yien YY, Bieker JJ. EKLF/KLF1, a tissue-restricted integrator of transcriptional control, chromatin remodeling, and lineage determination. *Mol Cell Biol*. 2013;33(1):4-13.
- Gnanapragasam MN, Bieker JJ. Orchestration of late events in erythropoiesis by KLF1/EKLF. *Curr Opin Hematol*. 2017;24(3):183-190.
- Perkins A, Xu X, Higgs DR, et al. Kruppeling erythropoiesis: an unexpected broad spectrum of human red blood cell disorders due to KLF1 variants. *Blood*. 2016;127(15):1856-1862.
- Waye JS, Eng B. Kruppel-like factor 1: hematologic phenotypes associated with KLF1 gene mutations. *Int J Lab Hematol*. 2015;37 Suppl 1:78-84.
- Borg J, Patrinos GP, Felice AE, Philipsen S. Erythroid phenotypes associated with KLF1 mutations. *Haematologica*. 2011;96(5):635-638.
- Tallack MR, Perkins AC. Three fingers on the switch: Kruppel-like factor 1 regulation of gamma-globin to beta-globin gene switching. *Curr Opin Hematol*. 2013;20(3):193-200.
- Singleton BK, Frayne J, Anstee DJ. Blood group phenotypes resulting from mutations in erythroid transcription factors. *Curr Opin Hematol*. 2012;19(6):486-493.
- Helias V, Saison C, Peyrard T, et al. Molecular analysis of the rare in(Lu) blood type: toward decoding the phenotypic outcome of haploinsufficiency for the transcription factor KLF1. *Hum Mutat*. 2013;34(1):221-228.
- Singleton BK, Burton NM, Green C, Brady RL, Anstee DJ. Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype. *Blood*. 2008;112(5):2081-2088.
- Liu D, Zhang X, Yu L, et al. KLF1 mutations are relatively more common in a thalassemia endemic region and ameliorate the severity of beta-thalassemia. *Blood*. 2014;124(5):803-811.
- Viprakasit V, Ekwattanakit S, Rioueang S, et al. Mutations in Kruppel-like factor 1 cause transfusion-dependent hemolytic anemia and persistence of embryonic globin gene expression. *Blood*. 2014;123(10):1586-1595.
- Huang J, Zhang X, Liu D, et al. Compound heterozygosity for KLF1 mutations is associated with microcytic hypochromic anemia and increased fetal hemoglobin. *Eur J Hum Genet*. 2015;23(10):1341-1348.
- Magor GW, Tallack MR, Gillinder KR, et al. KLF1-null neonates display hydrops fetalis and a deranged erythroid transcriptome. *Blood*. 2015;125(15):2405-2417.
- Heruth DP, Hawkins T, Logsdon DP, et al. Mutation in erythroid specific transcription factor KLF1 causes Hereditary Spherocytosis in the Nan hemolytic anemia mouse model. *Genomics*. 2010;96(5):303-307.
- Siatecka M, Sahr KE, Andersen SG, Mezei M, Bieker JJ, Peters LL. Severe anemia in the Nan mutant mouse caused by sequence-selective disruption of erythroid Kruppel-like factor. *Proc Natl Acad Sci U S A*. 2010;107(34):15151-15156.
- Gillinder KR, Ilsley MD, Nebor D, et al. Promiscuous DNA-binding of a mutant zinc finger protein corrupts the transcriptome and diminishes cell viability. *Nucleic Acids Res*. 2017;45(3):1130-1143.
- Planutis A, Xue L, Trainor CD, et al. Neomorphic effects of the neonatal anemia (Nan-Eklf) mutation contribute to deficits throughout development. *Development*. 2017;144(3):430-440.
- Gnanapragasam MN, McGrath KE, Catherman S, Xue L, Palis J, Bieker JJ. EKLF/KLF1-regulated cell cycle exit is essential for erythroblast enucleation. *Blood*. 2016;128(12):1631-1641.
- Daheron L, D'Souza S. *Blood - SeV derived fibroblast generated iPSCs*. In: *StemBook*. 2013/05/10 ed. Cambridge, MA: Harvard Stem Cell Institute; 2008.
- Migliaccio G, Di Pietro R, di Giacomo V, et al. In vitro mass production of human erythroid cells from the blood of normal donors and of thalassemic patients. *Blood Cells Mol Dis*. 2002;28(2):169-180.
- Migliaccio G, Sanchez M, Masiello F, et al. Humanized culture medium for clinical expansion of human erythroblasts. *Cell Transplant*. 2010;19(4):453-469.
- Chen J, Peterson KR, Iancu-Rubin C, Bieker JJ. Design of embedded chimeric peptide nucleic acids that efficiently enter and accurately reactivate gene expression in vivo. *Proc Natl Acad Sci U S A*. 2010;107(39):16846-16851.
- Li J, Hale J, Bhagia P, et al. Isolation and transcriptome analyses of human erythroid progenitors: BFU-E and CFU-E. *Blood*. 2014;124(24):3636-3645.
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57.
- Borg J, Papadopoulos P, Georgitsi M, et al. Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nat Genet*. 2010;42(9):801-805.
- Zhou D, Liu K, Sun CW, Pawlik KM, Townes TM. KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. *Nat Genet*. 2010;42(9):742-744.
- Norton LJ, Funnell APW, Burdach J, et al. KLF1 directly activates expression of the novel fetal globin repressor ZBTB7A/LRF in erythroid cells. *Blood Adv*. 2017;1(11):685-692.
- Siatecka M, Lohmann F, Bao S, Bieker JJ. EKLF directly activates the p21WAF1/CIP1 gene by proximal promoter and novel intronic regulatory regions during erythroid differentiation. *Mol Cell Biol*. 2010;30(11):2811-2822.
- Russo R, Andolfo I, Manna F, et al. Increased levels of ERFE-encoding FAM132B in patients with congenital dyserythropoietic anemia type II. *Blood*. 2016;128(14):1899-1902.
- Pimentel H, Parra M, Gee S, et al. A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res*. 2014;42(6):4031-4042.
- An X, Schulz VP, Li J, et al. Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood*. 2014;123(22):3466-3477.
- Feng WC, Southwood CM, Bieker JJ. Analyses of β -thalassaemia mutant DNA interactions with erythroid Kruppel-like factor (EKLF), an erythroid cell-specific transcription factor. *J Biol Chem*. 1994;269:1493-1500.
- Klevit RE. Recognition of DNA by Cys2, His2 zinc fingers. *Science*. 1991;253:1367-1395.
- Letuve S, Lajoie-Kadoch S, Audusseau S, et al. IL-17E upregulates the expression of

- proinflammatory cytokines in lung fibroblasts. *J Allergy Clin Immunol*. 2006;117(3):590-596.
51. Zhang D, Xu C, Manwani D, Frenette PS. Neutrophils, platelets, and inflammatory pathways at the nexus of sickle cell disease pathophysiology. *Blood*. 2016;127(7):801-809.
 52. O'Brien KA, Farrar JE, Vlachos A, et al. Molecular convergence in ex vivo models of Diamond-Blackfan anemia. *Blood*. 2017; 129(23):3111-3120.
 53. Ulirsch JC, Lareau C, Ludwig LS, Mohandas N, Nathan DG, Sankaran VG. Confounding in ex vivo models of Diamond-Blackfan anemia. *Blood*. 2017;130(9):1165-1168.
 54. Farrar JE, Neuberger D, Triche T Jr, Bodine DM. Response: Making "perfect" the enemy of good. *Blood*. 2017;130(9):1168-1169.
 55. Masuda T, Wang X, Maeda M, et al. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science*. 2016;351(6270):285-289.
 56. Siatecka M, Xue L, Bieker JJ. Sumoylation of EKLf promotes transcriptional repression and is involved in inhibition of megakaryopoiesis. *Mol Cell Biol*. 2007; 27:8547-8560.
 57. Bouilloux F, Juban G, Cohet N, et al. EKLf restricts megakaryocytic differentiation at the benefit of erythrocytic differentiation. *Blood*. 2008;112(3):576-584.
 58. Frontelo F, Manwani D, Galdass M, et al. Novel role for EKLf in megakaryocyte lineage commitment. *Blood*. 2007;110(12):3871-3880.
 59. Isern J, Fraser ST, He Z, Zhang H, Baron MH. Dose-dependent regulation of primitive erythroid maturation and identity by the transcription factor Eklf. *Blood*. 2010; 116(19):3972-3980.
 60. Tallack MR, Perkins AC. Megakaryocyte-erythroid lineage promiscuity in EKLf null mouse blood. *Haematologica*. 2010; 95(1):144-147.
 61. Lu YC, Sanada C, Xavier-Ferrucio J, et al. The Molecular Signature of Megakaryocyte-Erythroid Progenitors Reveals a Role for the Cell Cycle in Fate Specification. *Cell Rep*. 2018;25(8):2083-2093.e4.
 62. Gallagher PG. Disorders of erythrocyte hydration. *Blood*. 2017;130(25):2699-2708.
 63. Mohandas N, Gallagher PG. Red cell membrane: past, present, and future. *Blood*. 2008; 112(10):3939-3948.
 64. Drissen R, Palstra RJ, Gillemans N, et al. The active spatial organization of the beta-globin locus requires the transcription factor EKLf. *Genes Dev*. 2004;18(20):2485-2490.
 65. Schoenfelder S, Sexton T, Chakalova L, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet*. 2010;42(1):53-61.
 66. Nebor D, Graber JH, Ciciotte SL, et al. Mutant KLF1 in Adult Anemic Nan Mice Leads to Profound Transcriptome Changes and Disordered Erythropoiesis. *Sci Rep*. 2018; 8(1):12793.
 67. Mendez-Enriquez E, Garcia-Zepeda EA. The multiple faces of CCL13 in immunity and inflammation. *Inflammopharmacology*. 2013;21(6):397-406.