

Erythro-megakaryocytic transcription factors associated with hereditary anemia

John D. Crispino¹ and Mitchell J. Weiss²

¹Division of Hematology/Oncology, Northwestern University, Chicago, IL; and ²Children's Hospital of Philadelphia, Philadelphia, PA

Most heritable anemias are caused by mutations in genes encoding globins, red blood cell (RBC) membrane proteins, or enzymes in the glycolytic and hexose monophosphate shunt pathways. A less common class of genetic anemia is caused by mutations that alter the functions of erythroid transcription factors (TFs). Many TF mutations associated with heritable anemia cause truncations or amino acid substitutions, resulting in the production of functionally altered proteins. Characterization of these mutant proteins has

provided insights into mechanisms of gene expression, hematopoietic development, and human disease. Mutations within promoter or enhancer regions that disrupt TF binding to essential erythroid genes also cause anemia and heritable variations in RBC traits, such as fetal hemoglobin content. Defining the latter may have important clinical implications for de-repressing fetal hemoglobin synthesis to treat sickle cell anemia and β thalassemia. Functionally important alterations in genes encoding TFs or their

cognate *cis* elements are likely to occur more frequently than currently appreciated, a hypothesis that will soon be tested through ongoing genome-wide association studies and the rapidly expanding use of global genome sequencing for human diagnostics. Findings obtained through such studies of RBCs and associated diseases are likely generalizable to many human diseases and quantitative traits. (*Blood*. 2014;123(20):3080-3088)

Introduction

The most common hereditary forms of anemia arise from mutations in genes that encode globins, red blood cell (RBC) membrane proteins, or enzymes. However, mutations in RBC transcription factor (TF) genes, or the *cis* elements through which they bind to regulate gene expression, account for a rare, but informative subset of anemias. Megakaryocytes (MEGs) and RBCs derive from a common bipotential MEG-erythroid progenitor (MEP), and their development is regulated by several common TFs, such as GATA binding protein 1 (GATA1) (Figure 1). Accordingly, *GATA1* gene mutations can present with anemia, thrombocytopenia, or both. Similarly, T-cell acute lymphoblastic leukemia 1 (TAL1), FOG family member 1 (FOG1), and nuclear factor, erythroid 2 (NF-E2) are erythro-megakaryocytic TFs that could hypothetically be altered in hereditary anemia/thrombocytopenia. In contrast, Fli-1 proto oncogene (FLI1) and Kruppel-like factor 1 (KLF1) drive mono-lineage production of MEGs and RBCs, respectively, and identified mutations in these genes affect only the single corresponding lineage.^{1,2} Extensive research over more than a decade has defined the biological properties of these and several other erythro-megakaryocytic TFs. Mutations that alter their structure and/or expression to cause human blood diseases provide valuable insights into the biology of hematopoiesis and hemoglobin switching. Further investigations into the mechanisms of these mutations should identify new strategies for treating more common congenital anemias, for example, by manipulating fetal hemoglobin expression in the setting of sickle cell anemia or β thalassemia.

Etiologies of hereditary anemias

Hereditary anemia encompasses a wide spectrum of RBC disorders that typically present in infants and children. These disorders can

be subclassified according to the RBC developmental stage most profoundly affected.

1. Shortened lifespan of circulating erythrocytes (hemolysis). This group includes the most common causes of hereditary anemia and is typically caused by defects in hemoglobin (sickle cell anemia and other hemoglobin variants), RBC membrane proteins (hereditary spherocytosis, elliptocytosis), or enzymatic pathways (glucose 6 phosphate dehydrogenase (G6PD) deficiency, glycolytic defects). Congenital dyserythropoietic anemia type II (CDA II) is also associated with early destruction of erythrocytes in the spleen.

2. Faulty maturation of erythroid precursors. Thalassemia syndromes are associated with impaired erythroid precursor maturation, termed "ineffective erythropoiesis." The CDAs are associated with accumulation of erythroid precursors that form RBCs inefficiently.³ Affected erythroblasts often exhibit bizarre dysplastic morphologies including multinuclearity, internuclear bridging, and megakaryoblastic features. Historically, CDAs have been classified into 3 types (I, II, and III) based on the appearance of erythroid precursors and serological studies. Causal mutations for these classical CDAs have been identified in *CDAN1* (CDA I), *C15ORF41* (CDA I), *SEC23B* (CDA II), and *KIF23* (CDA III).^{3,4} These genes do not encode typical hematopoietic TFs, although *CDAN1* is reported to regulate the transport of histones⁵ and is implicated in the impaired localization of HP1 α , a key component of heterochromatin.⁶ However, it is now clear that the CDAs are heterogeneous and some variant forms are caused by mutations in *GATA1* and *KLF1* (see below).

3. Reduced erythroid precursors (pure red cell aplasia/hypoplasia). The classical example is Diamond Blackfan anemia (DBA), a congenital anemia that typically presents in early childhood, often with associated developmental anomalies.⁷ DBA is genetically

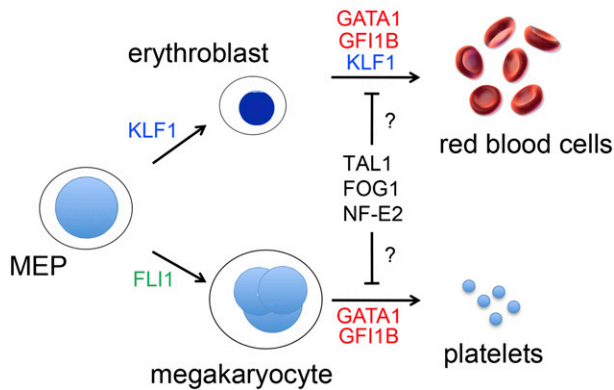


Figure 1. TFs that drive erythromegakaryopoiesis. RBCs and MEGs derive from the bipotential MEP. KLF1 and FLI1 regulate lineage determination of the MEP and subsequent erythroid or MEG maturation, respectively. GATA1, TAL1, FOG1, NF-E2, and GF11B regulate the maturation of both lineages. Germ-line mutations in KLF1 (shown in blue) cause isolated congenital anemia, whereas mutations in the TFs indicated in red affect RBCs and MEGs. Candidate TFs for which no germ-line human mutations have yet been discovered to be associated with anemia or thrombocytopenia are shown in black. Deficiency of FLI1 (shown in green) is associated with thrombocytopenia, but not anemia in Paris-Trousseau syndrome.¹¹²

heterogeneous, with about two-thirds of cases caused by dominantly inherited heterozygous loss-of-function mutations in 1 of ≥ 12 large or small ribosomal subunit proteins. It is not certain how ribosomal protein haploinsufficiency causes selective defects in the accumulation of erythroid precursors. This may occur in part via perturbation of TF networks. For example, imbalanced ribosome assembly may activate p53, a TF that induces apoptosis and/or cell cycle arrest. In addition, *GATA1* gene mutations are associated with an X-linked form of DBA⁸ (discussed in “How do defects in transcriptional regulation cause disease?”).

The majority of common hereditary anemias, including most hemoglobinopathies, membrane disorders, enzymopathies, and classical forms of CDA, are not caused by TF mutations. Once the clinical hematologist has ruled out these common causes, then mutations in erythroid or erythro-megakaryocytic TFs or their target *cis* elements become more likely and should be considered.

How do defects in transcriptional regulation cause disease?

Most essential erythroid or erythro-megakaryocytic TFs (so-called “master regulators”) activate or repress gene transcription globally to promote terminal maturation and suppress alternate lineages, respectively. Mutations that completely destroy the function of a TF by eliminating its expression or destabilizing the protein cause profound effects. In their homozygous state, these mutations are most likely embryonic lethal in humans, as predicted by numerous murine models. However, many mutations in human TF genes generate an altered, partially functional protein and are compatible with postnatal life, causing selective defects in gene expression and various clinical manifestations with a range of severity.

cis element mutations

TFs regulate gene expression by binding to specific *cis* element motifs in promoters and enhancers. Numerous human hematopoietic phenotypes are caused by polymorphisms or mutations in *cis* elements

that recruit key TFs to DNA. For example, a subset of β thalassemias are caused by mutations that affect TF binding sites in the proximal β globin gene promoter and deletions of an upstream enhancer termed the locus control region.⁹ More recently, mutations in the DNA regulatory regions of numerous other erythroid genes have been shown to cause anemia.

Mechanisms associated with mutations in TF genes or their *cis* elements in DNA regulatory regions are described generally in Figure 2, and specific salient examples are discussed in subsequent sections.

Mutations in erythro-megakaryocytic TFs required for RBC and platelet production. Congenital anemia that is not believed to be caused by common mutations in genes encoding globins, membrane proteins, enzymes, or classical CDA should alert the clinical hematologist to the potential for mutations in genes encoding TFs. Accompanying thrombocytopenia may suggest a mutation in a TF that is shared by erythroid and MEG lineages. TF genes discovered to be mutated in human diseases are described in this section.

GATA1. GATA1 is an X-linked TF that regulates maturation of multiple lineages including RBCs and MEGs.¹⁰ The full-length isoform contains 2 zinc fingers and an acidic N-terminal activation domain (NAD) (Figure 3A). The NAD is named by virtue of its ability to activate gene expression in transient reporter assays using heterologous cells.¹¹ The C-terminal zinc finger (Cf) is responsible for the bulk of DNA binding activity, whereas the N-terminal zinc finger (Nf) stabilizes DNA binding at more complex GATA elements, including palindromic sites.¹² The GATA1 Nf and Cf are similar in structure and both mediate interactions with multiple cofactor proteins. The GATA1 Nf specifically recruits the essential cofactor FOG1.¹³

The first report of *GATA1* mutations in defective erythropoiesis was published in 2000.¹⁴ Two half-brothers were born with severe anemia and thrombocytopenia, whereas their mother showed mild

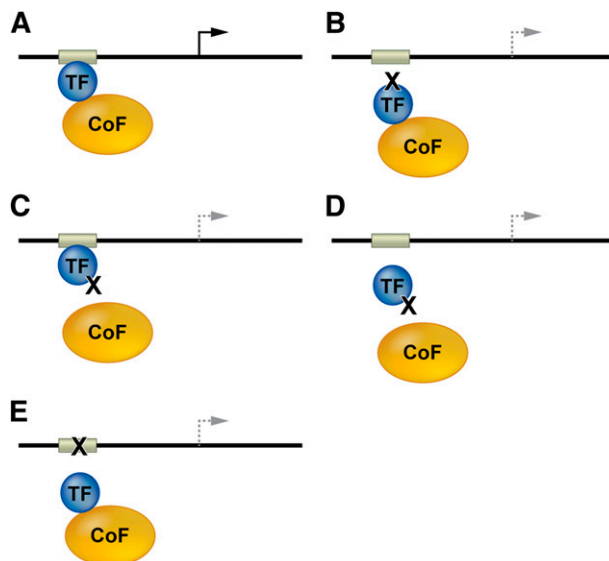


Figure 2. Mutations affecting transcription factor function. (A) TFs and bound cofactors (Cofs) regulate gene expression by binding cognate *cis* elements within DNA regulatory regions including promoters or enhancers (gray box). The arrow indicates gene transcription. (B) Mutations in the TF DNA binding domain inhibit TF interactions with some or all *cis* elements. The dashed lines indicate altered transcription. (C) TF mutations that inhibit Cof binding alter gene expression at some or all target genes. (D) Some Cof interactions are required for in vivo DNA occupancy by TFs. Thus, TF mutations outside of the DNA binding domain can inhibit gene occupancy. (E) Mutations in DNA *cis* elements can impair TF occupancy at a single site within one gene, thereby altering its transcription.

thrombocytopenia, initially thought to be caused by immune destruction. The patient's bone marrow displayed dyserythropoiesis and accumulation of dysplastic MEGs. Based on the known functions of GATA1 in these lineages and a pedigree configuration consistent with X-linked disease, mutated *GATA1* was suspected and confirmed by DNA sequencing. Affected individuals harbored a mutation that substituted methionine for valine at amino acid position 205 (V205M) within the GATA1 Nf. Simultaneously, it was determined that V205 participates in the binding of GATA1 to FOG1¹⁵ and that this interaction is inhibited by the V205M substitution. These findings demonstrated that the GATA1-FOG1 interaction is necessary for terminal maturation of both RBCs and MEGs, explaining the patient phenotype. Subsequently, several other groups identified different *GATA1* mutations in patients with variable degrees of thrombocytopenia and anemia (Figure 3A). These mutations most commonly introduce amino acid replacements into the Nf, affecting its ability to bind FOG1, TAL1, and associated proteins (discussed below) or DNA.^{10,16-20} Affected patients exhibit a range of distinct and overlapping phenotypes including CDA, thrombocytopenia, erythropoietic porphyria, thalassemia, and Gray platelet syndrome. Recently, a mutation in the C terminus of GATA1 was identified in a family with X-linked Lutheran (a-b-) blood group phenotype (absent expression of the Lutheran glycoprotein) and mild macrothrombocytopenia.²¹ Thus, alterations in *GATA1* cause a spectrum of red cell and platelet disorders whose precise characteristics relate to specific structure-function properties of the protein, many of which are not yet fully determined. Follow-up mechanistic studies have provided insights into this issue. For example, 2 missense mutations that alter the GATA1 Nf and impair DNA binding in vitro were found to have no effects on target gene occupancy when overexpressed in a cellular model for murine erythropoiesis.²⁰ Rather, these mutations inhibited recruitment of the TAL1 complex. Understanding how different *GATA1* mutations produce distinct phenotypes represents a current research challenge.

Mutations resulting in amino-truncated GATA1 (GATA1s)

Although complete loss of GATA1 leads to profound defects in the erythroid and MEG lineages (at least in mice), loss of the NAD causes less severe defects. Mutations that eliminate the NAD contribute to Down syndrome (DS)-associated transient myeloproliferative disorder (TMD) or acute megakaryoblastic leukemia (AMKL), underscoring its essential role in controlling megakaryopoiesis, particularly in the context of trisomy 21.²² In all cases, the acquired mutations lead to expression of a shortened isoform of GATA1, named GATA1s, which lacks amino acids 1-83 (Figure 3A). Loss of the N-terminal domain is associated with aberrant regulation of GATA1 target genes, defective differentiation activity, and expansion of a distinct fetal liver progenitor in vivo.²²

Germ-line GATA1s-type mutations occur in individuals without DS. In these cases, the mutation does not cause TMD or AMKL but rather leads to erythroid hypoplasia. For example, a germ-line *GATA1s* mutation was reported in a Brazilian pedigree with defective erythropoiesis, mild defects in MEGs, and neutropenia.²³ Last year, Gazda and colleagues performed exome sequencing on DBA patients lacking the typical ribosomal protein gene mutations. They discovered *GATA1s* mutations in 3 affected individuals from 2 unrelated pedigrees.⁸ A second study identified a *GATA1* mutation in a different unrelated patient.²⁴ These *GATA1* mutations identify a variant form of DBA, underscore the genetic and clinical heterogeneity of this disorder, and suggest a potential mechanistic link between ribosome biogenesis and GATA1 activity.²⁵

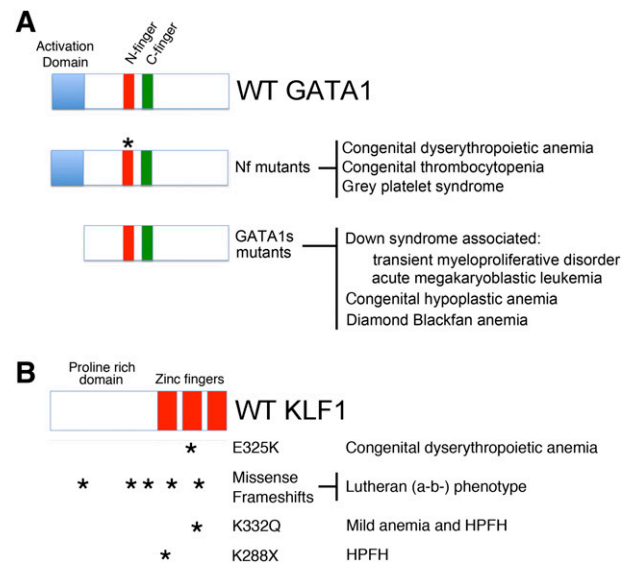


Figure 3. GATA1 and KLF1 domain modules, mutations, and associated diseases. (A) (Top) Three domains of GATA1 are shown: the N and C-terminal zinc fingers (Nf and Cf) and the NAD. (Middle) Mutations in the Nf (noted by asterisks) disrupt DNA binding or associations with the essential cofactors Friend of GATA1 (FOG-1) or TAL1 complex. These germ-line mutations cause a variety of inherited anemias and/or thrombocytopenias. (Bottom) Distinct mutations within exon 2 lead to predominant expression of GATA1s, which retains both zinc fingers, but is missing the NAD. In infants and young children with DS (trisomy 21), somatic mutations that favor GATA1s production cause TMD and AMKL. In euploid patients, germ-line GATA1s mutations are associated with congenital anemia, including DBA. (B) KLF1 includes an N-terminal proline-rich domain and 3 C-terminal zinc finger domains. Various missense and frameshift mutations (noted by asterisks) throughout the gene are associated with hereditary anemias, altered Lutheran blood group expression, and/or persistence of fetal hemoglobin expression. Note that E325K has a dominant negative effect. Adapted from Singleton et al.⁵⁸

Mutations that disrupt GATA1 binding to DNA regulatory regions

Several mutations in DNA *cis* elements recognized by GATA1 have been identified. For example, the RBC Duffy antigen null phenotype, a relatively common allele in Africans that confers resistance to *Plasmodium vivax* malaria, is caused by a mutation that disrupts a GATA1 binding site in the Duffy antigen/chemokine receptor gene (*DARC*) promoter.²⁶ X-linked sideroblastic anemia in some patients is associated with a mutation that disrupts a GATA1 binding motif in an enhancer that activates transcription of *ALAS2*, which encodes the enzyme catalyzing the first step in heme synthesis.^{27,28} Disruption of a GATA1 binding element in the promoter of the gene encoding uroporphyrinogen III synthase accounts for some cases of congenital erythropoietic porphyria.²⁹

How do GATA1 mutations cause disease?

Many questions remain regarding GATA1-associated blood disorders. How do different *GATA1* mutations confer distinct phenotypes in patients? What are the relevant target genes? How do GATA1s mutations cause anemia in children without DS, while cooperating with T21 to cause megakaryocytic defects in children with DS? Is there a relationship between GATA1 and ribosomal function in DBA?

Studies with animal models show that GATA1 is required for both RBC and MEG maturation. *Gata1*-null mouse embryos die in midgestation of anemia with premature cell death and/or impaired maturation of erythroid precursors. GATA1 activates and represses

hundreds of genes,³⁰⁻³³ and its loss presumably interferes with both induction of the entire erythroid maturation program and the suppression of genes regulating stem/progenitor and or alternate lineage development. The phenotypic variability in patients with different *GATA1* mutations is likely due to subtle alterations in binding to specific gene targets and/or various interacting proteins, such as FOG1, TAL1, and others,²⁰ although the specific biochemical and structural details are largely unknown.

Interestingly, adult mice with *GATA1*s-type mutations do not have any obvious hematopoietic defects, although fetal hematopoiesis is perturbed,³⁴ indicating that the manifestations of this mutation differ in humans who exhibit postnatal phenotypes. Consistent with these interspecies differences, *GATA1*s is produced naturally along with the full-length protein during human, but not murine, hematopoiesis. How *GATA1*s mutations cause anemia is not clear. One possibility is that these mutations cause reduced overall *GATA1* expression (*GATA1* full length + *GATA1*s), leading to net deficiency of the protein. In agreement, high levels of transgenic *GATA1*s expression can rescue anemia *Gata1* gene-ablated mice.³⁵ Alternatively, loss of the *GATA1* N terminus may qualitatively alter some of its functions that are particularly important for erythroid development.

GFI1B

Growth factor independent 1 (GFI1) is a zinc finger Snail/Gfi domain (SNAG) domain-containing transcriptional repressor that controls erythroid and MEG development, in part by forming distinct complexes with *GATA1* and chromatin-modifying enzymes that inhibit gene expression.³⁶⁻³⁸ Mice embryos lacking *Gfi1b* die in utero of anemia caused by defective erythroblast maturation and they also harbor developmentally arrested MEGs,³⁹ similar to the phenotype caused by loss of *GATA1*. GFI1B regulates proliferation and differentiation of MEPs through effects on transforming growth factor- β signaling.⁴⁰ One study found a heterozygous frameshift mutation in the *GFI1B* gene of a family with an autosomal dominant bleeding disorder, macrothrombocytopenia, and platelet dysfunction.⁴¹ Affected individuals also exhibited mild RBC abnormalities including anisopoikilocytosis and increased red cell distribution width. In addition, a *GFI1B* mutation causing the production of a dominant interfering protein was recently described to cause Gray platelet syndrome.⁴²

KLF1

KLF1 is the founding member of a family of Sp1 transcription factor (Sp1)-related TFs that contains 3 C-terminal C2H2 zinc finger DNA binding domains, as well as an N-terminal proline-rich region (Figure 3B).^{43,44} *Klf1*-null embryos die in midgestation from anemia with a β -thalassemia-like syndrome.^{45,46} KLF1 activates expression of key erythroid genes including those that encode β -globin, heme biosynthetic enzymes, RBC membrane proteins, and cell cycle regulators.⁴⁷⁻⁴⁹ Chromatin immunoprecipitation coupled with next-generation sequencing detected KLF1 binding at the regulatory regions of numerous erythroid genes, frequently in close proximity to *GATA1*-bound sites.⁴⁴ Another study found that *GATA1*, KLF1, and TAL1 co-occupy many erythroid-expressed genes.⁵⁰ KLF1 undergoes numerous post-translational modifications, including phosphorylation, sumoylation,

acetylation, and ubiquitinylation. These changes are believed to regulate its interactions with partner proteins and to control its activity in vivo.⁵⁰ In MEPs, KLF1 expression contributes to the specification of erythroid cell fate by competing with the MEG factor FLI1.^{1,2}

Mutations affecting KLF1 function cause isolated anemia without thrombocytopenia. Mutations that disrupt KLF1 binding to a cognate element in the human β globin gene (*HBB*) promoter cause β^+ thalassemia.⁹ Mutations in the human *KLF1* gene itself were first discovered in individuals with the Lutheran (a-b-) blood group phenotype,⁵¹ which also occurs in *GATA1*-mutant patients (discussed previously and reviewed in ref. 52). Twenty-one of 24 Lutheran (a-b-) blood group individuals studied harbored 9 different heterozygous *KLF1* loss-of-function mutations (Figure 3B). These individuals exhibited no anemia or overt RBC abnormalities. Therefore, 1 allele of *KLF1* appears to suffice for normal erythropoiesis, similar to what occurs in mice. Presumably, the gene encoding the Lutheran glycoprotein is positively regulated by *GATA1* and KLF1 and is highly sensitive to dosage of the latter.

Nan (neonatal anemia) mice harbor a *Klf1* missense mutation resulting in an E339D substitution within the second zinc finger of the protein.^{53,54} Homozygous embryos die in midgestation of anemia, whereas heterozygous animals display severe hemolytic anemia resembling hereditary spherocytosis.⁵⁵ *Nan* heterozygous mutant RBCs showed reduced expression of membrane cytoskeleton proteins including band3 and proteins 4.1 and 4.2, which are commonly deficient in hereditary spherocytosis.⁵³ In contrast to loss of a single allele of *Klf1*, which produces minimal effects on erythropoiesis, the *Nan* mutation acts dominantly, suggesting that the E339D substitution somehow interferes with the functions of the coexpressed wild-type KLF1 protein.

Analogous to the *Nan* mouse, a heterozygous human *KLF1* mutation at E325 (within the second zinc finger of the KLF protein) is associated with CDA⁵⁶ (Figure 3B). Affected individuals display severe anemia, jaundice, and high levels of circulating nucleated RBCs. Interestingly, *KLF1* E325K patients also express high levels of fetal hemoglobin (HbF, $\alpha 2\gamma 2$), suggesting a role for KLF1 in the γ to β globin switch that occurs perinatally (discussed in the next section). Of note, a recent study demonstrated that *KLF1* mutations cause severe congenital hemolytic anemia associated with a deficiency of pyruvate kinase in red cells and persistence of embryonic globin synthesis.⁵⁷ Affected individuals also displayed the Lutheran (a-b-) blood group phenotype.

How do KLF1 mutations cause disease?

Structure function assays examining various KLF1 mutations have provided interesting insights. For example, expression of the E339D mutant was found to interfere with the ability of KLF1 to bind and regulate a subset of targets depending on 1 specific nucleotide residing in the middle of the KLF1 consensus DNA binding motif.⁵³ Binding to genes with the consensus containing a cytidine, such as *Ahsp*, was maintained by E339D, whereas binding to genes with the consensus containing thymidine, including *Hbb* and *E2f2*, was disrupted by the mutation. Similar to murine E339D, the human E325K mutation produces a dominant-negative effect on KLF1-mediated transcription, including failure to bind and properly regulate a subset of target genes including *CD44* and *AQP1*.⁵⁶ Another report compared the DNA binding of mutations associated with the Lutheran (a-b-) blood group phenotype to ones seen in CDA or mild anemia.⁵⁸ This study found that mutants associated with the Lutheran (a-b-) phenotype (eg, R328H; Figure 3B) failed to bind DNA,

resulting in an overall reduction in the level of functional KLF1 and a mild phenotype. A mutant associated with mild anemia (K332Q) showed reduced binding to a number of DNA targets, but near normal binding to *Hbb*. Finally, the CDA E325K mutant displayed reduced binding to all promoters examined. In this case, however, the mutant evokes a strong phenotype even in the presence of the wild-type allele. This effect may reflect the ability of the mutant to destabilize TF complexes.⁵⁸

KLF1/BCL11A regulatory axis in hemoglobin switching

Different β globin-like genes are expressed sequentially during ontogeny.⁵⁹ γ globin is expressed predominantly during most of human fetal gestation to generate HbF ($\alpha_2\gamma_2$). Shortly after birth, there is a gradual switch from γ to β globin gene transcription with the resultant production of adult hemoglobin ($\alpha_2\beta_2$). Haploinsufficiency of *KLF1* and the E325K mutation are associated with hereditary persistence of fetal hemoglobin (HPFH),⁶⁰ a condition characterized by continued high-level expression of HbF in adult RBCs. Given that increased levels of HbF alleviate common hemoglobinopathies such as β -thalassemia and sickle cell anemia, understanding and manipulating the γ (*HBG*) to β (*HBB*) globin gene switch is of major clinical importance. Borg et al used genetic linkage analysis to identify chromosomal regions that are associated with HPFH in a Maltese family.⁶⁰ They discovered a heterozygous K288X *KLF1* mutation that is predicted to disrupt the DNA binding zinc finger domain (Figure 3B). Transcriptional profiling in patient erythroblasts demonstrated downregulation of many mRNAs, including *HBB* (a known KLF1 target) and B-cell lymphoma 11A (*BCL11A*), which encodes a TF that represses γ globin after birth.⁶¹⁻⁶⁵ Subsequent chromatin immunoprecipitation studies showed that wild-type KLF1 binds *BCL11A* promoter elements in adult, but not fetal liver erythroid progenitors.⁶⁰ Thus, KLF1 appears to mediate γ to β globin switching by binding the *BCL11A* gene and activating its transcription. Consistent with this model, deletion of an upstream *Klf1* enhancer in mice harboring a human β globin locus bacterial artificial chromosome inhibited expression of the endogenous murine *Klf1* and *Bcl11a* genes and increased the human γ -globin/ β -globin ratios in erythroblasts.⁶⁶ These genetic findings indicate that alterations in a KLF1-BCL11A transcription axis (Figure 4) account for some human HPFH. Several studies suggest that inhibition of human *BCL11A* (through KLF1-dependent or -independent mechanisms) can de-repress γ globin transcription and thereby alleviate the clinical manifestations of some human hemoglobinopathies, including sickle cell anemia and β -thalassemia.^{67,68}

Predicting new human hematopoietic disease alleles through genetic and biochemical studies

GATA1, KLF1, and Gfi1b, which are disrupted in human blood disorders (see previous sections), bind several other hematopoietic TFs to activate and repress gene expression during erythromegakaryopoiesis. Mutations in genes encoding any members of these complexes could potentially result in anemia and/or thrombocytopenia. Salient candidate genes include the following:

1. *TAL1* (previously also called stem cell leukemia [*SCL*]) encodes a basic helix-loop-helix TF that participates in the formation of

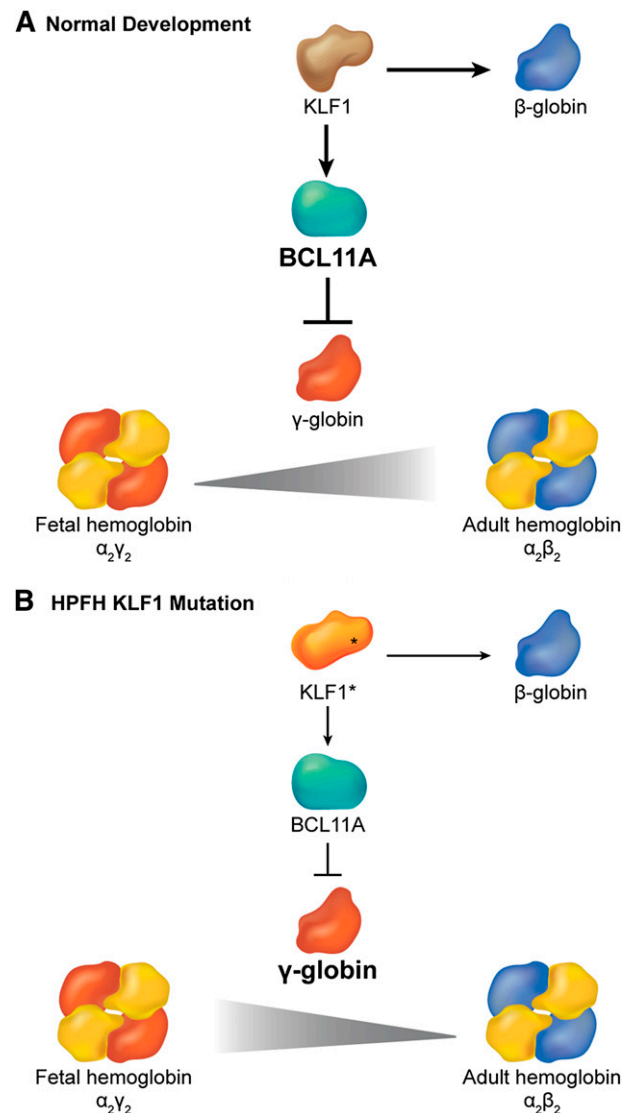


Figure 4. Interplay between KLF1 and BCL11A in the regulation of hemoglobin switching. (A) During normal development, KLF1 activates transcription of the *BCL11A* gene, which in turn represses γ globin expression promoting the switch from fetal ($\alpha_2\gamma_2$) to adult ($\alpha_2\beta_2$) hemoglobin. Simultaneously, high levels of KLF1 activate β -globin gene expression. (B) In some cases of HPFH, haplo-insufficiency of KLF1 (KLF1*) is associated with reduced BCL11A expression, which allows for persistence of fetal hemoglobin. Simultaneously, reduced KLF1 fails to activate β -globin expression.

hematopoietic stem cells during murine embryogenesis and erythromegakaryopoiesis postnatally.⁶⁹⁻⁷² TAL1 regulates gene expression via formation of a multisubunit complex that includes E2A, LIM binding domain 1 (LDB1), LIM domain only 2 (LMO2), and GATA proteins.⁷³ Somatic gene rearrangements resulting in TAL1 overexpression are associated with a subset of childhood T-cell acute lymphoblastic leukemias.⁷⁴ Given the essential role for TAL1 in erythropoiesis in mice,⁶⁹ it is possible that inherited hypomorphic or neomorphic mutations could cause CDA and/or thrombocytopenia. Such mutations may be associated with pleomorphic phenotypes because TAL1 is expressed in numerous non-hematopoietic tissues including endothelial and nerve cells.

2. GATA binding protein 2 (*GATA2*) encodes a GATA1-related protein with essential roles in normal and malignant hematopoiesis.^{75,76} *Gata2* knockout embryos die of anemia with defective production of

hematopoietic stem and progenitor cells.^{77,78} Germ-line heterozygous *GATA2* loss-of-function mutations cause several related diseases associated with myelodysplasia and/or immunodeficiency, including Mono-Mac syndrome, Emburger syndrome, and familial predisposition to acute myeloid leukemia.⁷⁹⁻⁸¹ The *GATA2* gene is expressed in early RBC precursors and downregulated during subsequent phases of maturation, in part through direct repression by *GATA1*. This “GATA switch,” whereby *GATA1* replaces *GATA2* at various sites in the genome, appears to be essential for normal erythropoiesis.⁸²⁻⁸⁴ Thus, *cis* elements in the *GATA2* gene that selectively mediate its downregulation during erythropoiesis represent potential target regions for mutations that cause CDA syndromes.⁸⁵

3. *FOG1* encodes a multi-zinc finger nuclear protein that does not bind DNA on its own, but provides essential functions in RBC and MEG development via physical interactions with the *GATA1* Nf.^{13,86} *Fog1*^{-/-} mice die in midgestation with impaired erythroid progenitor and MEG maturation, phenocopying loss of *GATA1*.⁸⁷ *GATA1* Nf missense mutations that impair binding to *FOG1* cause CDA and/or thrombocytopenia, emphasizing the importance of this association.²⁰ Based on these findings, germ-line *FOG1* mutations may cause anemia and/or thrombocytopenia. Such mutations could also be associated with congenital heart disease, as indicated by conditional gene targeting studies of *Fog1* in mice.⁸⁸

4. *LDB1* and *LMO2* encode nuclear adaptor proteins that form complexes with *GATA1*, *GATA2*, and *TAL1* in numerous hematopoietic cell types including stem/progenitor, erythroid, and MEGs.⁸⁹ These adaptors colocalize with *GATA1*, *KLF1*, and *TAL1* at numerous RBC- and MEG-expressed genes. *LDB1* self-oligomerizes and this property mediates long-range promoter-enhancer DNA looping interactions that activate or repress gene expression. Consistent with these findings, ablation of the *Ldb1* or *Lmo2* genes in mice cause severe anemia. Germ-line mutations in the corresponding human genes could cause inherited anemia, thrombocytopenia, or other blood abnormalities. Such mutations may cause additional developmental abnormalities, because *LDB1* and *LMO2* appear to have essential functions in many tissues.

5. The *NFE2* gene encodes nuclear factor-erythroid 2, discovered as protein heterodimer that binds conserved functional elements present in globins and many other erythroid and megakaryocytic genes.^{90,91} The NF-E2 complex is comprised of a hematopoietic specific 45-kDa basic leucine zipper protein and a widely expressed 18-kDa subunit of the Maf family. The *NFE2* gene is upregulated in polycythemia vera where it is believed to promote erythropoietin-independent erythropoiesis and expansion of the hematopoietic stem cell and common myeloid lineages.^{92,93} Mice engineered to overexpress NF-E2 protein develop similar phenotypes.⁹⁴ Moreover, patients with myeloproliferative neoplasms harbor somatic *NFE2* mutations that may activate protein function.⁹⁵ Thus, excessive NF-E2 activity appears to drive abnormal proliferation of erythroid, myeloid, and MEG lineages.

Nfe2^{-/-} mice exhibit profound thrombocytopenia with minimal anemia, most likely due to the presence of similar factors in erythroid cells with redundant functions.⁹⁶⁻⁹⁸ Thus, murine studies predict that human *NFE2* inactivating mutations would cause thrombocytopenia but not anemia. However, it is important to note that murine mutagenesis studies do not always predict human phenotypes. A germane example is *SEC23B*, which encodes a member of the coat protein complex II that facilitates exit of secreted proteins from the endoplasmic reticulum. Mutations in *SEC23B* cause CDA type II,⁹⁹

whereas ablation of the same gene in mice causes pancreatic secretory dysfunction but no anemia.¹⁰⁰ These disparate phenotypes may be caused by species-specific shifts in the functions of *SEC23B* and its closely related paralog *SEC23A*.

Quantitative trait loci-hematopoietic TF interactions

As discussed earlier in this review, mutations in TF binding DNA *cis* elements residing within promoters and enhancers can cause human anemias. Interestingly, mutations of this type can also exert subtle effects that are nonetheless clinically significant. Many quantitative trait loci identified by genome-wide association studies map to proximal *cis* elements that regulate the expression of nearby (or distant) genes,¹⁰¹ a concept that is relevant to RBC disorders. For example, genome-wide association studies have identified numerous loci that regulate RBC traits, including persistence of fetal hemoglobin expression in adults.^{102,103} Of particular interest, a DNA sequence variant that inhibits binding of *GATA1* and *TAL1* to a *BCL11A* gene enhancer causes elevated fetal hemoglobin in adult RBCs.¹⁰⁴ This finding is of major clinical significance because disruption of this enhancer region by genome editing¹⁰⁵ represents a potential means to de-repress γ globin synthesis for treating common hemoglobinopathies including sickle cell anemia and β thalassemia. Another interesting quantitative trait locus residing within the cyclin D3 (*CCND3*) gene regulates human RBC size and number. In this case, a nucleotide substitution modifies a *TAL1* binding site within an erythroid-specific enhancer to alter *CCND3* expression, thereby impacting cell division and cell size during erythroid maturation.¹⁰⁶ These recent discoveries of *cis* element mutations/polymorphisms that are associated with human RBC traits and anemias (discussed earlier and in Figure 2E) likely represent the tip of the iceberg. Genetic alterations in *cis* elements that regulate TF binding probably represent a relatively common etiology for blood diseases and an even more widespread mechanism for variation of RBC traits within populations. Of course, these concepts are generalizable to virtually all medically relevant quantitative traits and heritable disorders. Importantly, most *cis* element mutations would not be ascertained through whole genome exome sequencing.

Conclusions

Mutations in genes encoding hematopoietic TFs or their cognate *cis* elements are becoming increasingly appreciated as causes for congenital anemias and/or thrombocytopenias. Analysis of these mutations has taught us a great deal about TF function and mechanisms of gene expression. Two scientific advances have facilitated the recognition of these mutations: First, well-characterized phenotypes arising from murine loss-of-function studies (ie, gene knock-outs) identify candidate TF genes for targeted examination in unexplained inherited blood disorders. Second, nonbiased genome sequencing of affected human patients and pedigrees is increasingly used for genetic diagnosis.¹⁰⁷ Both technologies are rapidly evolving. For example, new genome editing methods, such as transcription activator-like effector nucleases and the clustered regulatory interspaced short palindromic repeats/Cas system, facilitate gene targeting in mice and human pluripotent stem cells,¹⁰⁸ which should increase the repertoire of candidate TF mutations predicted to

affect human hematopoiesis. Second, global exome and whole genome sequencing are becoming more available and cost effective for diagnostics. The latter, in particular, permits identification of *cis* element mutations outside of protein coding regions. Moreover, functional characterization of human and murine genomes, largely by The Encyclopedia of DNA Elements Consortium, is identifying thousands of potential tissue-specific enhancers that represent candidate regions for disease-associated mutations.¹⁰⁹⁻¹¹¹ Together, these technologies should identify causal mutations for most Mendelian disorders and many multigenic ones in the near future, including those that affect hematopoiesis. Understanding precisely how these mutations regulate TF function and gene expression represents an exciting and formidable long-term challenge.

Acknowledgments

The authors thank Gerd Blobel, Amy Campbell, and Vijay Sankaran for reviewing this manuscript.

Work on erythropoiesis in M.J.W.'s laboratory is supported by National Institutes of Health, National Institute of Diabetes

and Digestive and Kidney Diseases grants R01DK092318, P30DK090969, and R01DK065806, US Department of Defense grant BM090168, US Israel Binational Science Foundation grant 2009239, and the Jane Fishman-Grinberg Endowed Chair for Stem Cell Research. Research on erythropoiesis in J.D.C.'s laboratory is supported by National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases grant DK101329 and the Samuel Waxman Cancer Research Foundation.

Authorship

Contribution: J.D.C. and M.J.W. wrote the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: John D. Crispino, Northwestern University, Division of Hematology/Oncology, 303 East Superior St, Lurie Bldg, Room 5-113, Chicago, IL 60611; e-mail: j-crispino@northwestern.edu; and Mitchell Weiss, Children's Hospital of Philadelphia, Division of Hematology, Room 316B ARC, 3615 Civic Center Blvd, Philadelphia PA 19104; e-mail: weissmi@email.chop.edu.

References

- Boulloux F, Juban G, Cohet N, et al. EKLK restricts megakaryocytic differentiation at the benefit of erythrocytic differentiation. *Blood*. 2008;112(3):576-584.
- Frontelo P, Manwani D, Galdass M, et al. Novel role for EKLK in megakaryocyte lineage commitment. *Blood*. 2007;110(12):3871-3880.
- Iolascon A, Heimpel H, Wahlin A, Tamary H. Congenital dyserythropoietic anemias: molecular insights and diagnostic approach. *Blood*. 2013;122(13):2162-2166.
- Babbs C, Roberts NA, Sanchez-Pulido L, et al; WGS500 Consortium. Homozygous mutations in a predicted endonuclease are a novel cause of congenital dyserythropoietic anemia type I. *Haematologica*. 2013;98(9):1383-1387.
- Ask K, Jasencakova Z, Menard P, Feng Y, Almouzni G, Groth A. Codanin-1, mutated in the anaemic disease CDAL, regulates Asf1 function in S-phase histone supply. *EMBO J*. 2012;31(8):2013-2023.
- Renella R, Roberts NA, Brown JM, et al. Codanin-1 mutations in congenital dyserythropoietic anemia type 1 affect HP1alpha localization in erythroblasts. *Blood*. 2011;117(25):6928-6938.
- Horos R, von Lindern M. Molecular mechanisms of pathology and treatment in Diamond Blackfan Anaemia. *Br J Haematol*. 2012;159(5):514-527.
- Sankaran VG, Ghazvinian R, Do R, et al. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *J Clin Invest*. 2012;122(7):2439-2443.
- Thein SL. The molecular basis of β -thalassemia. *Cold Spring Harb Perspect Med*. 2013;3(5):a011700.
- Crispino JD. GATA1 in normal and malignant hematopoiesis. *Semin Cell Dev Biol*. 2005;16(1):137-147.
- Martin DI, Orkin SH. Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev*. 1990;4(11):1886-1898.
- Trainor CD, Omichinski JG, Vandergon TL, Gronenborn AM, Clore GM, Felsenfeld G. A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high-affinity interaction. *Mol Cell Biol*. 1996;16(5):2238-2247.
- Tsang AP, Visvader JE, Turner CA, et al. FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell*. 1997;90(1):109-119.
- Nichols KE, Crispino JD, Poncz M, et al. Familial dyserythropoietic anaemia and thrombocytopenia due to an inherited mutation in GATA1. *Nat Genet*. 2000;24(3):266-270.
- Crispino JD, Lodish MB, MacKay JP, Orkin SH. Use of altered specificity mutants to probe a specific protein-protein interaction in differentiation: the GATA-1:FOG complex. *Mol Cell*. 1999;3(2):219-228.
- Del Vecchio GC, Giordani L, De Santis A, De Mattia D. Dyserythropoietic anemia and thrombocytopenia due to a novel mutation in GATA-1. *Acta Haematol*. 2005;114(2):113-116.
- Tubman VN, Levine JE, Campagna DR, et al. X-linked gray platelet syndrome due to a GATA1 Arg216Gln mutation. *Blood*. 2007;109(8):3297-3299.
- Phillips JD, Steensma DP, Pulsipher MA, Spangrude GJ, Kushner JP. Congenital erythropoietic porphyria due to a mutation in GATA1: the first trans-acting mutation causative for a human porphyria. *Blood*. 2007;109(6):2618-2621.
- Kratz CP, Niemeyer CM, Karow A, Volz-Fleckenstein M, Schmitt-Gräff A, Strahm B. Congenital transfusion-dependent anemia and thrombocytopenia with myelodysplasia due to a recurrent GATA1(G208R) germline mutation. *Leukemia*. 2008;22(2):432-434.
- Campbell AE, Wilkinson-White L, Mackay JP, Matthews JM, Blobel GA. Analysis of disease-causing GATA1 mutations in murine gene complementation systems. *Blood*. 2013;121(26):5218-5227.
- Singleton BK, Roxby DJ, Stirling JW, et al. A novel GATA1 mutation (Stop414Arg) in a family with the rare X-linked blood group Lu(a-b-) phenotype and mild macrothrombocytic thrombocytopenia. *Br J Haematol*. 2013;161(1):139-142.
- Khan I, Malinge S, Crispino J. Myeloid leukemia in Down syndrome. *Crit Rev Oncog*. 2011;16(1-2):25-36.
- Hollanda LM, Lima CS, Cunha AF, et al. An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis. *Nat Genet*. 2006;38(7):807-812.
- Parrella SAA, Aspesi A, Quarello P, et al. Loss of GATA-1 full length as a cause of Diamond-Blackfan anemia phenotype [published online ahead of print January 22, 2014]. *Pediatr Blood Cancer*.
- Weiss MJ, Mason PJ, Bessler M. What's in a name? *J Clin Invest*. 2012;122(7):2346-2349.
- Tournamille C, Colin Y, Cartron JP, Le Van Kim C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet*. 1995;10(2):224-228.
- Campagna DR, de Bie CI, Schmitz-Abe K, et al. X-linked sideroblastic anemia due to ALAS2 intron 1 enhancer element GATA binding site mutations. *Am J Hematol*. 2013;89(3):315-319.
- Kaneko K, Furuyama K, Fujiwara T, et al. Identification of the novel erythroid-specific enhancer for ALAS2 gene and its loss-of-function mutation associated with congenital sideroblastic anemia. *Haematologica*. 2014;99(20):252-261.
- Solis C, Aizencang GI, Astrin KH, Bishop DF, Desnick RJ. Uroporphyrinogen III synthase erythroid promoter mutations in adjacent GATA1 and CP2 elements cause congenital erythropoietic porphyria. *J Clin Invest*. 2001;107(6):753-762.
- Welch JJ, Watts JA, Vakoc CR, et al. Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood*. 2004;104(10):3136-3147.
- Yu M, Riva L, Xie H, et al. Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol Cell*. 2009;36(4):682-695.
- Cheng Y, Wu W, Kumar SA, et al. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy,

- histone modifications, and mRNA expression. *Genome Res.* 2009;19(12):2172-2184.
33. Fujiwara T, O'Geen H, Keles S, et al. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell.* 2009;36(4):667-681.
 34. Li Z, Godinho FJ, Klusmann JH, Garriga-Canut M, Yu C, Orkin SH. Developmental stage-selective effect of somatically mutated leukemogenic transcription factor GATA1. *Nat Genet.* 2005;37(6):613-619.
 35. Shimizu R, Takahashi S, Ohneda K, Engel JD, Yamamoto M. In vivo requirements for GATA-1 functional domains during primitive and definitive erythropoiesis. *EMBO J.* 2001;20(18):5250-5260.
 36. Phelan JD, Shroyer NF, Cook T, Gebelein B, Grimes HL. Gfi1-cells and circuits: unraveling transcriptional networks of development and disease. *Curr Opin Hematol.* 2010;17(4):300-307.
 37. van der Meer LT, Jansen JH, van der Reijden BA. Gfi1 and Gfi1b: key regulators of hematopoiesis. *Leukemia.* 2010;24(11):1834-1843.
 38. Rodriguez P, Bonte E, Krijgsvelde J, et al. GATA-1 forms distinct activating and repressive complexes in erythroid cells. *EMBO J.* 2005;24(13):2354-2366.
 39. Saleque S, Cameron S, Orkin SH. The zinc-finger proto-oncogene Gfi-1b is essential for development of the erythroid and megakaryocytic lineages. *Genes Dev.* 2002;16(3):301-306.
 40. Randrianarison-Huetz V, Laurent B, Bardet V, Blobe GC, Huetz F, Duménil D. Gfi-1B controls human erythroid and megakaryocytic differentiation by regulating TGF-beta signaling at the bipotent erythro-megakaryocytic progenitor stage. *Blood.* 2010;115(14):2784-2795.
 41. Stevenson WS, Morel-Kopp MC, Chen Q, et al. GF11B mutation causes a bleeding disorder with abnormal platelet function. *J Thromb Haemost.* 2013;11(11):2039-2047.
 42. Monteferrario D, Bolar NA, Marneth AE, et al. A dominant-negative GF11B mutation in the Gray platelet syndrome. *N Engl J Med.* 2014;370(3):245-253.
 43. 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(5):2776-2786.
 44. 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.
 45. Perkins AC, Sharpe AH, Orkin SH. Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature.* 1995;375(6529):318-322.
 46. Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosfeld F. Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature.* 1995;375(6529):316-318.
 47. 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.
 48. Nilson DG, Sabatino DE, Bodine DM, Gallagher PG. Major erythrocyte membrane protein genes in EKLF-deficient mice. *Exp Hematol.* 2006;34(6):705-712.
 49. Tallack MR, Magor GW, Dartigues B, et al. Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq. *Genome Res.* 2012;22(12):2385-2398.
 50. Wontakal SN, Guo X, Smith C, et al. A core erythroid transcriptional network is repressed by a master regulator of myelo-lymphoid differentiation. *Proc Natl Acad Sci U S A.* 2012;109(10):3832-3837.
 51. 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.
 52. Singleton BK, Frayne J, Anstee DJ. Blood group phenotypes resulting from mutations in erythroid transcription factors. *Curr Opin Hematol.* 2012;19(6):486-493.
 53. 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.
 54. 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.
 55. Lyon MF. Dominant haemolytic anaemia. *Mouse News Lett.* 1993;68:68.
 56. 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.
 57. 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.
 58. 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.
 59. Sankaran VG, Orkin SH. The switch from fetal to adult hemoglobin. *Cold Spring Harb Perspect Med.* 2013;3(1):a011643.
 60. 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.
 61. Menzel S, Garner C, Gut I, et al. A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. *Nat Genet.* 2007;39(10):1197-1199.
 62. Lettre G, Sankaran VG, Bezerra MA, et al. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. *Proc Natl Acad Sci U S A.* 2008;105(33):11869-11874.
 63. Uda M, Galanello R, Sanna S, et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. *Proc Natl Acad Sci U S A.* 2008;105(5):1620-1625.
 64. Sankaran VG, Menne TF, Xu J, et al. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science.* 2008;322(5909):1839-1842.
 65. Sankaran VG, Xu J, Ragoczy T, et al. Developmental and species-divergent globin switching are driven by BCL11A. *Nature.* 2009;460(7259):1093-1097.
 66. 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.
 67. Wilber A, Hargrove PW, Kim YS, et al. Therapeutic levels of fetal hemoglobin in erythroid progeny of beta-thalassemic CD34+ cells after lentiviral vector-mediated gene transfer. *Blood.* 2011;117(10):2817-2826.
 68. Xu J, Peng C, Sankaran VG, et al. Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. *Science.* 2011;334(6058):993-996.
 69. Shivdasani RA, Mayer EL, Orkin SH. Absence of blood formation in mice lacking the T-cell leukaemia oncogene tal-1/SCL. *Nature.* 1995;373(6513):432-434.
 70. Robb L, Lyons I, Li R, et al. Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene. *Proc Natl Acad Sci U S A.* 1995;92(15):7075-7079.
 71. Mikkola HK, Klintman J, Yang H, et al. Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature.* 2003;421(6922):547-551.
 72. Hall MA, Curtis DJ, Metcalf D, et al. The critical regulator of embryonic hematopoiesis, SCL, is vital in the adult for megakaryopoiesis, erythropoiesis, and lineage choice in CFU-S12. *Proc Natl Acad Sci U S A.* 2003;100(3):992-997.
 73. Wadman IA, Osada H, Grütz GG, et al. The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* 1997;16(11):3145-3157.
 74. O'Neil J, Look AT. Mechanisms of transcription factor deregulation in lymphoid cell transformation. *Oncogene.* 2007;26(47):6838-6849.
 75. Bresnick EH, Katsumura KR, Lee HY, Johnson KD, Perkins AS. Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. *Nucleic Acids Res.* 2012;40(13):5819-5831.
 76. Vicente C, Conchillo A, Garcia-Sánchez MA, Odero MD. The role of the GATA2 transcription factor in normal and malignant hematopoiesis. *Crit Rev Oncol Hematol.* 2012;82(1):1-17.
 77. Tsai FY, Keller G, Kuo FC, et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. *Nature.* 1994;371(6494):221-226.
 78. Ling KW, Ottersbach K, van Hamburg JP, et al. GATA-2 plays two functionally distinct roles during the ontogeny of hematopoietic stem cells. *J Exp Med.* 2004;200(7):871-882.
 79. Hahn CN, Chong CE, Carmichael CL, et al. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet.* 2011;43(10):1012-1017.
 80. Hsu AP, Sampaio EP, Khan J, et al. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood.* 2011;118(10):2653-2655.
 81. Ostergaard P, Simpson MA, Connell FC, et al. Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat Genet.* 2011;43(10):929-931.
 82. Briegel K, Lim KC, Plank C, Beug H, Engel JD, Zenke M. Ectopic expression of a conditional GATA-2/estrogen receptor chimera arrests erythroid differentiation in a hormone-dependent manner. *Genes Dev.* 1993;7(6):1097-1109.
 83. Heyworth C, Gale K, Dexter M, May G, Enver TA. A GATA-2/estrogen receptor chimera

- functions as a ligand-dependent negative regulator of self-renewal. *Genes Dev.* 1999; 13(14):1847-1860.
84. Grass JA, Boyer ME, Pal S, Wu J, Weiss MJ, Bresnick EH. GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proc Natl Acad Sci U S A.* 2003; 100(15):8811-8816.
 85. Snow JW, Trowbridge JJ, Fujiwara T, et al. A single cis element maintains repression of the key developmental regulator Gata2. *PLoS Genet.* 2010;6(9):e1001103.
 86. Chlon TM, Crispino JD. Combinatorial regulation of tissue specification by GATA and FOG factors. *Development.* 2012;139(21):3905-3916.
 87. Tsang AP, Fujiwara Y, Horn DB, Orkin SH. Failure of megakaryopoiesis and arrested erythropoiesis in mice lacking the GATA-1 transcriptional cofactor FOG. *Genes Dev.* 1998; 12(8):1176-1188.
 88. Katz SG, Williams A, Yang J, et al. Endothelial lineage-mediated loss of the GATA cofactor Friend of GATA 1 impairs cardiac development. *Proc Natl Acad Sci U S A.* 2003;100(24): 14030-14035.
 89. Love PE, Warzecha C, Li L. Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends Genet.* 2014;30(1):1-9.
 90. Andrews NC, Erdjument-Bromage H, Davidson MB, Tempst P, Orkin SH. Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. *Nature.* 1993;362(6422): 722-728.
 91. Andrews NC. The NF-E2 transcription factor. *Int J Biochem Cell Biol.* 1998;30(4):429-432.
 92. Goerttler PS, Kreutz C, Donauer J, et al. Gene expression profiling in polycythaemia vera: overexpression of transcription factor NF-E2. *Br J Haematol.* 2005;129(1):138-150.
 93. Bogeska R, Pahl HL. Elevated nuclear factor erythroid-2 levels promote epo-independent erythroid maturation and recapitulate the hematopoietic stem cell and common myeloid progenitor expansion observed in polycythemia vera patients. *Stem Cells Transl Med.* 2013;2(2): 112-117.
 94. Kaufmann KB, Gründer A, Hadlich T, et al. A novel murine model of myeloproliferative disorders generated by overexpression of the transcription factor NF-E2. *J Exp Med.* 2012; 209(1):35-50.
 95. Jutzi JS, Bogeska R, Nikoloski G, et al. MPN patients harbor recurrent truncating mutations in transcription factor NF-E2. *J Exp Med.* 2013; 210(5):1003-1019.
 96. Shivdasani RA, Rosenblatt MF, Zucker-Franklin D, et al. Transcription factor NF-E2 is required for platelet formation independent of the actions of thrombopoietin/MGDF in megakaryocyte development. *Cell.* 1995;81(5):695-704.
 97. Shivdasani RA, Orkin SH. Erythropoiesis and globin gene expression in mice lacking the transcription factor NF-E2. *Proc Natl Acad Sci U S A.* 1995;92(19):8690-8694.
 98. Kotkow KJ, Orkin SH. Complexity of the erythroid transcription factor NF-E2 as revealed by gene targeting of the mouse p18 NF-E2 locus. *Proc Natl Acad Sci U S A.* 1996;93(8): 3514-3518.
 99. Schwarz K, Iolascon A, Verissimo F, et al. Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II. *Nat Genet.* 2009;41(8):936-940.
 100. Tao J, Zhu M, Wang H, et al. SEC23B is required for the maintenance of murine professional secretory tissues. *Proc Natl Acad Sci U S A.* 2012;109(29):E2001-E2009.
 101. Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* 2008;24(8): 408-415.
 102. van der Harst P, Zhang W, Mateo Leach I, et al. Seventy-five genetic loci influencing the human red blood cell. *Nature.* 2012;492(7429):369-375.
 103. Thein SL, Menzel S. Discovering the genetics underlying foetal haemoglobin production in adults. *Br J Haematol.* 2009;145(4):455-467.
 104. Bauer DE, Kamran SC, Lessard S, et al. An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science.* 2013;342(6155):253-257.
 105. Gaj T, Gersbach CA, Barbas CF III. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 2013; 31(7):397-405.
 106. Sankaran VG, Ludwig LS, Sicinska E, et al. Cyclin D3 coordinates the cell cycle during differentiation to regulate erythrocyte size and number. *Genes Dev.* 2012;26(18): 2075-2087.
 107. Sankaran VG, Gallagher PG. Applications of high-throughput DNA sequencing to benign hematology. *Blood.* 2013;122(22):3575-3582.
 108. Wang H, Yang H, Shivalila CS, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell.* 2013;153(4):910-918.
 109. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M; ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012; 489(7414):57-74.
 110. Consortium EP; ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 2011;9(4): e1001046.
 111. Mortazavi A, Pepke S, Jansen C, et al. Integrating and mining the chromatin landscape of cell-type specificity using self-organizing maps. *Genome Res.* 2013;23(12):2136-2148.
 112. Raslova H, Komura E, Le Couédic JP, et al. FLI1 monoallelic expression combined with its hemizygous loss underlies Paris-Trousseau/Jacobsen thrombopenia. *J Clin Invest.* 2004; 114(1):77-84.