

BRIEF DEFINITIVE REPORT

Impaired human hematopoiesis due to a cryptic intronic *GATA1* splicing mutation

Nour J. Abdulhay^{1,2*}, Claudia Fiorini^{1,2*}, Jeffrey M. Verboon^{1,2*}, Leif S. Ludwig^{1,2*} , Jacob C. Ulirsch^{1,2,3}, Barbara Zieger^{4,5,6}, Caleb A. Lareau^{1,2,3}, Xiaoli Mi^{1,2}, Anindita Roy^{7,8} , Esther A. Obeng^{1,2,9,10} , Miriam Erlacher^{4,5,6} , Namrata Gupta², Stacey B. Gabriel², Benjamin L. Ebert^{2,10,11}, Charlotte M. Niemeyer^{4,5,6}, Rami N. Khoriaty¹², Philip Ancliff⁷, Hanna T. Gazda^{2,13}, Marcin W. Wlodarski^{4,5,6}, and Vijay G. Sankaran^{1,2} 

Studies of allelic variation underlying genetic blood disorders have provided important insights into human hematopoiesis. Most often, the identified pathogenic mutations result in loss-of-function or missense changes. However, assessing the pathogenicity of noncoding variants can be challenging. Here, we characterize two unrelated patients with a distinct presentation of dyserythropoietic anemia and other impairments in hematopoiesis associated with an intronic mutation in *GATA1* that is 24 nucleotides upstream of the canonical splice acceptor site. Functional studies demonstrate that this single-nucleotide alteration leads to reduced canonical splicing and increased use of an alternative splice acceptor site that causes a partial intron retention event. The resultant altered *GATA1* contains a five-amino acid insertion at the C-terminus of the C-terminal zinc finger and has no observable activity. Collectively, our results demonstrate how altered splicing of *GATA1*, which reduces levels of the normal form of this master transcription factor, can result in distinct changes in human hematopoiesis.

Introduction

While hematopoiesis is arguably one of the best-understood paradigms of cellular differentiation in physiology, many facets of this process remain to be characterized. Genetic blood disorders provide an opportunity to learn more about hematopoiesis, even in cases where the mutated gene has been previously well studied, since allelic variation can provide novel biological insights (Casanova et al., 2014; Basak et al., 2015; Kuehn et al., 2016; Polfus et al., 2016; Kim et al., 2017; Zmajkovic et al., 2018). Advances in sequencing technologies have enabled rapid and efficient mutation identification, but deciphering the pathogenicity of noncoding variation can be immensely challenging. Noncoding transcriptional regulatory elements can be interrogated using a number of functional approaches, including the use of genome editing and exogenous assays of regulatory function (Ulirsch et al., 2016; Wakabayashi et al., 2016; Gasperini et al., 2017). However, deciphering the pathogenicity of cryptic splicing mutations can present substantial challenges (Rosenberg et al., 2015; Cummings et al., 2017).

Here we have identified two unrelated patients with a unique form of dyserythropoietic anemia that is associated with other abnormalities in hematopoiesis. Both of the patients harbor identical intronic mutations in *GATA1*. Through functional studies, we show that this single-nucleotide change leads to altered splice acceptor usage, along with a decrease in canonical *GATA1* splicing and expression, thereby resulting in the disease phenotype. This mutation disrupts the activity of the U2 splicing complex in the last intron of *GATA1*. While this complex is also mutated in more commonly observed cases of myelodysplastic syndrome (MDS), which can be attributable to somatic mutations in this splicing regulatory complex itself (Yoshida et al., 2011; Sperling et al., 2017), we were not able to identify a connection between the altered splicing observed in these two cases of dyserythropoietic anemia and more common cases of MDS. Nonetheless, our study illustrates how decreasing the splicing efficiency of the hematopoietic master regulator *GATA1* can impair selective aspects of human hematopoiesis and lead to

¹Division of Hematology/Oncology, The Manton Center for Orphan Disease Research, Boston Children's Hospital and Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA; ²Broad Institute of MIT and Harvard, Cambridge, MA; ³Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA; ⁴Division of Pediatric Hematology and Oncology, Department of Pediatrics and Adolescent Medicine, Faculty of Medicine, Medical Center-University of Freiburg, Freiburg, Germany; ⁵German Cancer Consortium, Freiburg, Germany; ⁶German Cancer Research Center, Heidelberg, Germany; ⁷Department of Paediatric Haematology, Great Ormond Street Hospital for Children, London, UK; ⁸Department of Paediatrics, University of Oxford, Children's Hospital, John Radcliffe Hospital, Oxford, UK; ⁹Division of Molecular Oncology, St. Jude Children's Research Hospital, Memphis, TN; ¹⁰Division of Hematology, Brigham and Women's Hospital, Boston, MA; ¹¹Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA; ¹²Division of Hematology and Oncology, Department of Internal Medicine, Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, MI; ¹³Division of Genetics and Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Harvard Medical School, Boston, MA.

*N.J. Abdulhay, C. Fiorini, J.M. Verboon, and L.S. Ludwig contributed equally to this paper; Correspondence to Vijay G. Sankaran: sankaran@broadinstitute.org.

© 2019 Abdulhay et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms/>). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).

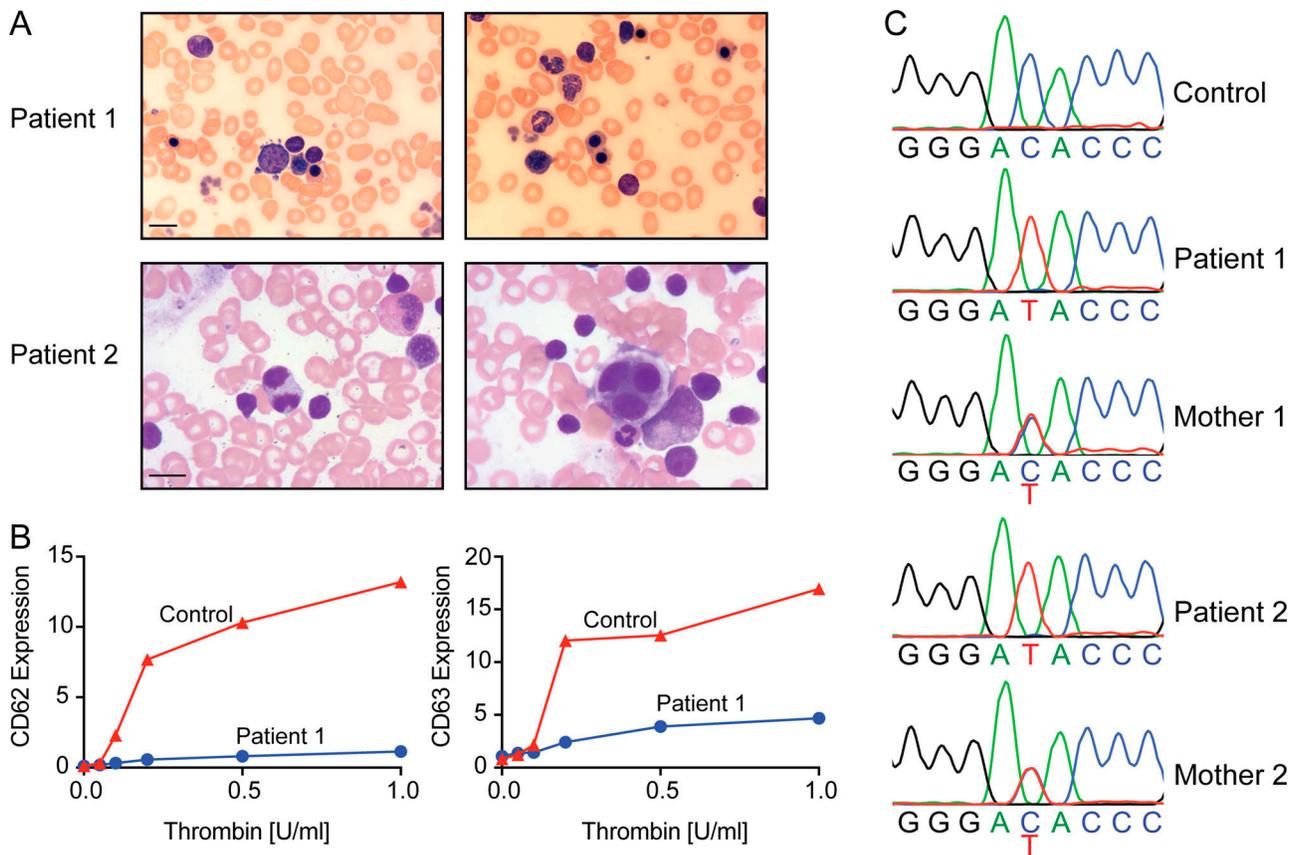


Figure 1. **Identification of a *GATA1* intronic mutation in two unrelated patients with dyserythropoietic anemia.** (A) Images of bone marrow aspirates from patients 1 and 2 are shown as images taken with 40× and 60× objectives, respectively. Notable features in both patients include mild to moderate dyserythropoiesis and other alterations, including hypolobated dysplastic megakaryocytes. (B) The thrombin-induced expression of CD62 and CD63 on platelets from patient 1 or a healthy control was detected by flow cytometry (shown as a percentage). There is defective expression of these antigens, indicating an α - δ -granule deficiency. (C) Sequencing chromatograms of mutation in *GATA1* (chrX:48,652,176 C>T in hg19) from a healthy control, the two patients, and their mothers. Scale bars: 10 μ m.

distinct phenotypes in comparison with other pathogenic mutations affecting the same gene (Sankaran et al., 2012; Campbell et al., 2013; Crispino and Horwitz, 2017).

Results and discussion

In the course of studying a cohort of patients with rare genetic blood disorders, we encountered two patients who both had a distinct form of dyserythropoietic anemia that necessitated transfusions in the intrauterine period and/or in early infancy with subsequent evolution into a milder anemia that was noted to worsen in the setting of intercurrent illnesses. Bone marrow evaluation of the patients revealed moderate dyserythropoiesis, along with frequent, small hypolobated megakaryocytes and occasional dysplastic myeloid cells (Figs. 1 A and S1 A). Platelet function testing in patient 1, who was noted to have clinical bleeding, revealed specific impairments, including defective aggregation and α - δ -granule secretion (as indicated by decreased expression of P-selectin/CD62 and LAMP-3/CD63), which are suggestive of defective thrombopoiesis (Figs. 1 B and S1 B). Patient 2 was also noted to have abnormal platelet function testing. Further clinical details of the two patients are described

in Materials and methods. To assess the etiology of the impaired hematopoiesis in these patients, whole-exome sequencing (WES) and targeted mutation analysis were performed. No known blood disorder-associated mutations were identified (Materials and methods), but both patients were noted to have a unique mutation in the fifth intron of *GATA1* (chrX:48,652,176 C>T in hg19) that was carried in the mothers of these patients (Figs. 1 C and S1 C). This mutation was absent from the 123,136 exomes and 15,496 genomes in the gnomAD database, despite excellent coverage of this region (Materials and methods).

The identified mutation was located 24 nucleotides upstream of the canonical splice acceptor site in the fifth intron of *GATA1*. While the mutation was not predicted to disrupt known splicing elements, we hypothesized that it may impair appropriate splicing of this gene. To directly interrogate this, we used a minigene reporter assay (Fig. 2, A and B; Kishore et al., 2008). We found that the mutation reduced normal splicing of this region of *GATA1* and promoted an intron retention event of 15 nucleotides involving an alternative splice acceptor site (Fig. 2 B). Semiquantitative RT-PCR analysis revealed a reduction of canonical *GATA1* splicing to 42% of normal levels, with the mutant mRNA at 36% of normal levels (Fig. 2 C). Importantly, we

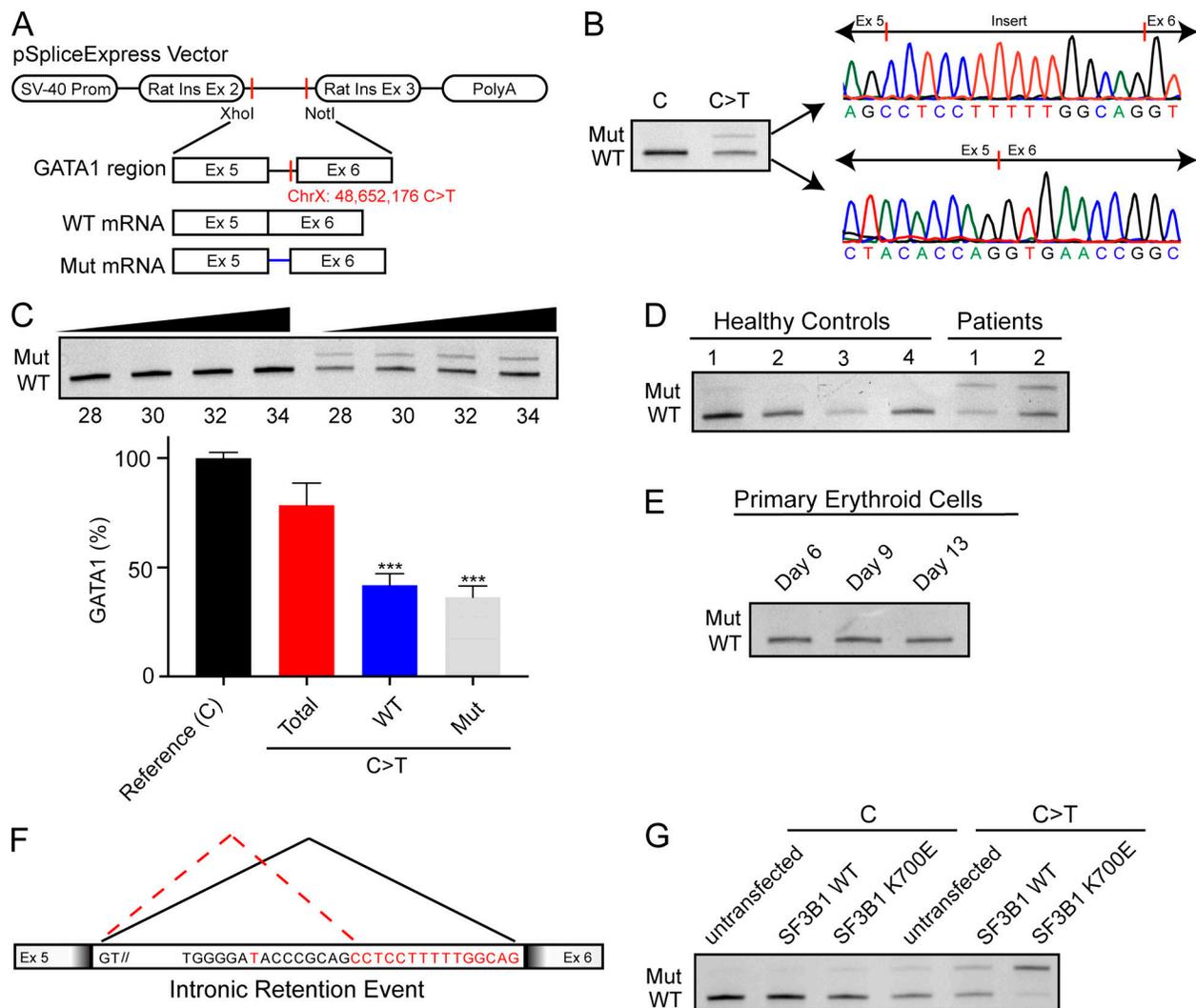


Figure 2. Decreased canonical splicing and intron retention due to a pathogenic *GATA1* mutation. (A) Schematic of the minigene assay involving exons 5 and 6 and the intervening intron of *GATA1*. The minigene vector, pSpliceExpress, includes two exons from rat insulin (Rat Ins Ex) as a control for splicing, a SV-40 promoter (Prom), and a polyadenylation site (PolyA). (B) Representative RT-PCR from the minigene assay in the presence or absence of the mutation shows the reduction of canonical splicing and presence of an intron retention event, which was confirmed through cloning and Sanger sequencing of the resultant products. (C) Semiquantitative RT-PCR analysis of the minigene assay using 28, 30, 32, and 34 PCR cycles. The bar graph (below) depicts quantified levels of the various products, including the total amount of *GATA1* (red), the amount of the wild-type band (blue), or the amount of the mutant (Mut) band in the setting of the mutation involving an intron retention event (***, $P < 0.001$; $n = 4$). (D) RT-PCR analysis of *GATA1* exon 5 and exon 6 from control and patient peripheral blood mononuclear cells. (E) Representative RT-PCR analysis of *GATA1* exons 5 and 6 from human HSPCs undergoing erythroid differentiation. (F) Schematic demonstrating alternative splicing in the fifth intron of *GATA1*, including the location of the intronic mutation and the nucleotides involved in intron retention (red). The splicing patterns are depicted above with a red dashed (alternative) or solid black (canonical) line. (G) RT-PCR analysis of the minigene assay performed as above or with increased expression of wild-type SF3B1 or the K700E SF3B1 mutated protein in 293T cells.

could confirm that this altered splicing was present in the patient samples but was found at significantly lower levels in healthy controls (Figs. 2 D and 3, A and B). By sequencing, we confirmed that the patients had an identical intron retention event of the same 15 nucleotides noted in the minigene assay. We did not identify any major variation in *GATA1* splicing over the course of erythroid differentiation of primary adult CD34⁺ hematopoietic stem and progenitor cells (HSPCs; Hu et al., 2013; Ulirsch et al., 2016; Figs. 2 E and 3), suggesting that the significantly altered splicing observed in these cases was attributable to the distinctive pathogenic variants. Importantly, we note that this altered splice variant is found in only 1.9–4.6% of *GATA1*

transcripts during normal human erythropoiesis, emphasizing its minor contribution to total *GATA1* mRNA levels (Fig. 3 B).

We next wanted to further investigate the underlying mechanisms for the altered splicing present due to these pathogenic mutations. Interestingly, the observed alternative splice acceptor site usage with resultant intron retention (Fig. 2 F) was similar to what is frequently observed in SF3B1-mutated MDS, where there is alternative splice acceptor site usage in a number of transcripts between ~15 and 25 nucleotides upstream of the canonical splice acceptor site (DeBoever et al., 2015; Obeng et al., 2016). Given the phenotypic similarity between MDS and the disordered hematopoiesis observed in these patients, as well as

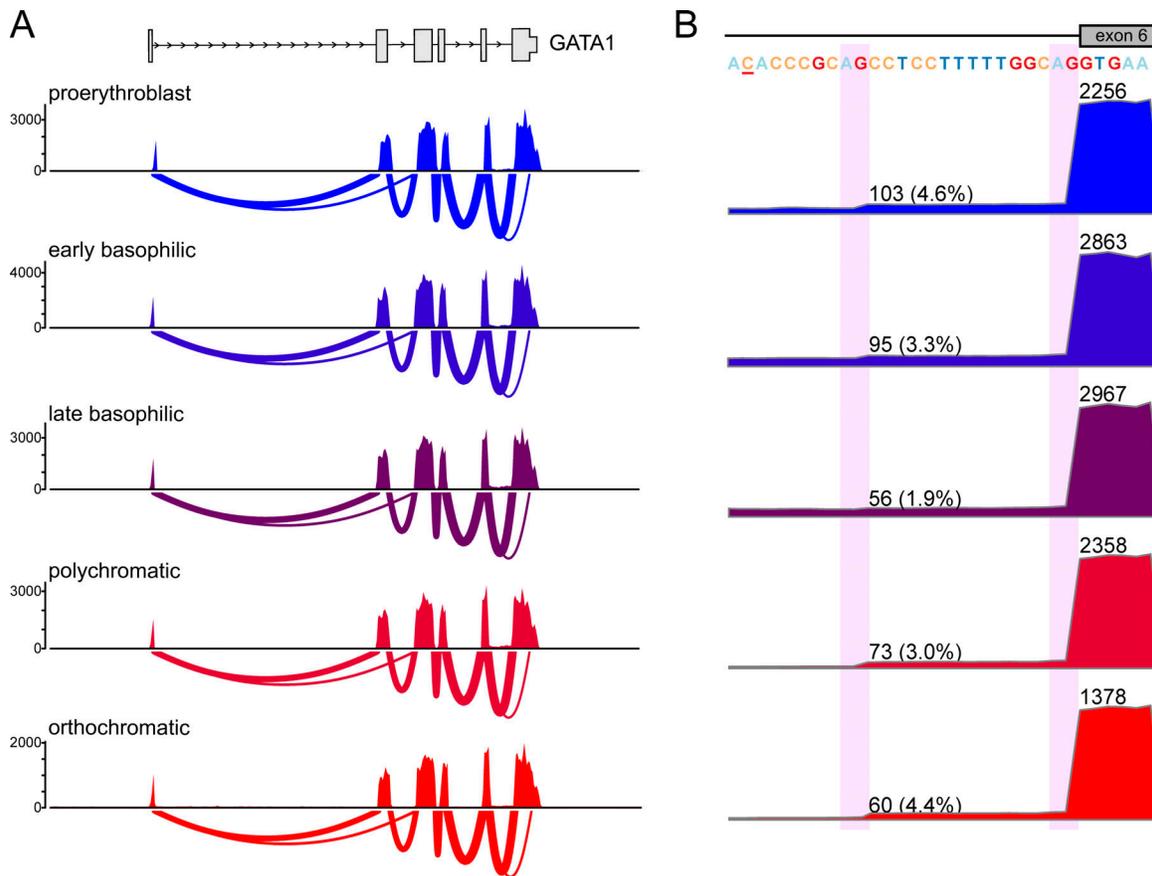


Figure 3. **Rare intron retention events during physiological human erythroid differentiation.** (A) Sashimi plots showing splicing events in the *GATA1* locus across the indicated stages of human erythropoiesis as determined by RNA-seq analysis. (B) Quantification of rare events of intron retention at indicated stages as counts of junction-spanning reads with the percentage of total transcript reads this represents shown in parentheses. Alternate and canonical splice acceptors are highlighted in pink. One representative example shown per stage of erythropoiesis.

the likely involvement of the U2 splicing complex in the region where this mutation resided, we tested whether the commonly observed SF3B1 K700E mutant, which alters U2 splicing activity, would further diminish the observed altered splicing due to this intronic *GATA1* mutation. Using the minigene assay described above, we found that expression of the SF3B1 K700E mutant markedly impaired canonical splicing of the mutant *GATA1* and led to near-complete alternative splice acceptor usage (Fig. 2 G). This finding suggests that the observed germline mutation in these patients specifically alters the activity of the U2 splicing complex in this region of *GATA1*. Given the involvement of this complex in MDS, we examined whether SF3B1-mutated MDS cases (with a diagnosis of refractory anemia with ringed sideroblasts) may have altered splicing or levels of *GATA1*, but we did not observe consistent defects, suggesting altered splicing or the presence of reduced levels of *GATA1* mRNA, even in sorted stage-matched bone marrow populations (Fig. S2). This finding is not surprising, given the minimal effect of the SF3B1 K700E mutation on splicing of the wild-type form of *GATA1* in the minigene assay discussed above (Fig. 2 G). Nonetheless, given the phenotypic and molecular connections, we felt it worthwhile to explore possible connections between these conditions.

To gain further insight into the mechanism by which the *GATA1* intronic mutant disrupts human hematopoiesis, we examined whether the protein produced from the intron retention event—which would add five additional amino acids at the C-terminus of the C-terminal zinc finger of *GATA1*—may have altered activity (Fig. 4 A). In human embryonic kidney 293T cells, we found that both RNA levels (as inferred through the expression of a linked GFP molecule that is translated from an internal ribosome entry site on the same transcript and through direct measurement) and protein levels (as measured directly) of this mutated cDNA were stable and similar to what is observed with wild-type *GATA1* cDNA (Fig. 4, B and C). Moreover, in the mouse *Gata1*-null erythroid cell line, G1E, we found similar expression of the mutated and wild-type proteins (Fig. 4 D). Exogenous expression of the mutated cDNA in primary HSPCs induced to undergo erythroid differentiation revealed that the mutated form had little effect on promoting precocious erythroid differentiation, as occurs with the wild-type cDNA, and importantly there was no dominant-negative activity observed in this setting where wild-type *GATA1* protein is present (Fig. 4 E; Ludwig et al., 2014; Giani et al., 2016; Wakabayashi et al., 2016). To directly assess whether the protein formed by this mutation is inactive or hypomorphic,

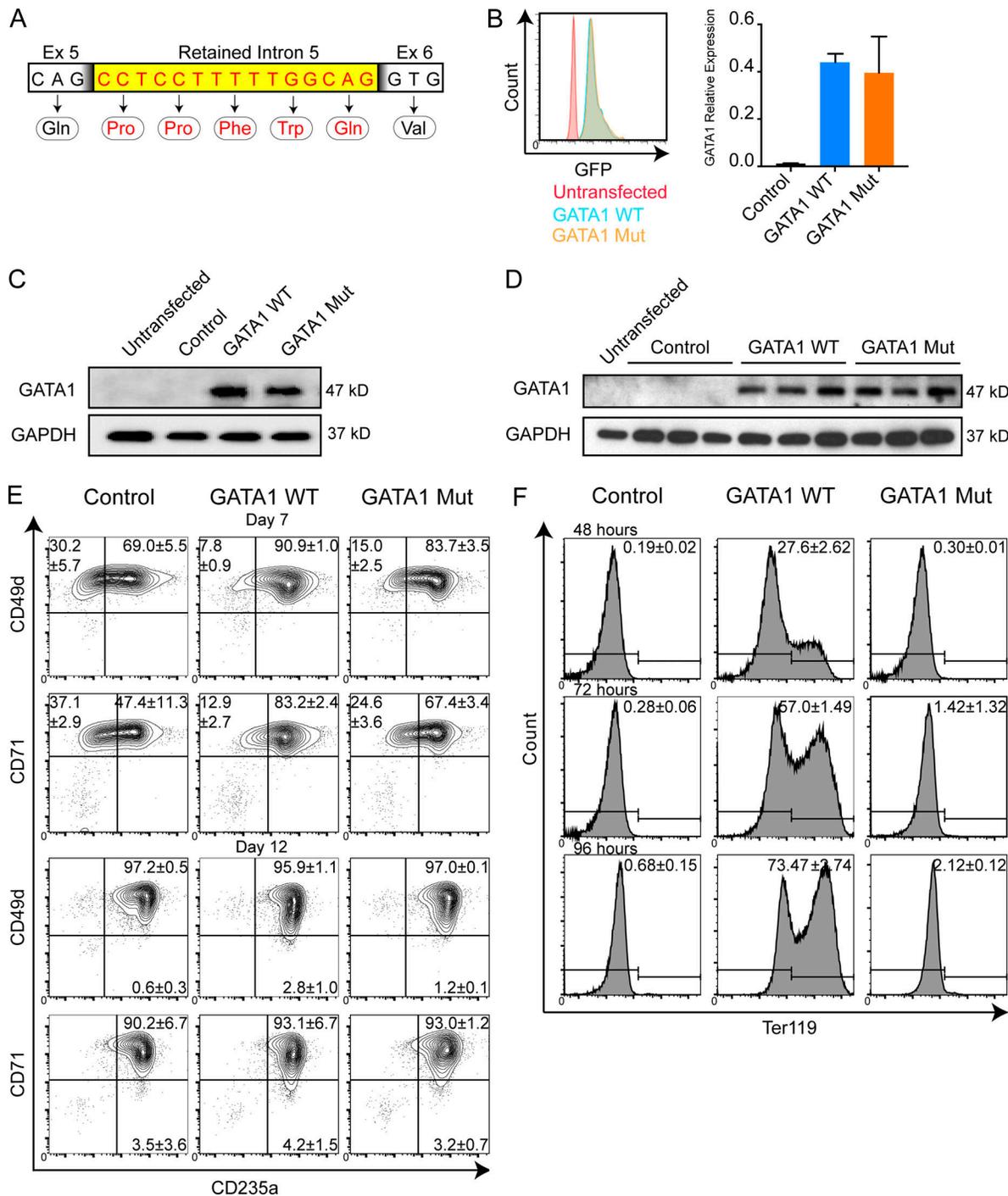


Figure 4. GATA1 variant produced through the intron retention event is stably expressed, but shows no function. (A) Schematic shows insertion of five amino acids as a result of the intron retention event. (B) A histogram plot (left) shows GFP levels indicating robust expression and equivalent transfection efficiency (GFP linked to GATA1 cDNA by internal ribosomal entry site). Mutant and wild-type GATA1 mRNA are expressed at equivalent levels (right). (C) A representative Western blot of GATA1 levels from exogenous cDNA expression in 293T cells. Loading control is GAPDH. (D) Western blot of GATA1 levels from exogenous cDNA expression in erythroid G1E cells (three biological replicates per indicated construct). Loading control is GAPDH. (E) Representative flow cytometric assessment (of three replicates) of erythroid differentiation (using CD235a, CD71, and CD49d surface markers) on day 7 and day 12 of culture of human HSPCs upon exogenous expression of wild-type GATA1 or the GATA1 mutant protein. (F) Representative histogram plots (of three replicates) assessing the marker of mouse erythroid differentiation, Ter119, in G1E cells infected with the empty vector (control), GATA1 wild type, or GATA1 mutant cDNAs vectors after 48, 72, or 96 h.

we used the GIE *Gata1*-null cell line, to complement this phenotype with either the wild-type or mutant cDNAs (Rylski et al., 2003; Campbell et al., 2013). While the wild-type cDNA robustly promoted erythroid differentiation in this setting based on Ter119 marker induction, the mutant form failed to do so, suggesting that this mutated protein produced from an intron retention event may result in loss of function (Figs. 4 F and S3).

To globally assess the potentially altered transcriptional activity of the mutant splice variant of GATA1, we infected GIE cells with either the wild-type or mutant cDNA and conducted RNA sequencing (RNA-seq) 72 h after infection. While wild-type GATA1 either activated or repressed a large number of canonical target genes (Fig. 5, A, E, and F), expression of the mutant form had little change compared with the control lentivirus (Fig. 5, B, E, and F). Importantly, expression of the cDNAs was at comparable levels (Fig. 5, C and D). Interestingly, we did note some subtle repression of select GATA1-repressed genes by the mutant form (cluster K3, Fig. 5 F). To investigate this further, we assessed GATA1 chromatin occupancy near the transcription start sites (TSSs) of these genes as measured by chromatin immunoprecipitation sequencing (ChIP-seq; ENCODE Project Consortium, 2012) but noted very little occupancy in this set of genes, in contrast to the majority of GATA1 activated genes (Fig. 5, F and G, cluster K1 vs. K3). Together with the data shown above from the interrogation of splicing, these results demonstrate that the impaired hematopoiesis in these cases emerges due to reduced expression of wild-type GATA1 to ~40% of normal levels with the production of an inactive mutant protein due to an intron retention event.

Our findings have several important and broad implications. While determining the functional consequences of cryptic splicing mutations can be challenging, we illustrate how by using a series of functional assays, the mechanisms underlying such mutations can be more fully understood. Exactly how the activity of the U2 splicing complex is altered in these cases remains uncertain, but given the lack of consensus binding sequences, no clear-cut mechanisms have been apparent. Given the finding of interactions between a mutant form of *SF3B1* commonly observed in MDS cases and the germline mutation we identified, as well as the phenotypic similarity between these disorders, it is interesting to speculate about potential connections between these observations. However, we have not been able to demonstrate aberrant GATA1 mRNA splicing or expression in MDS cases in our analyses, but given the phenotypic similarities between this form of dyserythropoietic anemia and MDS, the reported perturbation of GATA1 protein levels in MDS patient samples (Frisan et al., 2012), and the myelodysplasia seen in mouse models with reduced *Gata1* levels (McDevitt et al., 1997; Takahashi et al., 1998), this will be an important area for further investigation in the future.

Our findings also demonstrate how by reducing the overall levels of GATA1, a distinct defect in human hematopoiesis can emerge. Importantly, the observed hematologic phenotype in the two patients described here is different from other cases of GATA1-mutated blood phenotypes. Missense mutations in GATA1 can result in other forms of dyserythropoiesis, thalassemia, and

a variety of thrombopoietic defects (Crispino and Horwitz, 2017). Lack of the full-length form of GATA1 with continued production of the short isoform can cause Diamond-Blackfan anemia (DBA; Sankaran et al., 2012). We have recently found that impaired translation of GATA1 in early hematopoietic progenitors underlies the most commonly observed cases of DBA due to ribosomal protein mutations (Ludwig et al., 2014; Khajuria et al., 2018). Moreover, such aberrant translation of GATA1 may also occur in other blood disorders, such as myelofibrosis (Gilles et al., 2017). The results we describe in this paper extend the spectrum of GATA1-related disease and show how distinct alleles in a single key regulator of hematopoiesis can cause pleiotropic phenotypes, illuminating the numerous functions of GATA1 in human hematopoiesis.

Materials and methods

Patients and family

Patient 1 was a boy born at full term to healthy unrelated parents from Bulgaria (mother) and Togo (father). On the first day of life, anemia with a hemoglobin (Hb) of 9.4 g/dL and concomitant jaundice with elevated lactate dehydrogenase were noted. Besides second-degree hypospadias, no syndromic features were present. The patient required transfusions at 6 wk and 3 mo of age because of Hb decreases to the 5-g/dL range with an inadequate reticulocyte response and lower gastrointestinal bleeding. Subsequently, the child maintained a stable Hb of 8–10 g/dL with persistent red cell macrocytosis and a few giant platelets. At the age of 3.5 yr, a respiratory tract infection with febrile seizure occurred with transiently worsened anemia (Hb 6.3 g/dL, reticulocytes at 18%) and a mild thrombocytopenia (133,000 cells/ μ l). The patient had an elevated lactate dehydrogenase with no signs of active hemolysis. Extensive evaluations performed during and between the acute episodes for nonimmune hemolytic anemias; infectious, metabolic, and autoimmune disorders; and bone marrow failure syndromes (including mitomycin C-induced DNA damage testing and telomere length measurements) were unrevealing. An erythrocyte adenosine deaminase level was elevated at 2.79 U/g Hb. Multiple bone marrow examinations revealed normal cellularity with signs of dyserythropoiesis with macrocytic and occasional megaloblastic differentiation, mild dysplasia of the megakaryocytic lineage with hypolobated forms, and hypogranulated neutrophils. During infection-free periods, the Hb remained stable at levels of 9–10 g/dl with macrocytosis (MCV ~100 fl) and a persistently elevated fetal Hb (~20–23%), while the platelet counts remained in the low normal range. At the age of 7 yr, when presenting with *Mycoplasma pneumoniae*, there was severe epistaxis and moderate thrombocytopenia (45,000 cells/ μ l), as well as platelet function abnormalities that persisted even after the platelet counts normalized. Platelet agglutination/aggregation was severely impaired after stimulation with ristocetin, collagen, ADP, and epinephrine. Severely decreased expression of CD62 and CD63, respectively, were consistent with a α -granule and δ -granule secretion defect. The family medical history was unremarkable with the exception of a maternal first cousin with a chronic anemia of unclear etiology. When examined, the mother

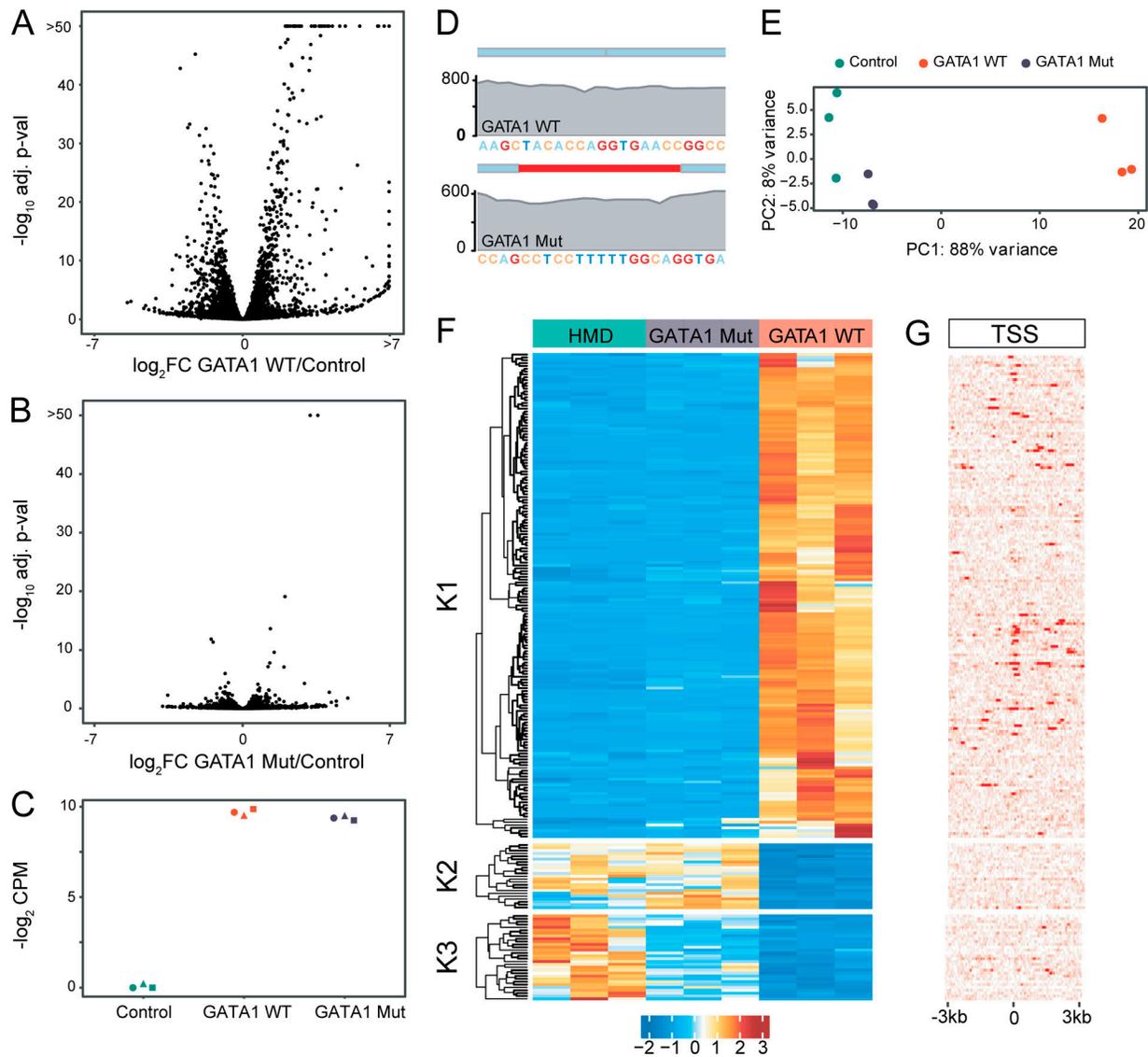


Figure 5. GATA1 variant shows complete loss of transcriptional activity. (A) Volcano plot showing \log_2 fold change (FC) of differentially expressed genes between GATA1 wild type–infected and control cells at indicated P values. (B) Volcano plot showing \log_2 FC of differential genes between GATA1 mutant infected and control cells at indicated P values. (C) Dot plot showing equivalent expression of GATA1 wild type and mutant based on RNA-seq for three biological replicates per condition. cpm, counts per million. (D) RNA-seq coverage of the region containing the intron event showing GATA1 wild-type cDNA and GATA1 mutant cDNA (biological replicate 1 for each condition shown). (E) Principal component plot of RNA-seq replicates (symbol) by condition (color) showing PC1 and PC2 and the percent variance explained by each. This reveals that the GATA1 mutant clusters with the control, while the GATA1 wild type clusters separately. (F) Heatmap showing clustering for the top genes ($n = 232$) differentially expressed between GATA1 wild type and empty HMD control ($\log_2\text{FC} > 2$, $p_{\text{adj}} < 0.01$), which shows three distinct clusters. (G) Depiction of GATA1 chromatin occupancy in 60-bp bins in a 6-kb window surrounding the TSS for the principal isoform of the indicated genes as in F shows an enrichment in cluster K1 genes. The darker red color in this plot indicates greater chromatin occupancy by GATA1 in the G1E cells. For RNA-seq experiments, three biological replicates shown unless otherwise indicated.

was noted to have normal peripheral blood counts. Only limited samples from the healthy mother could be obtained.

Patient 2 was a boy born to unrelated parents of African and Asian ethnicity. The child was diagnosed antenatally with severe anemia (Hb 2.4 g/dL) leading to hydrops fetalis, which was treated with intrauterine transfusion at 22 wk gestation with good response. The child was born by normal vaginal delivery at 40 wk gestation with no obvious dysmorphic features. The family history was unremarkable with the exception of a maternal first cousin with thalassemia trait. The patient required

two further red cell transfusions at week 4 and week 8 of life for anemia (Hb 6–7 g/dL with normal WBC and platelet counts). Subsequently, he maintained a stable Hb of 8–10 g/dL and a raised MCV (~100 fl), with transient worsening of anemia associated with febrile illnesses. A bone marrow examination performed at 18 mo of age showed normal cellularity with mild trilineage dysplasia with signs of dyserythropoiesis and small dysplastic megakaryocytes. Investigations performed for hemolytic anemias, infections, and bone marrow failure syndromes (diepoxybutane chromosome fragility test, RPS19

mutation analysis, MDS and juvenile myelomonocytic leukemia mutation/cytogenetic screen) were unremarkable. He had marginally elevated HbF (6.5%) and erythrocyte adenocytase deaminase (123) levels. His platelet counts remained in the low normal range with large platelets seen on the blood smear, and because of a history of prolonged bleeding after circumcision and recurrent mild epistaxis associated with respiratory infections, platelet function tests were done at the age of 6 yr. This showed abnormal platelet function with reduced aggregation after stimulation with low-dose collagen (1 $\mu\text{g}/\text{ml}$) and ADP (2 μM and 5 μM). ADP release was reduced on nucleotide studies, hinting at a δ -granule secretion defect. A repeat bone marrow examination done at the age of 7 yr because of falling platelet counts (lowest 113,000 cells/ μl) showed no new changes. Subsequently his Hb and platelet counts have remained stable at 9–10 g/dL and 115,000–160,000 cells/ μl , respectively. He remains clinically well with occasional mild epistaxis and bruising after injury. The mother of the patient had a reportedly normal peripheral blood count, but we were unable to obtain samples from this individual for further study.

Study approval

All family members provided written informed consent to participate in this study. The institutional review boards of Boston Children's Hospital, Massachusetts Institute of Technology, the University of Michigan, and the University of Freiburg approved the study protocols.

Platelet aggregometry and flow cytometry

Platelet agglutination/aggregation were analyzed using the following agonists: ristocetin (1.2 mg/ml), collagen (2.0 $\mu\text{g}/\text{ml}$), ADP (4.0 $\mu\text{mol}/\text{l}$), and epinephrine (8 $\mu\text{mol}/\text{l}$), as described previously (Born and Cross, 1963). For flow cytometric analyses, diluted platelet rich plasma (5×10^7 platelets/ml) was stimulated with different concentrations of thrombin (0.025–1.0 U/ml) in the presence of 1.25 mM Gly-Pro-Arg-Pro. Platelets were stained by monoclonal anti-CD62P (CLB-thromb/6-FITC) and anti-CD63 (CLB-gran/12-FITC) antibodies and analyzed by flow cytometry as previously described (Lahav et al., 2002).

WES and related genetic analyses

The patients described in this paper are part of a rare blood disorder cohort that has been studied through the use of WES. WES was performed as previously described (Polfus et al., 2016; Kim et al., 2017; Khajuria et al., 2018; Ulirsch et al., 2018). WES in these cases was performed using genomic DNA obtained from peripheral blood samples of the patients. The resultant variant call file (in hg19 coordinates) was annotated with VEP v89 (McLaren et al., 2016) and rare variants (based on ExAC v0.3.1 and GnomAD r2.0.2; Lek et al., 2016; <http://gnomad.broadinstitute.org/>) were identified using a combination of the Genome Analysis Toolkit, Bcftools, and Gemini (McKenna et al., 2010; Li, 2011; Paila et al., 2013). No rare (<0.01% allele frequency in ExAC and GnomAD) loss-of-function or missense variants were identified in any known red blood cell disorder genes (ANK1, SPTB, SPTA1, SLC4A1, EPB42, EPB41, PIEZO1, KCNN4, GLUT1, G6PD, PKLR, NT5C3A, HK1, GPI, PGK1, ALDOA, TPII, PFKM,

ALAS2, FECH, UROS, CDAN1, SEC23B, KIF23, KLF1, GATA1, HBB, HBA1, HBA2, RPS7, RPS10, RPS15A, RPS17, RPS19, RPS20, RPS24, RPS26, RPS27, RPS28, RPS29, RPS31, RPL5, RPL11, RPL15, RPL18, RPL26, RPL27, RPL35, RPL35A, TSR2, and EPO), including those known to cause dyserythropoietic anemia. We next expanded our search to include all rare variants in these genes, and noted chrX: 48,652,176 C>T in the fifth intron of GATA1 harbored by these two patients, but absent from ExAC and GnomAD. The variant was hemizygous in both patients and carried in both mothers, suggesting a model of complete penetrance (the mother of patient 1 was studied by exome sequencing, while the mother of patient 2 had sequence analysis by Sanger sequencing). All mutations were confirmed from genomic DNA samples of the patients or family members by Sanger sequencing.

Minigene assay and splicing analysis

The GATA1 region flanking exon 5 and exon 6 was PCR amplified from genomic DNA from patient 1 and from a healthy unrelated individual with the addition of XhoI and NotI restriction enzyme sites (primers are as follows: GATA1 minigene forward, 5'-ATC ATCCTCGAGTCTTGGGTCTCTGACATC-3' and GATA1 minigene reverse, 5'-ATCATCGCGGCCGCCACATGGTCACACATT GCAG-3') for cloning into the pSpliceExpress vector (Addgene; Kishore et al., 2008). The constructs were transfected into human embryonic kidney 293T cells (ATCC) with FuGENE (Promega). After reverse transcription of RNA obtained 48 h after transfection using an RNAeasy Mini kit (Qiagen), RT-PCR was performed, and amplified fragments were cloned using the TOPO TA Kit (Thermo Fisher Scientific) to enable confirmation of sequences through Sanger sequencing. For RT-PCR analyses, the segment of interest in GATA1 was amplified with the following primers (GATA1 Exon 5, 5'-AGTGGGGATCCCGTGTG-3' and GATA1 Exon 6, 5'-ATCCTTCCGCATGGTCAGT-3'). Amplified products were linearized with IX TBE-Urea (Bio-Rad), incubated for 10 min at 95°C, run on a 10% TBE-Urea Gel (Bio-Rad), and subsequently visualized on a GelDoc instrument (Bio-Rad) after 15 min of staining in ethidium bromide/1 \times TBE buffer. For semiquantitative analyses, bands were quantified from scanned images using ImageJ software (National Institutes of Health).

Cell culture and lentiviral transduction

Human primary adult bone marrow-derived CD34⁺ HSPCs were obtained from Fred Hutchinson Cancer Research Institute and cultured as previously described (Hu et al., 2013; Ludwig et al., 2014). The wild-type or mutated (involving a five-amino acid insertion corresponding with the observed intron retention event) GATA1 cDNA constructs were cloned into the HMD vector and transfected into 293T cells for lentiviral production, along with helper plasmids, as described previously (Ludwig et al., 2014). The primary HSPCs undergoing erythroid differentiation were infected with lentiviruses on day 2 of differentiation. Multiple donors were used in different experiments, which all yielded similar results. On day 5 of differentiation, cells were sorted for GFP and maintained in culture for further terminal erythroid differentiation (Ulirsch et al., 2016). Differentiating cells were collected on days 7, 9, and 12 for flow

cytometric analysis with erythroid markers CD235a (HIR2; BD Biosciences), CD71 (OKT9; Invitrogen), and CD49d (9F10 or MZ18-24A9; BioLegend; Miltenyi) using a Fortessa Instrument (BD Biosciences). Cells were gated on live events using PI. G1E cells were kindly provided by the laboratory of Dr. Mitchell Weiss (St. Jude Children's Research Hospital, Memphis, TN) and cultured with 15% FBS, 1-thioglycerol, murine stem cell factor (PeproTech), and human erythropoietin (Amgen). G1E cells were infected with the lentiviruses discussed above and were either subjected to flow cytometric analysis with the mouse erythroid marker Ter119 and GFP using an Accuri cytometer (BD Biosciences) or sorted and subjected to downstream analyses. Lentiviruses expressing SF3B1 wild-type or K700E mutants, as described (Obeng et al., 2016), were used to infect 293T cells 1 d before transfection, as above for the minigene assay. Cytocentrifugation was performed with ~100,000–200,000 cells using a Shandon Cytospin 4 onto polylysine-coated slides at 300× RPM for 4 min. After drying, samples were stained with May-Grünwald for 5 min and then with Giemsa for 15 min, as previously described (Fiorini et al., 2017). Images were taken using a Nikon Eclipse E800 microscope and Nikon Elements Software.

Western blotting

Approximately 1 million G1E cells were collected 72 h after infection and lysed in radioimmunoprecipitation assay buffer supplemented with protease and phosphatase inhibitor cocktails (Santa Cruz). Proteins were quantified according to the DC Protein assay (Bio-Rad), run using a Mini-Protean TGX gel system (Bio-Rad) and transferred onto a polyvinylidene fluoride membrane. Signal was detected by ECL Amersham Hyperfilm (GE Healthcare). Western blotting was performed using GATA1 (M20) and GAPDH (6C5) primary antibodies at 1:1,000 dilution (Santa Cruz). An HRP-conjugated goat anti-mouse was used at 1:20,000 dilution as a secondary antibody (Bio-Rad). For the 293T cells, a GAPDH antibody (0411; Santa Cruz) was used.

G1E RNA-seq

Cells were lysed in RLT lysis buffer (Qiagen) supplemented with β -mercaptoethanol, and RNA was isolated using a RNeasy Mini Plus kit (Qiagen) according to the manufacturer's instructions. RNA was quantified using NanoDrop OneC (Thermo Fisher Scientific). 30 ng of RNA was used as input to a modified SMART-seq2 (Picelli et al., 2014) protocol, and after reverse transcription, eight cycles of PCR were used to amplify the transcriptome library. The quality of whole-transcriptome libraries was validated using a High Sensitivity DNA Chip run on a Bioanalyzer 2100 system (Agilent), followed by library preparation using the Nextera XT kit (Illumina) and custom index primers according to the manufacturer's instructions. Final libraries were quantified using a Qubit dsDNA HS Assay kit (Invitrogen) and a High Sensitivity DNA chip run on a Bioanalyzer 2100 system.

Sequencing

All libraries were sequenced using Nextseq High Output Cartridge kits and a Nextseq 500 sequencer (Illumina). Libraries

were consistently sequenced paired end (2×38 bp) and demultiplexed using the bcl2fastq program.

Alignment and differential expression

For each library, raw fastq reads were aligned using STAR version 2.5.1b (RNA-seq; Dobin et al., 2013) using the 2-Pass mode to the GRCh38.p4 reference genome, and gene expression counts were generated using quantMode in STAR. Human GATA1 cDNA (wild type and mutant) in the HMD vector, as described above, were appended as an additional chromosome. Differential expression was performed using the DESeq2 method (Love et al., 2014), and gene-wise statistical significance was determined using independent hypothesis weighting of 0.01 (Ignatiadis et al., 2016).

Clustering

A consistent set of features for the RNA-seq data were established by choosing the top differentially expressed genes between control (HMD) and GATA1 wild-type G1E cells ($p_{\text{adj}} < 0.01$, $\log_2\text{FC} > 2$, $n = 223$). A top genes by sample matrix of $\log_2\text{CPM}$ values was created and transformed to Z scores. The optimal cluster number (3) was determined by NbClust, computed by K-means with three centers, and Z scores were plotted using ComplexHeatmap R package (Gu et al., 2016).

Principal components analysis

Principal components of the G1E RNA-seq experiment were computed using the rlog transformation of the gene count matrix and plotted using the plotPCA function of the DESeq2 package.

ChIP-seq analysis

ChIP-seq data were obtained from the NCBI Gene Expression Omnibus (GEO; accession no. GSE36029) and aligned using Bowtie2 version 2.3.3 (Langmead and Salzberg, 2012). The 30-h 17 β -estradiol-treated G1E-ER4 cell sample (GSM995448) was used, as it is most representative of our experiment. The GATA1 binding profiles were determined for the top genes determined in the clustering analysis in a ± 3 -kb window around the canonical isoform TSSs, and Z scores were plotted as a heatmap using Complex Heatmap.

Human erythropoiesis RNA-seq analysis

Single-end RNA-seq datasets for five distinct maturation stages of primary human erythroid cells in biological triplicates were obtained from the NCBI GEO database (accession no. GSE53983; An et al., 2014). These datasets were aligned to the Ensembl GRCh37 r75 genome and gene annotations using 2-Pass STAR alignment. Resultant alignments were imported into R and visualized with Gviz.

Analysis of MDS patient samples

Cryopreserved bone marrow mononuclear cells from MDS patients with refractory anemia with ringed sideroblasts and healthy controls were thawed, washed, and recovered before staining with fluorescently conjugated antibodies targeting CD235a (HIR2), CD71 (OKT9), and CD34 (4H11). SF3B1 mutations

had been analyzed and confirmed in other investigations on these samples but were not specifically done for this study. Flow cytometric cell sorting for CD235a⁺CD71⁺ and CD34⁺ was performed using a FACSaria II Cell Sorter (BD Biosciences). After sorting, the cells were washed, and RNA was isolated using the Ambion small RNA extraction kit (Zymo). cDNA was synthesized from the recovered RNA samples (20–40 ng) using the iScript cDNA Kit (Bio-Rad). RT-PCR was performed with GATA1 primers and normalization controls (primers GATA1 Exon 5, 5'-AGTGGGGATCCCGTGTG-3', GATA1 Exon 6, 5'-ATCCTTCCGCATGGTCAGT-3', GAPDH Exon 3, 5'-CACCAGGGCTGCTTTTAACT-3', and GAPDH Exon 4, 5'-GACAAGCTTCCCGTTCTCAG-3'). Bands were visualized on denaturing gels and quantified as noted above. Samples were also analyzed with quantitative RT-PCR for GATA1 using a CFX96 Real Time Thermocycler (Bio-Rad) with SYBR Green (Bio-Rad; primers GATA1 Exon 2, 5'-CCCAGTTTTGTGGATCCT-3' and GATA1 Exon 3, 5'-CACAGTTGAGGCAGGGTAGAG-3', human ACTB Exon 3, 5'-AGAAAATCTGGCACCACACC-3', and human ACTB Exon 4, 5'-GGGGTGTGAAGGTCTCAA-3').

Statistical analysis

All pairwise comparisons were performed using the two-tailed Student's *t* test, unless otherwise indicated. Differences were considered significant if the *P* value was <0.05.

Data availability

The WES data are available in the dbGaP database (<http://www.ncbi.nlm.nih.gov/gap>) under the accession no. phs000474. RNA-seq data are available in the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) under accession no. GSE124830.

Online supplemental material

Fig. S1 shows clinical phenotypes and WES data for affected patients. Fig. S2 shows the GATA1 mRNA levels of SF3B1 MDS patients in sorted marrow populations. Fig. S3 shows cyto-centrifuge analysis of G1E cells expressing wild-type or mutated GATA1.

Acknowledgments

We are grateful to the patients and families for their interest and willingness to participate in this study. We thank members of the Sankaran and Wlodarski laboratories, as well as S. Orkin and D. Nathan, for valuable comments and assistance. We thank the ENCODE Consortium and specifically the Hardison laboratory for generating the GATA1 ChIP-seq data in G1E-ER4 cells and providing it to the community as a public resource.

This work was supported by National Institutes of Health grants R01 DK103794 (to V.G. Sankaran), R33 HL120791 (to V.G. Sankaran), R01 HL107558 (to H.T. Gazda), and K02 HL111156 (to H.T. Gazda), as well as the New York Stem Cell Foundation (to V.G. Sankaran). V.G. Sankaran is a New York Stem Cell Foundation-Robertson Investigator.

The authors declare no competing financial interests.

Author contributions: N.J. Abdulhay, C. Fiorini, J.M. Verboon, and L.S. Ludwig performed the research, analyzed data, and

wrote the manuscript. J.C. Ulirsch, C.A. Lareau, and X. Mi performed the research and analyzed data. B. Zieger, A. Roy, E.A. Obeng, M. Erlacher, N. Gupta, S.B. Gabriel, B.L. Ebert, C.M. Niemeyer, R.N. Khoriaty, P. Ancliff, H.T. Gazda, and M.W. Wlodarski provided clinical data and experimental reagents. V. G. Sankaran supervised the research, analyzed data, and wrote the manuscript. All authors edited the manuscript.

Submitted: 23 August 2018

Revised: 11 January 2019

Accepted: 6 March 2019

References

- An, X., V.P. Schulz, J. Li, K. Wu, J. Liu, F. Xue, J. Hu, N. Mohandas, and P.G. Gallagher. 2014. Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood*. 123:3466–3477. <https://doi.org/10.1182/blood-2014-01-548305>
- Basak, A., M. Hancarova, J.C. Ulirsch, T.B. Balci, M. Trkova, M. Pelisek, M. Vlckova, K. Muzikova, J. Cermak, J. Trka, et al. 2015. BCL11A deletions result in fetal hemoglobin persistence and neurodevelopmental alterations. *J. Clin. Invest.* 125:2363–2368. <https://doi.org/10.1172/JCI81163>
- Born, G.V., and M.J. Cross. 1963. The Aggregation of Blood Platelets. *J. Physiol.* 168:178–195. <https://doi.org/10.1113/jphysiol.1963.sp007185>
- Campbell, A.E., L. Wilkinson-White, J.P. Mackay, J.M. Matthews, and G.A. Blobel. 2013. Analysis of disease-causing GATA1 mutations in murine gene complementation systems. *Blood*. 121:5218–5227. <https://doi.org/10.1182/blood-2013-03-488080>
- Casanova, J.L., M.E. Conley, S.J. Seligman, L. Abel, and L.D. Notarangelo. 2014. Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies. *J. Exp. Med.* 211:2137–2149. <https://doi.org/10.1084/jem.20140520>
- Crispino, J.D., and M.S. Horwitz. 2017. GATA factor mutations in hematologic disease. *Blood*. 129:2103–2110. <https://doi.org/10.1182/blood-2016-09-687889>
- Cummings, B.B., J.L. Marshall, T. Tukiainen, M. Lek, S. Donkervoort, A.R. Foley, V. Bolduc, L.B. Waddell, S.A. Sandaradura, G.L. O'Grady, et al.; Genotype-Tissue Expression Consortium. 2017. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* 9:eal5209. <https://doi.org/10.1126/scitranslmed.aal5209>
- DeBoever, C., E.M. Ghia, P.J. Shepard, L. Rassenti, C.L. Barrett, K. Jepsen, C.H. Jamieson, D. Carson, T.J. Kipps, and K.A. Frazer. 2015. Transcriptome sequencing reveals potential mechanism of cryptic 3' splice site selection in SF3B1-mutated cancers. *PLOS Comput. Biol.* 11:e1004105. <https://doi.org/10.1371/journal.pcbi.1004105>
- Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 489:57–74. <https://doi.org/10.1038/nature11247>
- Fiorini, C., N.J. Abdulhay, S.K. McFarland, M. Munschauer, J.C. Ulirsch, R. Chiarle, and V.G. Sankaran. 2017. Developmentally-faithful and effective human erythropoiesis in immunodeficient and Kit mutant mice. *Am. J. Hematol.* 92:E513–E519. <https://doi.org/10.1002/ajh.24805>
- Frisan, E., J. Vandekerckhove, A. de Thonel, C. Pierre-Eugène, A. Sternberg, J. B. Arlet, C. Floquet, E. Gyan, O. Kosmider, F. Dreyfus, et al. 2012. Defective nuclear localization of Hsp70 is associated with dyserythropoiesis and GATA-1 cleavage in myelodysplastic syndromes. *Blood*. 119:1532–1542. <https://doi.org/10.1182/blood-2011-03-343475>
- Gasparini, M., G.M. Findlay, A. McKenna, J.H. Milbank, C. Lee, M.D. Zhang, D.A. Cusanovich, and J. Shendure. 2017. CRISPR/Cas9-Mediated Scanning for Regulatory Elements Required for HPRT1 Expression via Thousands of Large, Programmed Genomic Deletions. *Am. J. Hum. Genet.* 101:192–205. <https://doi.org/10.1016/j.ajhg.2017.06.010>
- Giani, F.C., C. Fiorini, A. Wakabayashi, L.S. Ludwig, R.M. Salem, C.D. Jobaliya, S.N. Regan, J.C. Ulirsch, G. Liang, O. Steinberg-Shemer, et al. 2016. Targeted Application of Human Genetic Variation Can Improve Red Blood Cell Production from Stem Cells. *Cell Stem Cell*. 18:73–78. <https://doi.org/10.1016/j.stem.2015.09.015>

- Gilles, L., A.D. Arslan, C. Marinaccio, Q.J. Wen, P. Arya, M. McNulty, Q. Yang, J.C. Zhao, K. Konstantinoff, T. Lasho, et al. 2017. Downregulation of GATA1 drives impaired hematopoiesis in primary myelofibrosis. *J. Clin. Invest.* 127:1316–1320. <https://doi.org/10.1172/JCI82905>
- Gu, Z., R. Eils, and M. Schlesner. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 32: 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313>
- Hu, J., J. Liu, F. Xue, G. Halverson, M. Reid, A. Guo, L. Chen, A. Raza, N. Galili, J. Jaffray, et al. 2013. Isolation and functional characterization of human erythroblasts at distinct stages: implications for understanding of normal and disordered erythropoiesis in vivo. *Blood.* 121:3246–3253. <https://doi.org/10.1182/blood-2013-01-476390>
- Ignatiadis, N., B. Klaus, J.B. Zaugg, and W. Huber. 2016. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods.* 13:577–580. <https://doi.org/10.1038/nmeth.3885>
- Khajuria, R.K., M. Munschauer, J.C. Ulirsch, C. Fiorini, L.S. Ludwig, S.K. McFarland, N.J. Abdulhay, H. Specht, H. Keshishian, D.R. Mani, et al. 2018. Ribosome levels selectively regulate translation and lineage commitment in human hematopoiesis. *Cell.* 173:90–103.e119.
- Kim, A.R., J.C. Ulirsch, S. Wilmes, E. Unal, I. Moraga, M. Karakukcu, D. Yuan, S. Kazerounian, N.J. Abdulhay, D.S. King, et al. 2017. Functional selectivity in cytokine signaling revealed through a pathogenic EPO mutation. *Cell.* 168:1053–1064.e1015.
- Kishore, S., A. Khanna, and S. Stamm. 2008. Rapid generation of splicing reporters with pSpliceExpress. *Gene.* 427:104–110. <https://doi.org/10.1016/j.gene.2008.09.021>
- Kuehn, H.S., B. Boisson, C. Cunningham-Rundles, J. Reichenbach, A. Stray-Pedersen, E.W. Gelfand, P. Maffucci, K.R. Pierce, J.K. Abbott, K.V. Voelkerding, et al. 2016. Loss of B Cells in Patients with Heterozygous Mutations in IKAROS. *N. Engl. J. Med.* 374:1032–1043. <https://doi.org/10.1056/NEJMoal512234>
- Lhav, J., K. Jurk, O. Hess, M.J. Barnes, R.W. Farndale, J. Luboshitz, and B.E. Kehrel. 2002. Sustained integrin ligation involves extracellular free sulfhydryls and enzymatically catalyzed disulfide exchange. *Blood.* 100: 2472–2478. <https://doi.org/10.1182/blood-2001-12-0339>
- Langmead, B., and S.L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
- Lek, M., K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J.S. Ware, A.J. Hill, B.B. Cummings, et al.; Exome Aggregation Consortium. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 536:285–291. <https://doi.org/10.1038/nature19057>
- Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 27:2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15: 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Ludwig, L.S., H.T. Gazda, J.C. Eng, S.W. Eichhorn, P. Thiru, R. Ghazvinian, T.I. George, J.R. Gotlib, A.H. Beggs, C.A. Steff, et al. 2014. Altered translation of GATA1 in Diamond-Blackfan anemia. *Nat. Med.* 20:748–753. <https://doi.org/10.1038/nm.3557>
- McDevitt, M.A., R.A. Shivdasani, Y. Fujiwara, H. Yang, and S.H. Orkin. 1997. A “knockdown” mutation created by cis-element gene targeting reveals the dependence of erythroid cell maturation on the level of transcription factor GATA-1. *Proc. Natl. Acad. Sci. USA.* 94:6781–6785. <https://doi.org/10.1073/pnas.94.13.6781>
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytzky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M.A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>
- McLaren, W., L. Gil, S.E. Hunt, H.S. Riat, G.R. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. 2016. The Ensembl Variant Effect Predictor. *Genome Biol.* 17:122. <https://doi.org/10.1186/s13059-016-0974-4>
- Obeng, E.A., R.J. Chappell, M. Seiler, M.C. Chen, D.R. Campagna, P.J. Schmidt, R.K. Schneider, A.M. Lord, L. Wang, R.G. Gambe, et al. 2016. Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. *Cancer Cell.* 30:404–417. <https://doi.org/10.1016/j.ccell.2016.08.006>
- Paila, U., B.A. Chapman, R. Kirchner, and A.R. Quinlan. 2013. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput. Biol.* 9:e1003153. <https://doi.org/10.1371/journal.pcbi.1003153>
- Picelli, S., O.R. Faridani, A.K. Björklund, G. Winberg, S. Sagasser, and R. Sandberg. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9:171–181. <https://doi.org/10.1038/nprot.2014.006>
- Polfus, L.M., R.K. Khajuria, U.M. Schick, N. Pankratz, R. Pazoki, J.A. Brody, M.H. Chen, P.L. Auer, J.S. Floyd, J. Huang, et al. 2016. Whole-Exome Sequencing Identifies Loci Associated with Blood Cell Traits and Reveals a Role for Alternative GFIIB Splice Variants in Human Hematopoiesis. *Am. J. Hum. Genet.* 99:481–488. <https://doi.org/10.1016/j.ajhg.2016.06.016>
- Rosenberg, A.B., R.P. Patwardhan, J. Shendure, and G. Seelig. 2015. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell.* 163:698–711. <https://doi.org/10.1016/j.cell.2015.09.054>
- Rylski, M., J.J. Welch, Y.Y. Chen, D.L. Letting, J.A. Diehl, L.A. Chodosh, G.A. Blobel, and M.J. Weiss. 2003. GATA-1-mediated proliferation arrest during erythroid maturation. *Mol. Cell. Biol.* 23:5031–5042. <https://doi.org/10.1128/MCB.23.14.5031-5042.2003>
- Sankaran, V.G., R. Ghazvinian, R. Do, P. Thiru, J.A. Vergilio, A.H. Beggs, C.A. Steff, S.H. Orkin, D.G. Nathan, E.S. Lander, and H.T. Gazda. 2012. Exome sequencing identifies GATA1 mutations resulting in Diamond-Blackfan anemia. *J. Clin. Invest.* 122:2439–2443. <https://doi.org/10.1172/JCI63597>
- Sperling, A.S., C.J. Gibson, and B.L. Ebert. 2017. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat. Rev. Cancer.* 17:5–19. <https://doi.org/10.1038/nrc.2016.112>
- Takahashi, S., T. Komeno, N. Suwabe, K. Yoh, O. Nakajima, S. Nishimura, T. Kuroha, T. Nagasawa, and M. Yamamoto. 1998. Role of GATA-1 in proliferation and differentiation of definitive erythroid and megakaryocytic cells in vivo. *Blood.* 92:434–442.
- Ulirsch, J.C., S.K. Nandakumar, L. Wang, F.C. Giani, X. Zhang, P. Rogov, A. Melnikov, P. McDonel, R. Do, T.S. Mikkelsen, and V.G. Sankaran. 2016. Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. *Cell.* 165:1530–1545. <https://doi.org/10.1016/j.cell.2016.04.048>
- Ulirsch, J.C., J.M. Verboon, S. Kazerounian, M.H. Guo, D. Yuan, L.S. Ludwig, R.E. Handsaker, N.J. Abdulhay, C. Fiorini, G. Genovese, et al. 2018. The Genetic Landscape of Diamond-Blackfan Anemia. *Am. J. Hum. Genet.* 103:930–947. <https://doi.org/10.1016/j.ajhg.2018.10.027>
- Wakabayashi, A., J.C. Ulirsch, L.S. Ludwig, C. Fiorini, M. Yasuda, A. Choudhuri, P. McDonel, L.I. Zon, and V.G. Sankaran. 2016. Insight into GATA1 transcriptional activity through interrogation of cis elements disrupted in human erythroid disorders. *Proc. Natl. Acad. Sci. USA.* 113:4434–4439. <https://doi.org/10.1073/pnas.1521754113>
- Yoshida, K., M. Sanada, Y. Shiraishi, D. Nowak, Y. Nagata, R. Yamamoto, Y. Sato, A. Sato-Otsubo, A. Kon, M. Nagasaki, et al. 2011. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature.* 478:64–69. <https://doi.org/10.1038/nature10496>
- Zmajkovic, J., P. Lundberg, R. Nienhold, M.L. Torgersen, A. Sundan, A. Waage, and R.C. Skoda. 2018. A Gain-of-Function Mutation in EPO in Familial Erythrocytosis. *N. Engl. J. Med.* 378:924–930. <https://doi.org/10.1056/NEJMoal709064>