Published online 7 October 2016

Nucleic Acids Research, 2017, Vol. 45, No. 2 ***657–671***
doi: 10.1093/nar/gkw901

Intron 1 GATA site enhances ALAS2 expression indispensably during erythroid differentiation

Yingchi Zhang[1,2,3,†], Jingliao Zhang[1,2,3,†], Wenbin An[1,2,3,†], Yang Wan[1,2,3], Shihui Ma[1,3], Jie Yin[4], Xichuan Li[5], Jie Gao[1,3], Weiping Yuan[1,3], Ye Guo[1,2,3], James Douglas Engel[6], Lihong Shi[1,3,*], Tao Cheng[1,3,7,*] and Xiaofan Zhu[1,2,3,*]

[1]State Key Laboratory of Experimental Hematology, Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China, [2]Division of Pediatric Blood Diseases Center, Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China, [3]Center for Stem Cell Medicine, Chinese Academy of Medical Sciences, Beijing 100730, China, [4]Department of Cell Biology, Tianjin Key Laboratory of Medical Epigenetics, Tianjin Medical University, Tianjin 300070, China, [5]Department of Immunology, Biochemistry and Molecular Biology, Tianjin Key Laboratory of Medical Epigenetics, Tianjin Medical University, Tianjin 300070, China, [6]Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA and [7]Department of Stem Cell & Regenerative Medicine, Peking Union Medical College, Tianjin 300020, China

Received July 8, 2016; Revised September 26, 2016; Accepted September 30, 2016

ABSTRACT

The first intronic mutations in the intron 1 GATA site (int-1-GATA) of 5-aminolevulinate synthase 2 (*ALAS2)* have been identified in X-linked sideroblastic anemia (XLSA) pedigrees, strongly suggesting it could be causal mutations of XLSA. However, the function of this int-1-GATA site during *in vivo* development remains largely unknown. Here, we generated mice lacking a 13 bp fragment, including this int-1-GATA site (TAGATAAAGCCCC) and found that hemizygous deletion led to an embryonic lethal phenotype due to severe anemia resulting from a lack of ALAS2 expression, indicating that this non-coding sequence is indispensable for ALAS2 expression *in vivo*. Further analyses revealed that this int-1-GATA site anchored the GATA site in intron 8 (int-8-GATA) and the proximal promoter, forming a long-range loop to enhance ALAS2 expression by an enhancer complex including GATA1, TAL1, LMO2, LDB1 and Pol II at least, in erythroid cells. However, compared with the int-8-GATA site, the int-1-GATA site is more essential for regulating ALAS2 expression through CRISPR/Cas9-mediated site-specific deletion. Therefore, the int-1-GATA site could serve as a valuable site for diagnosing XLSA in cases with unknown mutations.

INTRODUCTION

Congenital sideroblastic anemia (CSA) comprises a heterogeneous group of genetic disorders that are characterized by decreased heme biosynthesis and the presence of bone marrow (BM) ring sideroblasts in erythroid precursors and pathologic mitochondrial iron deposits (1). Seven types of CSA have thus far been identified at the molecular level; these disorders primarily involve heme biosynthesis, iron-sulfur [Fe-S] clusters and mitochondrial proteins, accounting for approximately 60% of the genetic defects in CSA (2,3). Among all genetically revealed CSA cases, the most common form is X-linked sideroblastic anemia (XLSA, OMIM#300751), a disease caused by mutations in the gene encoding 5-aminolevulinic acid synthase 2 (ALAS2), which was first identified in 1985 (4).

ALAS2 catalyzes the rate-limiting step of heme biosynthesis in erythroid cells (5). To date, at least 61 distinct pathogenic mutations have been identified in the catalytic domain encoded by exons 5–11 in over 100 unrelated individuals or families (6–8). In addition to many mutations identified in exons, mutations that occur within non-coding regions have also begun to attract attention.

Recently, novel non-coding mutations within the conserved GATA1 binding site in the first intron of *ALAS2* (int-1-GATA) were identified in seven XLSA pedigrees by two independent research groups (9,10). These findings strongly suggest that these int-1-GATA mutations comprise new causal XLSA mutations. One of these groups (10) further

*To whom correspondence should be addressed. Tel: +86 22 2390 9156, Fax: +86 2390 9156; Email: chengtao@ihcams.ac.cn
Correspondence may also be addressed to Lihong Shi. Tel: +86 22 2390 9448, Fax: +86 22 2390 9448; Email: shilihongxys@ihcams.ac.cn
Correspondence may also be addressed to Xiaofan Zhu. Tel: +86 22 2390 9001; Fax: +86 22 2390 9001; Email: xfzhu@ihcams.ac.cn
†These authors contributed equally to the paper as first authors.

demonstrated by an *in vitro* luciferase promoter assay that this int-1-GATA site might act as an enhancer that modulates ALAS2 expression. However, the relevance of this conserved *cis*-regulatory element *in vivo* has not yet been addressed.

GATA1 was the first cloned member of the GATA family (11–13), which consists of six structurally related transcription factors (GATA1-GATA6). In mice, *Gata1* deletion is embryonic lethal due to the development of severe anemia between embryonic day (E) 10.5–11.5 caused by the arrest of primitive erythropoiesis (14). Furthermore, *Gata1*-deficient mouse embryonic stem (ES) cells are unable to generate mature red blood cells during definitive erythropoiesis in chimeric mice (15). Therefore, GATA1 is a key transcription factor that regulates numerous erythroid-specific genes during erythropoiesis *in vivo* (16). In addition, ALAS2 expression is significantly reduced in GATA1 promoter-disrupted erythroid cells differentiated from ES cells *in vitro* (17) and in XLSA patients carrying these int-1-GATA mutations (9), indicating that *ALAS2* could be an important downstream target of GATA1.

GATA1 executes its key regulatory functions in erythroid cells via three functional domains: an N-terminal activation domain, an N-terminal zinc finger (N-finger) domain and a C-terminal zinc finger (C-finger) domain (16). The C-finger domain mediates sequence-specific DNA binding to (A/T)GATA(A/G) motifs (18,19), whereas the N-finger domain mediates important protein–protein interactions, such as with friend of GATA1 (FOG1) (20), erythroid Kruppel-like factor 1 (EKLF) (21), mediator complex subunit 1 (MED1) (22) and stem cell leukemia/T-cell acute lymphocytic leukemia-1 (Scl/TAL1), a master regulator of hematopoiesis that binds to E-boxes (23,24). When GATA1 interacts with TAL1, the complex recruits the non-DNA-binding components of LIM domain-binding protein 1 (LDB1) and LIM domain only 2 (LMO2), which are thought to mediate the long-range promoter/enhancer interactions by which GATA1 and TAL1 activate erythroid genes (14,23–28). Nevertheless, to date, direct *in vivo* evidence of the GATA1-mediated regulation of *ALAS2* activity in erythroid cells is absent.

In the present study, we coincidently identified the int-1-GATA mutation in a larger XLSA pedigree, as described previously (9,10) and employed the genome editing tool transcription activator-like effector nuclease (TALEN) to address the critical roles of this non-coding *cis*-regulatory element during *in vivo* development. We found that deletion of int-1-GATA and its flanking DNA fragment led to an anemia-induced embryonic lethality that phenocopied the *Alas2* null mutant mouse, demonstrating that the int-1-GATA site is an indispensable regulatory element for Alas2 expression. We then demonstrated the mechanism by which GATA1 activates *ALAS2* in erythroid cells. The int-1-GATA site acts as an anchor that links the GATA site in intron 8 (int-8-GATA) to the proximal promoter, forming a long-range loop to enhance ALAS2 expression via an enhancer complex that includes GATA1, TAL1, LMO2, LDB1, Pol II and possibly other proteins to fully activate *ALAS2* transcription.

MATERIALS AND METHODS

In this study, we assessed a large XLSA family pedigree in which six male patients were diagnosed (Figure 1A). DNA was isolated from the peripheral blood (PB) cells and oral epithelial cells of all members of this family for Sanger sequencing of the ALAS2 gene. The BMMCs of the patients (III 2, III 7 and III 9) and three healthy controls were prepared for RT-qPCR and western blotting. All patients and carriers involved in our study signed an informed consent form approved by the IRB of the Institute of Hematology & Blood Diseases Hospital, CAMS/PUMC (KT2013004-EC-1).

Luciferase reporter assay

The human *ALAS2* proximal promoter region (between 146 and +14 from the transcription start site) and intron 1 enhancer region (a 115 bp fragment containing *ALAS2* int-1-GATA), as shown in Figure 1E, were prepared from the genomic DNA of a healthy volunteer or patient III 7 and cloned into the multiple-cloning site of pGL3basic. These reporter vectors and pEF-RL were introduced into K562 cells. Luciferase activity was determined using the Dual-Luciferase reporter system (Promega, E1910).

Generation of TALEN-mediated *Alas2* int-1-GATA site knockout mice

TALEN repeats were designed to bind to the GATA1 binding region in intron 1 of the mouse *Alas2* gene (int-1-GATA) and fused to a FokI nuclease domain (Figure 2A). *In vitro*-transcribed Alas2 int-1-GATA-TALEN mRNAs were prepared and microinjected into fertilized eggs from C57BL/6 mice, which were transferred to pseudopregnant females. All mice were maintained in a specific pathogen-free facility. Newborn pups were genotyped by polymerase chain reaction (PCR) amplification and Sanger sequencing using tail-derived DNA. The sex of these embryos was determined using PCR amplification of the Y chromosome-specific *Zfy-1* gene. All animal protocols were approved by the Institutional Animal Care and Use Committee (IACUC), Institute of Hematology and Blood Disease Hospital, CAMS/PUMC. All surgeries were performed under sodium pentobarbital anesthesia, and all efforts were made to minimize suffering.

Generation of *ALAS2* int-1-GATA and int-8-GATA site deletion K562 cell lines using the CRISPR/Cas9 system

Guide sequences were designed to delete the GATA1 binding regions of intron 1 and intron 8 of the human *ALAS2* gene, as shown in Figure 7A and B. SgRNAs were cloned into the pX330 wild-type (WT) Cas9 vector, as previously described (29). Deletions of the int-1-GATA or int-8-GATA site were confirmed by PCR amplifications and Sanger sequencing. The sgRNAs used for cloning and the PCR primers are listed in Supplementary Table S1.

Mouse embryo preparation and flow cytometry

Single-cell suspensions of E11.5 yolk sac cells were prepared as previously described (30). For flow cytometric (FACS)

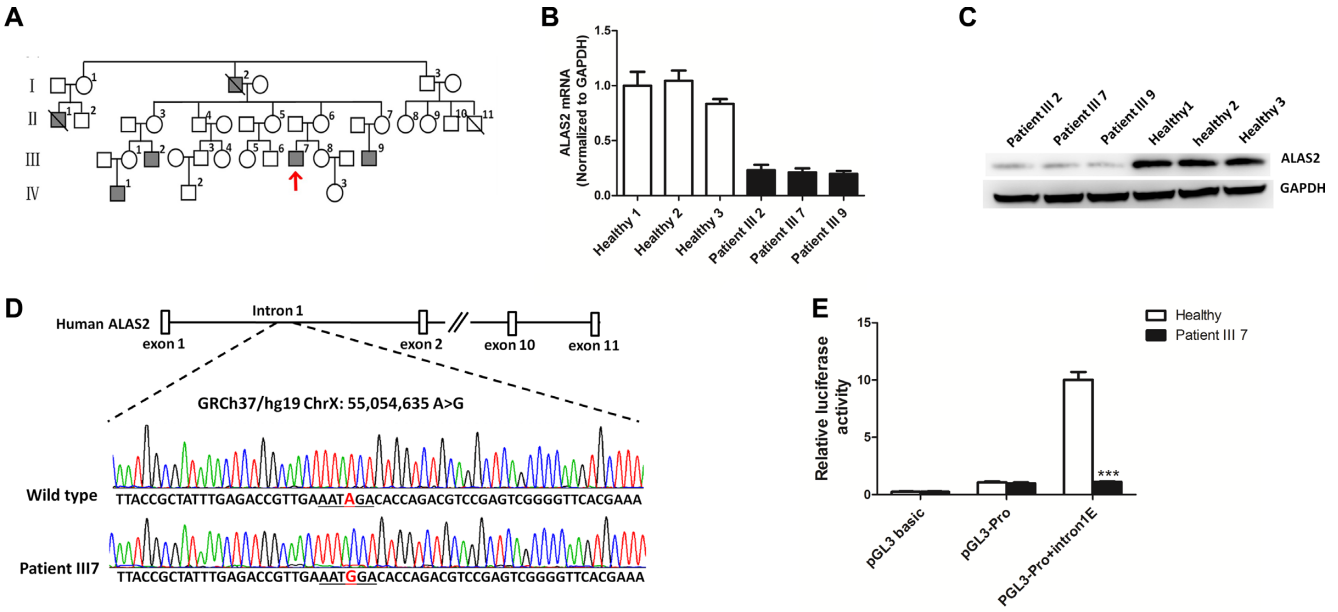


Figure 1. Identification of a GATA1 binding site mutation within the *ALAS2* intron 1 in an XLSA pedigree. (**A**) The XLSA family tree. Shaded boxes indicate affected individuals in this pedigree. The red arrow indicates the proband in this XLSA family. (**B** and **C**) The transcript (B) and protein (C) levels of ALAS2 were measured by RT-qPCR and western blotting, respectively, in patients III 2, III 7, III 9 and 3 healthy individuals. RNA and proteins were extracted from BMMCs in both groups. GAPDH transcripts (B) and protein (C) acted as reference gene and loading control for RT-qPCR and western blotting, respectively. (**D**) The proband harbors an A>G missense mutation (red bold) in the intron 1 positive strand. The putative GATA binding site is underlined. The reference sequence is GRCh37/hg19ChrX: 55054615–55054661. (**E**) The int-1-GATA site significantly promotes *ALAS2*-promoter driven transcription according to an *in vitro* luciferase assay. The proximal promoter region of the *ALAS2* gene from the genomes of normal family members and patient III7 was subcloned into pGL3 basic (abbreviated pGL3-Pro). The GATA1 binding region in the first intron of the *ALAS2* gene from the healthy and patient III7 genomes was subcloned into pGL3-Pro (abbreviated pGL3-Pro+intron 1E). Each of these reporter vectors was introduced individually into K562 cells. The results are expressed as relative activity compared with pGL3-Pro. $n = 3$ independent experiments. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

analysis, the single-cell suspensions were labeled with PE-Cy7-conjugated anti-Ter119 (eBioscience, 25–5921) and APC-conjugated anti-CD71 (eBioscience, 17–0711). Cell data were acquired using an LSRII flow cytometer (Becton Dickson). The data analysis was performed using FlowJo.

Colony-forming unit (CFU) assays

Single-cell suspensions of E11.5 yolk sac cells were prepared for colony-forming unit (CFU)-E and BFU-E assays using MethoCult complete medium (StemCell Technologies, M3334 and M3434). CFU-E colonies were scored on days 3–4, and BFU-E colonies were scored on days 8–10.

Ex vivo differentiation of purified human CD34^{+} cells

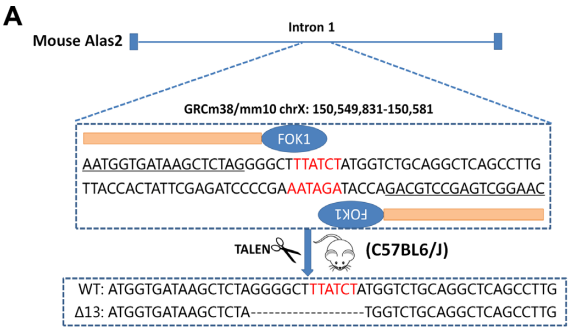
Human cord blood samples were obtained from the Tianjin Obstetric Central Hospital (Tianjin, China) under a protocol approved by the Ethical Committee on Medical Research at the Institute of Hematology. The CD34^{+} cells were cultured and differentiated *ex vivo* into the erythroid lineage within 18 days using a two-phase culture method, as previously described (31,32).

ChIP-qPCR assays

For ChIP-qPCR assays, cells differentiated from cord blood CD34^{+} cells on days 4, 8 and 11, int1Δ6 K562 cells, int8Δ K562 cells, WT K562 cells and yolk sac cells from E11.5 mice were fixed with 1% (vol/vol) formaldehyde for 10 min at room temperature. The ChIP assays were performed using a EZ-Magna ChIP A/G Kit (Millipore, 17–408). An anti-GATA1 antibody (Abcam, ab11852), anti-TAL1 antibody (Santa Cruz, sc12984), anti-LDB1 antibody (Santa Cruz, sc11198), anti-LMO2 (Abcam, ab72841), anti-FOG1 antibody (Santa Cruz, sc9361), anti-Pol II antibody (Abcam, ab817) and normal mouse and rabbit IgGs were used for immunoprecipitation. ChIP DNA fragments, which ranged from 100 to 500 bp according to agarose gel electrophoresis, were subjected to real-time PCR analysis using primers that targeted the promoter region, intron 1 enhancer region or intron 8 enhancer region of the human *ALAS2* gene. The primers used for the ChIP-qPCR analysis are listed in Supplementary Table S1.

Chromosome conformation capture (3C)

Chromosome conformation capture (3C) assays were performed as previously described (33). Cells differentiated from cord blood CD34^{+} cells on days 4, 8 and 11, int1Δ6 K562 cells, int8Δ K562 cells and WT K562 cells were cross-linked and lysed and the nuclei were digested with XbaI (Thermo Fisher Scientific, FD0684). After ligation and subsequent DNA purification, the cross-linking frequencies between the anchor and test fragments, as measured by the amount of the corresponding ligation product, were estimated by PCR relative to standards. The PCR products amplified from the human *ALAS2* gene (~20 kb) were ligated



B

Genotype of F1 progeny					
Age	No. of litters	No. of progeny with Alas2 genotype			
		X/X	X/Y	Δ13/X	Δ13/Y
E10.5	5	8	11	12	9
E11.5	6	9	13	10	6
E12.5	3	5	7	6	5*
E13.5-14.5	6	15	16	14	0
P30	31	56	52	44	0

*All pups were found dead.

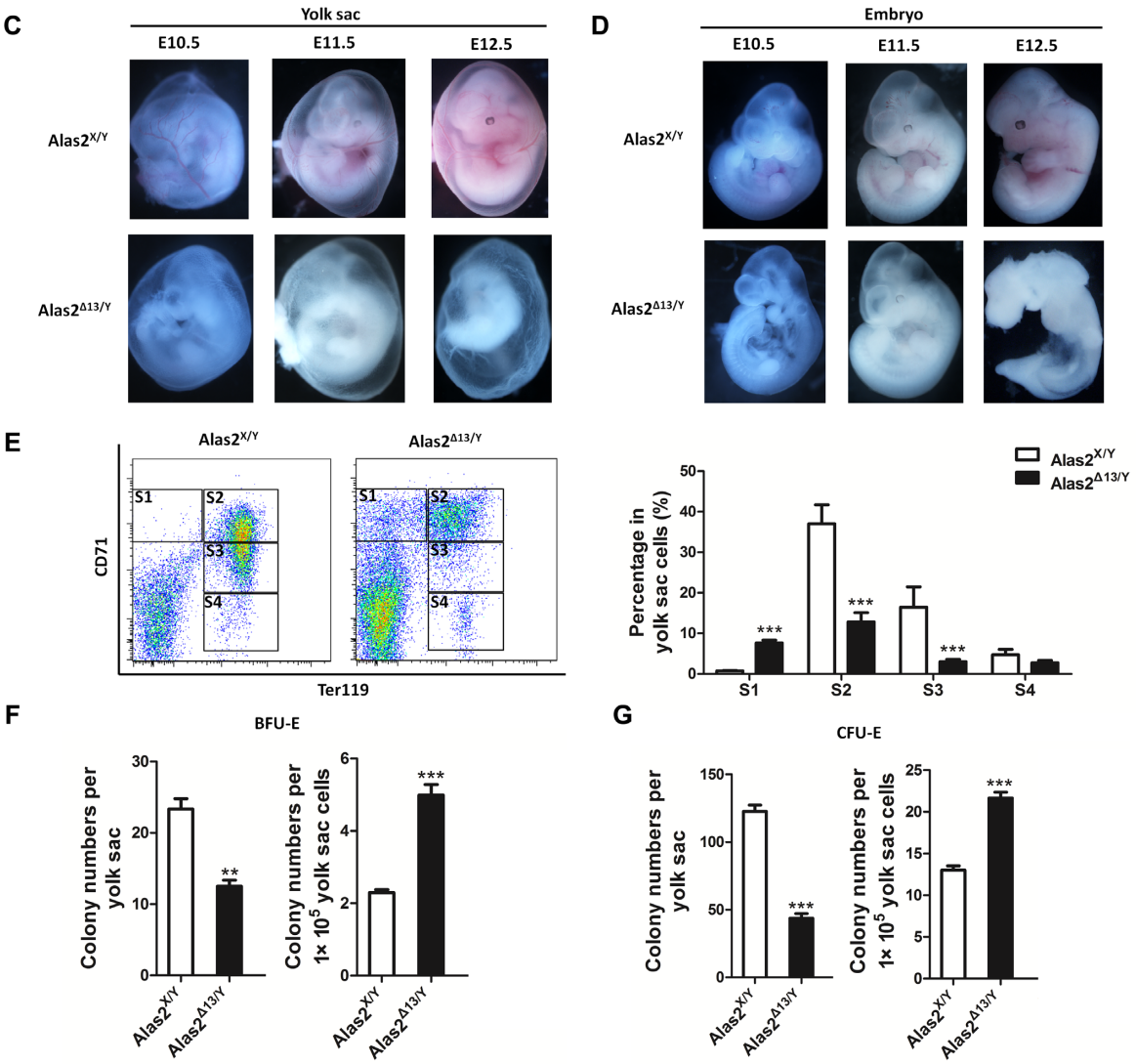


Figure 2. Deletion of the GATA1 binding site from intron 1 of the mouse *Alas2* gene leads to embryonic lethality. (**A**) Schematic showing the deletion of the GATA1 binding site from the first intron of the mouse *Alas2* gene using TALEN. The top panel shows the left (NTL) and right (NTR) TALEN DNA-binding sequences (underlined) with the spacer region; the bottom panel shows that 13 bp has been deleted from the target region of the first intron of the mouse *Alas2* gene. (**B**) Genotypes of F1 progeny. (**C** and **D**) The panels show embryos with yolk sacs (C) and embryos without yolk sacs (D) of int-1-GATA-deleted mice ($Alas2^{\Delta 13/Y}$) and their wild-type littermates ($Alas2^{X/Y}$) at E10.5-12.5. (**E**) Erythroid differentiation at E11.5 from yolk sac cells was examined by flow cytometry. (**F** and **G**) Colony-forming unit (CFU) assay of $Alas2^{\Delta 13/Y}$ and $Alas2^{X/Y}$ yolk sac cells at E11.5. Total yolk sac cells and 1×10^5 yolk sac cells were plated onto methylcellulose medium and colonies were counted on days 3–4 (CFU-E) (G) or days 9–10 (BFU-E) (F), respectively.

at a high concentration following XbaI digestion to generate equimolar mixtures of all possible ligation products and then used to generate standards. The cross-linking and ligation efficiencies between different samples and different experiments were normalized by setting the highest cross-linking frequency for each experimental series to 1.0. The error bars represent the SDs from the means of three to five independent experiments, as indicated in the figure legends. The primers used in this study are listed in Supplementary Table S1.

Western blotting

Cells differentiated from cord blood $CD34^+$ cells on days 4, 8 and 11, int1Δ6 K562 cells, int8Δ K562 cells, WT K562 cells and yolk sac cells from E11.5 mice were lysed, as previously described (34). The anti-ALAS2 antibody was obtained from Sigma-Aldrich (SAB1403541), anti-GATA1 was obtained from Abcam (ab11852) and anti-GAPDH was purchased from Cell Signaling Technology (#5174).

RNA purification and real-time PCR

RNA purification and real-time PCR were performed as previously described (34). The measured transcript copy numbers were normalized using GAPDH as a reference and fold changes were calculated according to the ΔΔCT method. The primers used in this study are listed in Supplementary Table S1.

Statistical analysis

Student's *t*-tests were used for comparisons between two groups, and ANOVA was used for multiple group comparisons. All results represent the average of at least three independent experiments and are presented as the mean ± SD. $P < 0.05$ was considered significant.

RESULTS

Identification of a GATA1 binding site mutation in *ALAS2* intron 1 from an XLSA pedigree

In 2010, our clinic diagnosed a large CSA pedigree that was suspected to be an XLSA pedigree due to the presence of multiple affected males in the family who were related through the maternal lineage (Figure 1A, proband III 7). Because ALAS2, which catalyzes the rate-limiting step of the heme biosynthesis pathway in erythroid cells (5), is the principle cause of XLSA, we then examined ALAS2 mRNA and protein abundance in bone marrow mononuclear cells (BMMCs) from three affected males (patient III 2, III 7 and III 9) in this pedigree and confirmed that ALAS2 expression was significantly decreased when compared with healthy individuals (Figure 1B and C). Thus, deficient ALAS2 expression could be the primary cause of XLSA in the affected males in this pedigree.

Next, we performed whole-genome sequencing of the *ALAS2* gene within this pedigree using genomic DNA extracted from PB and oral epithelial cells, which revealed an A>G transition at chromosomal position X: 55054635 [Chr X (GRCh37/hg19): g.55054635 A>G, NM_000032.4: c.-15–2187 T>C] in all the affected males (Figure 1A, gray-filled square); their mothers were heterozygous for this mutation, whereas the healthy males in this family were WT. The mutation occurs at the conserved GATA site of *ALAS2* intron 1, changing G**A**TA to G**G**TA (Figure 1D). A luciferase reporter assay confirmed the enhancer activity of this GATA site *in vitro* (Figure 1E). During the course of our study, the same mutation was also reported by other research groups (9,10). Our results and conclusions agree that this non-coding int-1-GATA sequence is an important *cis*-regulatory element of the *ALAS2* gene and that its mutation might be responsible for XLSA (9,10).

Hemizygous *Alas2* mutant embryos with intron 1 GATA site deletion are embryonic lethal due to severe anemia

In addition to the identified A>G mutation (G**A**TA>G**G**TA) in the intron 1 GATA site, mutations of G>C (G**A**TA>G**C**TA) and T>C (GA**T**A>GA**C**A) in the same GATA site have also been reported in XLSA pedigrees (9). Therefore, instead of generating a homologous point mutation (A>G) in the mouse, we used TALEN to delete the entire GATA site (Figure 2A) to clarify the potentially critical roles of this *cis*-regulatory element during *in vivo* development based on the high conservation of this region in mice and humans (Supplementary Figure S1).

Using TALEN-mediated mutagenesis, we generated a heterozygous founder female mouse bearing a 13 bp deletion ($Alas2^{\Delta 13/X}$) that removed the *Alas2* int-1-GATA site as well as its proximal flanking sequence (T**AGATA**AGCCCC) (Figure 2A). Astonishingly, when the female heterozygous mouse ($Alas2^{\Delta 13/X}$) was crossed with a WT male mouse ($Alas2^{X/Y}$), the only newborns recovered bore genotypes $Alas2^{\Delta 13/X}$, $Alas2^{X/X}$ and $Alas2^{X/Y}$, but no male hemizygous ($Alas2^{\Delta 13/Y}$) mice were obtained (Figure 2B). Because the *Alas2* gene is located on the X chromosome, we speculated that homozygous disruption of the int-1-GATA sequence led to embryonic lethality.

To investigate why the hemizygous genotype is embryonic lethal, we continuously followed development by collecting embryos from embryonic day 10.5 (E10.5) to E14.5 (Figure 2B). Embryos of all genotypes ($Alas2^{X/X}$, $Alas2^{X/Y}$, $Alas2^{\Delta 13/X}$ and $Alas2^{\Delta 13/Y}$) were observed from E10.5 to E11.5; however, hemizygous male embryos ($Alas2^{\Delta 13/Y}$) were not recovered beyond E12.5 (Figure 2B). All the $Alas2^{\Delta 13/Y}$ embryos were pale and severely anemic (Figure 2C and D), and few red blood cells were observed in the blood vessels of $Alas2^{\Delta 13/Y}$ yolk sacs at E10.5 and E11.5 (Figure 2C); at E12.5, the $Alas2^{\Delta 13/Y}$ embryos exhibited severe pallor and began to be resorbed (Figure 2D). FACS analyses and colony-forming assays showed that erythroid differentiation was arrested at an immature erythroid progenitor stage (Figure 2E, F and G).

Strikingly, we also observed that Alas2 transcription and protein levels were reduced to a nearly undetectable level in the yolk sac cells of the $Alas2^{\Delta 13/Y}$ mice (Figure 3A and B), suggesting that the int-1-GATA site is the key regulatory element driving *Alas2* transcription. Conversely, transcription of embryonic globins (εy and βh1) and Gata1 were unaffected (Figure 3A). Because ALAS2 catalyzes the rate-

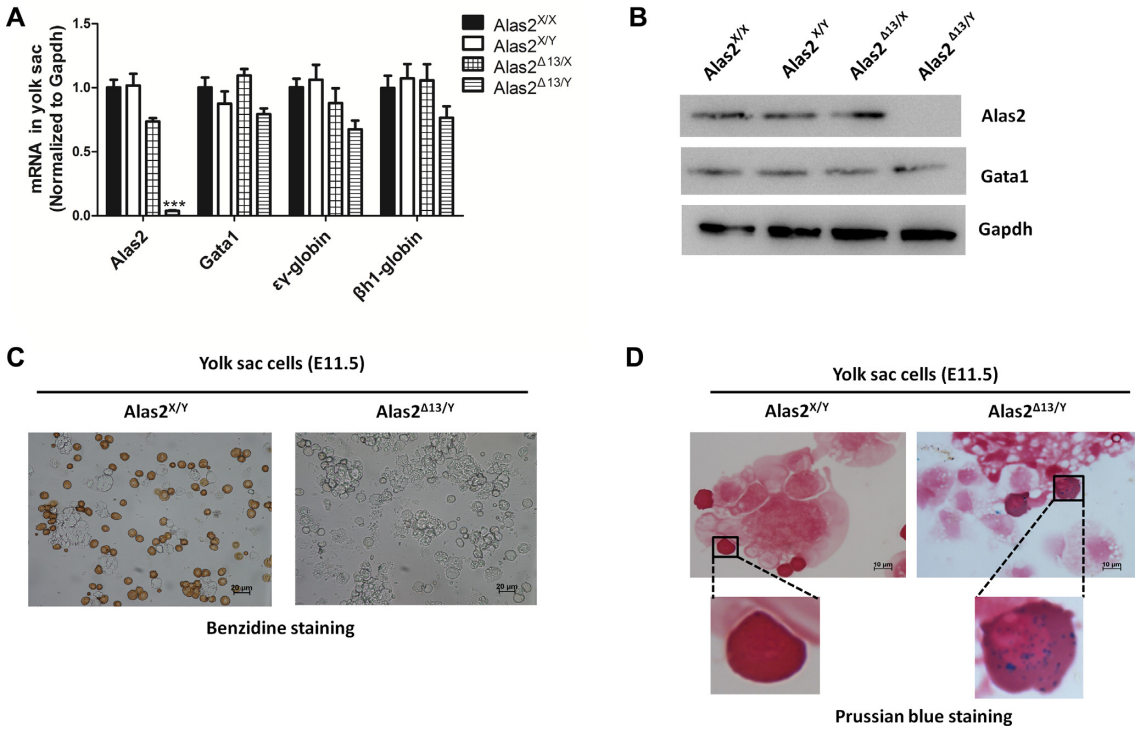


Figure 3. Deletion of the intron 1 GATA site of the mouse *Alas2* gene leads to reduced Alas2 expression and impaired heme biosynthesis. (**A**) The relative expression of Alas2, Gata1, εγ-globin and βh1-globin mRNA in yolk sac cells of $Alas2^{X/Y}$, $Alas2^{\Delta 13/X}$, $Alas2^{\Delta 13/Y}$ mice to that of $Alas2^{X/X}$ mice at E11.5 by RT-qPCR, respectively. Here Gapdh mRNA acted as the internal control. (**B**) Expression of Alas2 and Gata1 protein in yolk sac cells of $Alas2^{X/X}$, $Alas2^{X/Y}$, $Alas2^{\Delta 13/X}$, $Alas2^{\Delta 13/Y}$ mice at E11.5 by western blotting. Gapdh was used as a loading control. (**C**) Benzidine staining of yolk sac cells from $Alas2^{\Delta 13/Y}$ and $Alas2^{X/Y}$ mice at E11.5. (**D**) Yolk sac cells from $Alas2^{\Delta 13/Y}$ and $Alas2^{X/Y}$ mice at E11.5 were stained with Prussian blue. The experiments were performed in three independent experiments. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

limiting step of heme biosynthesis in erythroid cells (5), we performed benzidine staining to examine heme biosynthesis and found significantly fewer benzidine-positive cells in the $Alas2^{\Delta 13/Y}$ yolk sac cells at E11.5, indicating that heme biosynthesis was severely impaired in these knockout embryos (Figure 3C). As the impairment of heme biosynthesis perturbs iron metabolism, we also applied Prussian blue staining to monitor iron accumulation and detected iron in the cytoplasm surrounding the nucleus of in $Alas2^{\Delta 13/Y}$ yolk sac erythroid cells at E11.5, which resembled the typical pathological sideroblastic characteristic of CSA (Figure 3D). In summary, this study is the first to demonstrate *in vivo* that the DNA fragment containing the int-1-GATA site is so indispensable for ALAS2 expression that embryos bearing this germline deletion cannot survive to birth, a defect that results from heme biosynthesis impairment and erythroid differentiation arrest.

ALAS2 expression is regulated by an enhancer chromatin loop

We next investigated how the GATA site regulates ALAS2 expression. The luciferase reporter assay suggested that the int-1-GATA site acted as an enhancer to regulate ALAS2 expression (Figure 1E). In addition to this int-1-GATA site, another GATA site in *ALAS2* intron 8 (int-8-GATA) has also been reported to function as an enhancer through luciferase assays (35). To understand how these GATA sites integrate together to regulate ALAS2 expression, we initially examined the occupancies of GATA1 at the promoter and the intron 1 and 8 enhancer regions in K562 cells (Figure 4A) and detected their enrichment at these regions (Figure 4B). As FOG1 is an important partner of GATA1 that facilitates GATA1 occupancy at multiple chromatin sites (20) and TAL1 often co-localizes with GATA1 to promote transcription (23), we also investigated their occupancies. Our results showed that TAL1, but not FOG1, co-localized with GATA1 to the intron 1 and 8 enhancer regions (Figure 4C and D). Surprisingly, occupancy of Pol II was observed not only at the promoter but also at the intron 1 and 8 enhancer regions (Figure 4E).

A previous study reported that GATA1 and TAL1 preferentially recruit LDB1 and LMO2 to form a long-range chromatin loop to distally regulate erythroid-specific genes (27). In our study, we also found that LDB1 and LMO2 bound to the intron 1 and 8 enhancer regions (Figure 4F and G). Furthermore, the physical proximity between the int-1-GATA and int-8-GATA sites and between the promoter and the int-8-GATA site were verified by 3C assays (Figure 4H and I), suggesting that the promoter, the int-1-GATA site and the int-8-GATA site could form a potential chromatin enhancer loop.

The formation of a chromatin loop was also confirmed physiologically (Figure 5D and E) during *ex vivo* erythroid differentiation induced from human cord blood $CD34^{+}$ primary hematopoietic progenitor cells, as monitored by flow

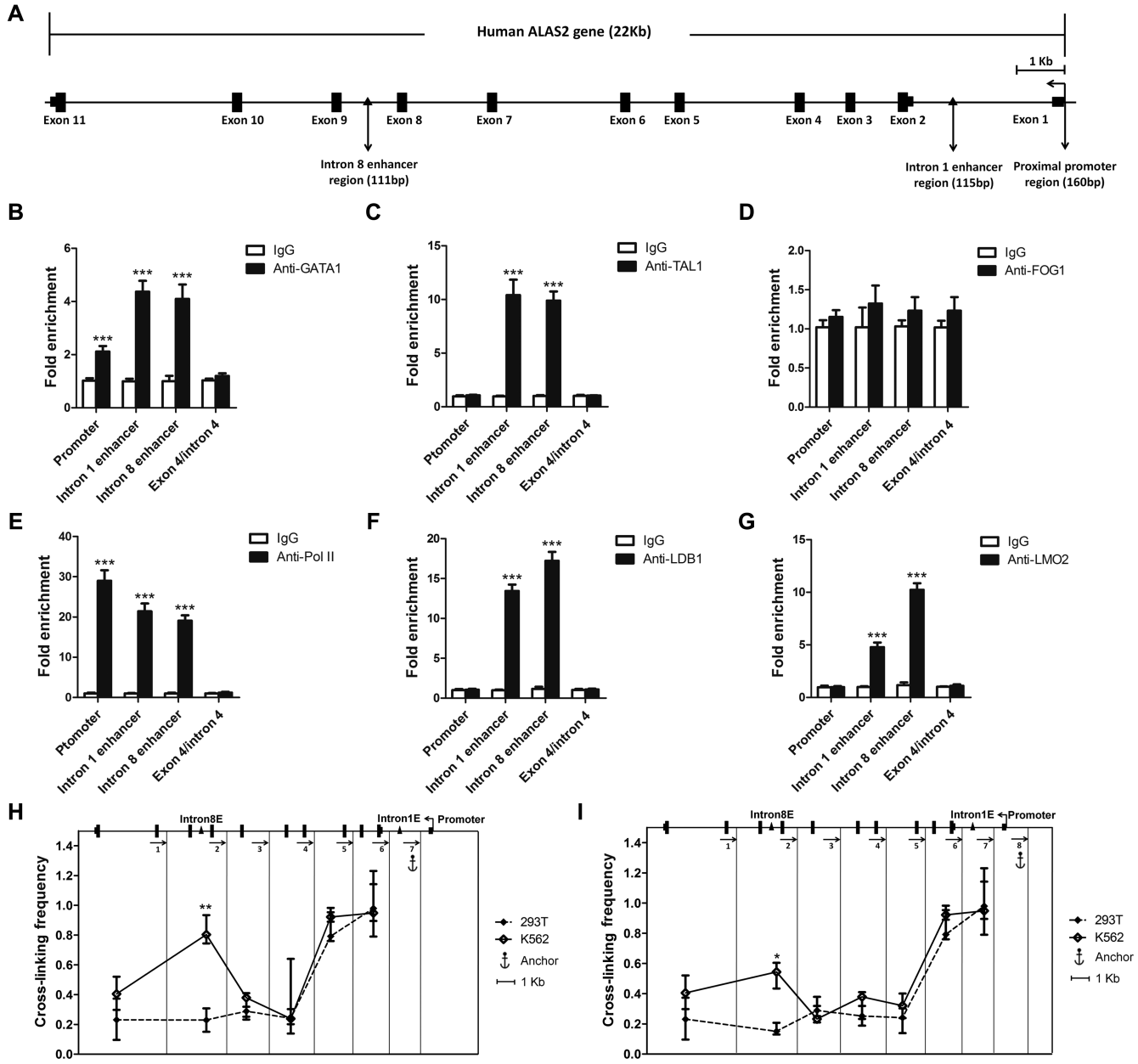


Figure 4. The intron 1 GATA site physically interacts with the promoter and the intron 8 GATA site to form a long-rang enhancer chromatin loop. (**A**) Schematic of the *ALAS2* proximal promoter region and intron 1 and 8 enhancer regions used for ChIP-qPCR. The examined promoter region in this study is located at chromosomal position X: 55057352–55057511, which includes the GATA site; the intron 1 enhancer region is located at chromosomal position X: 55054583–55054697, which includes the GATA site and its adjacent E-box; the intron 8 enhancer region is located at chromosomal position X: 55041577–55041687, which includes the GATA site and its adjacent E-box. The ChIP-qPCR primers for the proximal promoter region, intron 1 enhancer region and intron 8 enhancer region are shown in Supplementary Table S1. (**B**–**G**) The enrichment of GATA1, TAL1, FOG1, Pol II, LDB1 and LMO2 at the promoter, intron 1 and intron 8 enhancer regions was demonstrated by ChIP-qPCR in K562 cells. Normal rabbit IgG was employed as the control in all ChIP assays. The DNA sequences at the exon 4-intron 4 junction (exon 4/intron 4) acts as a negative control region for all ChIP-qPCR assays. (**H** and **I**) 3C assays indicate that the intron 8 GATA site physically interacts with the intron 1 GATA site (H) and promoter (I) in K562 cells. Vertical lines represent XbaI restriction sites; arrows indicate PCR primer sites and direction. Anchor symbols mark the anchoring primer. The experiments were performed independently in triplicate. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

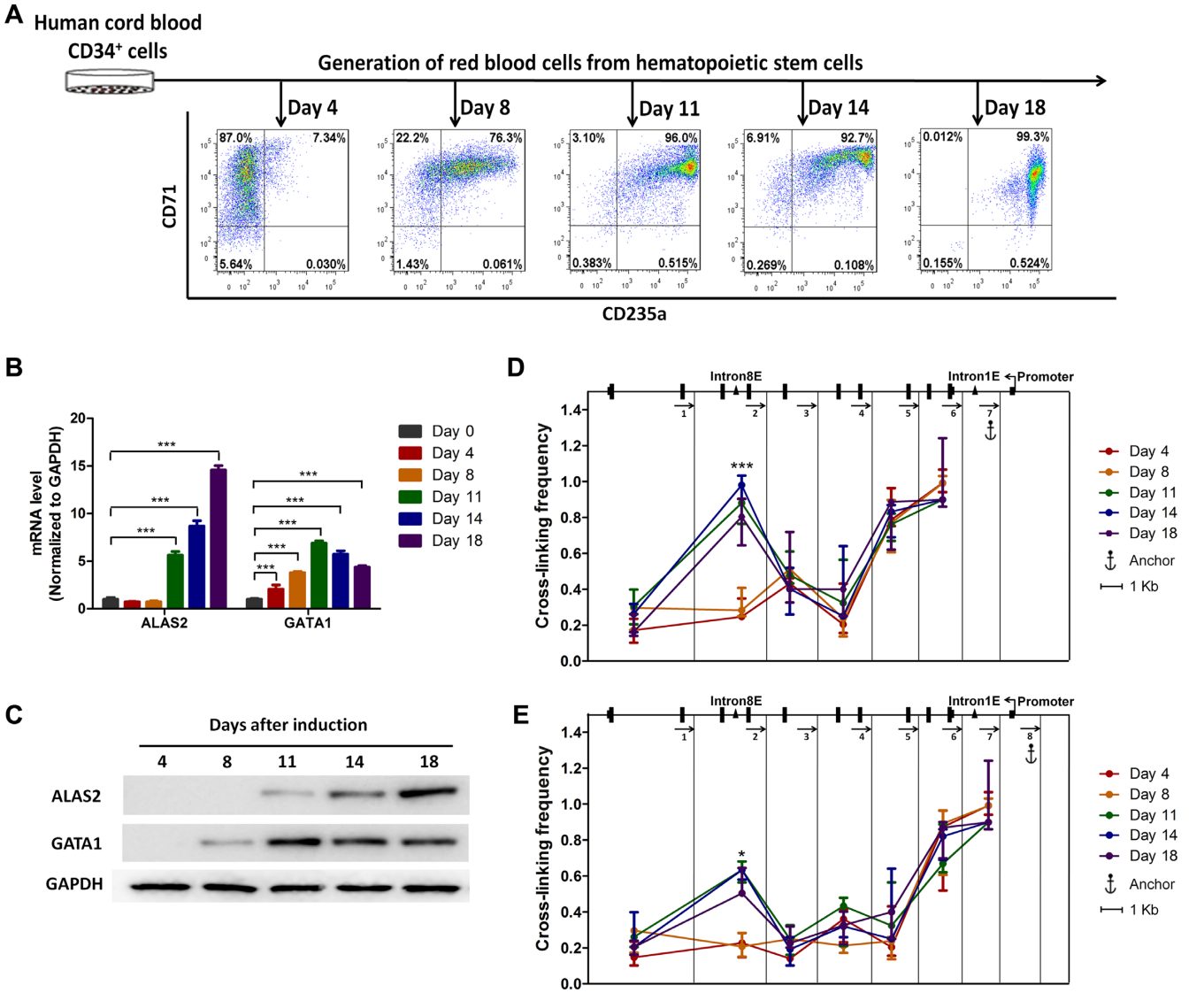


Figure 5. A chromatin loop linking the promoter, intron 1 and intron 8 GATA sites are formed during human erythroid progenitor CD34$^+$ differentiation *ex vivo*. (**A**) Erythroid differentiation of cord blood progenitor CD34$^+$ cells induced *ex vivo* was monitored by FACS (stained with anti-CD71 and anti-CD235a antibodies). (**B**) *ALAS2* and *GATA1* transcription was measured by RT-qPCR during erythroid differentiation and normalized to their transcription levels at day 0. GAPDH mRNA acted as the internal control. (**C**) The protein abundance of ALAS2 and GATA1 was measured during erythroid differentiation. GAPDH served as the loading control for Western blotting. (**D** and **E**) 3C assays indicated that the intron 8 GATA site physically interacts with intron 1 GATA site (D) and promoter (E) to form a chromatin loop from day 11 until day 18 in differentiated erythroid cells. Vertical lines represent XbaI restriction sites, and arrows indicate the PCR primer sites and direction. Anchor symbols mark anchoring primers. All experiments were performed independently in triplicate. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

cytometry (Figure 5A). Although GATA1 protein expression was detectable on day 8 during human primary erythroid differentiation (Figure 5C), we did not observe the enrichment of GATA1 and Pol II at the promoter or the intron 1 and intron 8 enhancer regions, nor the enrichment of TAL1, LDB1 and LMO2 at the intron 1 and 8 enhancer regions until day 11 of erythropoiesis (Figure 6A–E), a time when ALAS2 expression began to markedly increase (Figure 5B and C).

Taken together, our results indicate that to facilitate the dramatic increase of ALAS2 expression in erythroid cells, GATA sites in the promoter and introns 1 and 8 form an enhancer loop that is mediated by an enhancer protein complex centered by GATA1 and involved TAL1, LDB1, LMO2 and Pol II at least.

The two enhancer GATA sites in introns 1 and 8 unequally modulate ALAS2 expression

To further interrogate the specific role of proximal int-1-GATA and distal int-8-GATA enhancer elements in ALAS2 regulation, we employed CRISPR/Cas9 technology to separately delete the GATA1 binding sites from the intron 1 or 8 enhancer regions in K562 cells. We generated an int-1-GATA site knockout cell line (int1Δ6) harboring a 6 bp (A**GATA**A) deletion (Figure 7A) and an int-8-GATA-knockout line (int8Δ4) carrying a 4 bp deletion, including

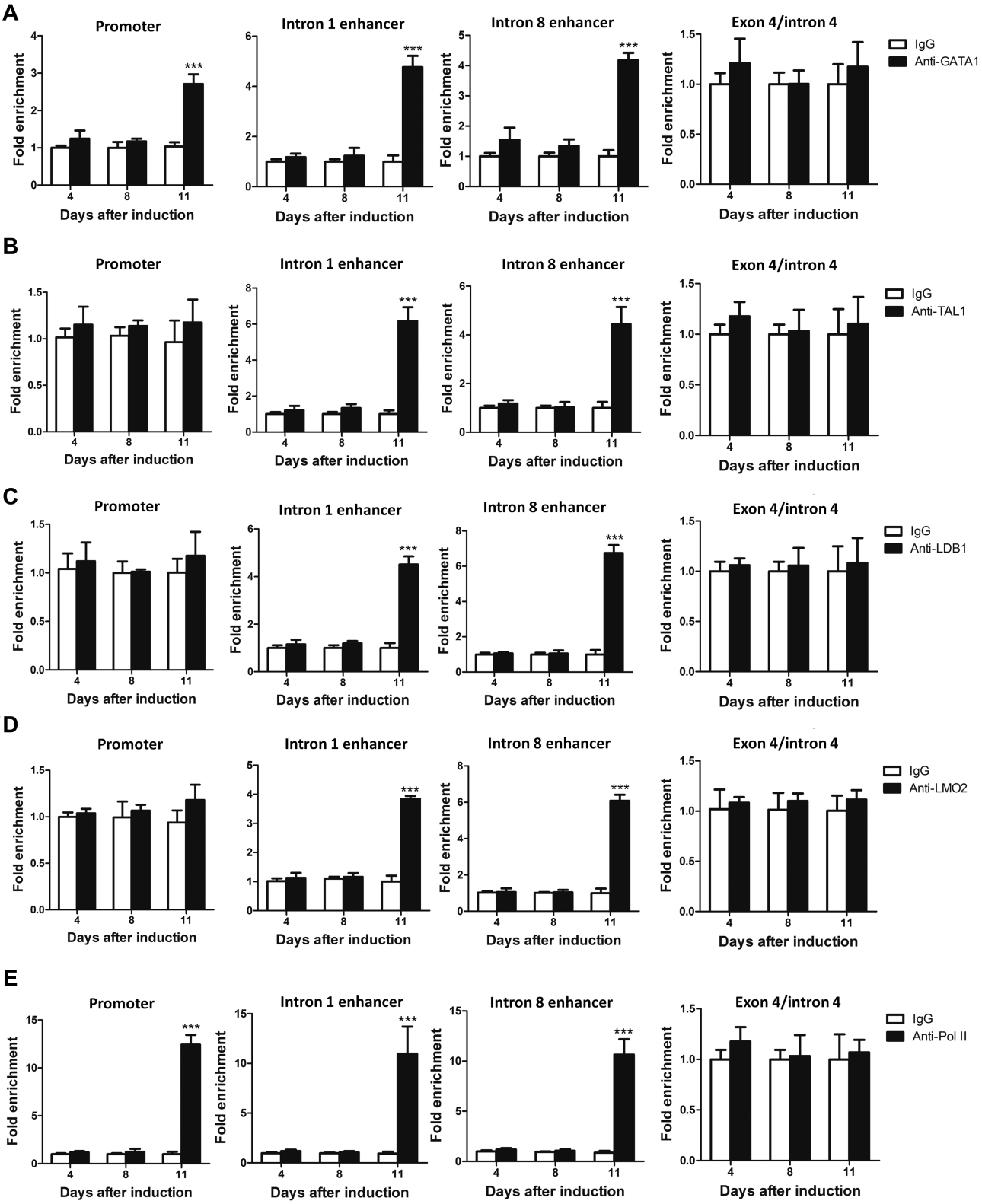


Figure 6. The enrichment of the enhancer protein complex at the promoter, intron 1 enhancer and intron 8 enhancer region during human erythroid progenitor $CD34^+$ differentiation *ex vivo*. (**A**–**E**) Enrichment of GATA1, TAL1, LDB1, LMO2 and Pol II at the promoter, intron 1 enhancer and intron 8 enhancer region in differentiated erythroid cells at day 11, according to ChIP-qPCR. Normal rabbit IgG was employed as the control in all ChIP assays. The DNA sequence at the exon 4-intron 4 junction (exon 4/intron 4) acts as a negative control region for all ChIP-qPCR assays. All experiments were performed independently in triplicate. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

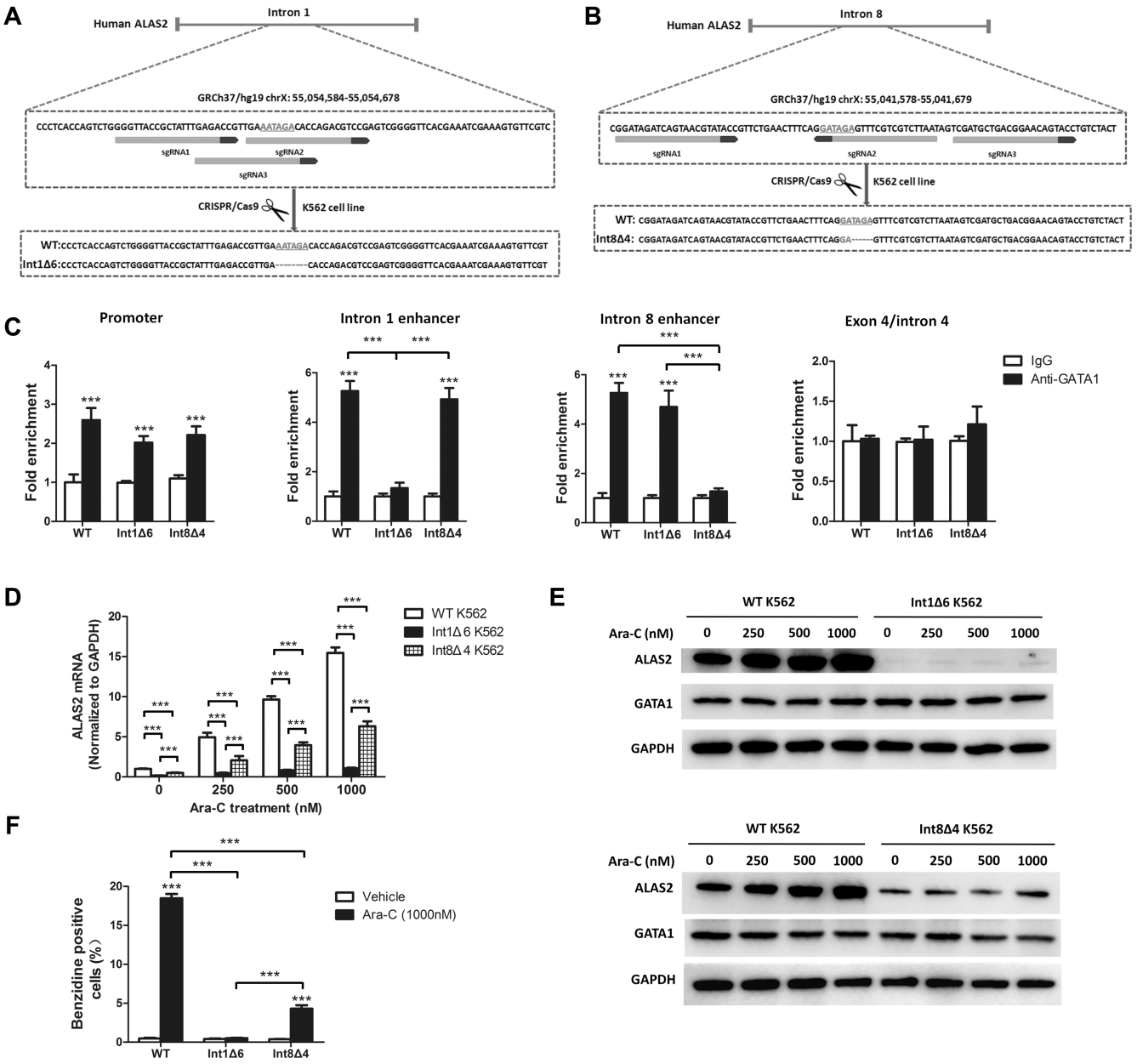


Figure 7. Deletion of the GATA1 binding sites from either intron 1 and 8 significantly decreases ALAS2 expression. (**A**) Schematic of the *ALAS2* int-1-GATA site deletion in K562 cells using CRISPR/Cas9. The top panel shows the sgRNA binding sequences; the bottom panel shows the deletion of a 6 bp fragment from the target region of intron 1 of the human *ALAS2* gene in K562 cells. (**B**) Schematic of the int-8-GATA site deletion. The top panel shows the sgRNA binding sequences; the bottom panel shows the deletion of a 4 bp fragment from the target region of intron 8 of the *ALAS2* gene. (**C**) The ChIP-qPCR assay showed the binding of GATA1 to the promoter and the intron 1 and 8 enhancer regions in int1Δ6, int8Δ4 and WT K562 cells. The DNA sequence at the exon 4-intron 4 junction (exon 4/intron 4) acts as a negative control region. Normal rabbit IgG was employed as the control in all ChIP assays. (**D** and **E**) ALAS2 transcript (D) and protein (E) abundance was measured by RT-qPCR and western blotting in WT K562, int1Δ6 and int8Δ4 cells after treatment with AraC for 72 h, respectively. GAPDH transcripts (D) and protein (E) acted as the reference gene and the loading control for RT-qPCR and western blotting, respectively. (**F**) The percentages of benzidine-positive cells in int1Δ6, int8Δ4 and their WT control after inducing K562 erythroid differentiation with AraC for 72 h. All experiments were performed independently in triplicate. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

the first 3 bp of the int-8-GATA site and 1 bp flanking sequence [**(T/A)GAT**A(A/G)] (Figure 7B). As reported previously, deletion of the first 3 nucleotides of the 'GATA' element is sufficient to disrupt GATA1 binding to chromatin (36,37); thus, we believe that the int8Δ4 line is suitable for studying the role of the int-8-GATA region. As expected, GATA site deletion in int1Δ6 and int8Δ4 cells entirely disrupted GATA1 occupancy at the intron 1 and 8 enhancer regions, respectively, whereas each deletion did not affect GATA1 occupancy at the other intronic GATA site and the promoter region (Figure 7C).

We detected significant reductions in ALAS2 expression in both int1Δ6 and int8Δ4 cells during cytosine arabinoside (AraC)-induced K562 erythroid differentiation (Fig-

ure 7D and E). However, these two enhancer GATA sites contributed unequally to ALAS2 expression. For example, ALAS2 expression was almost entirely abolished in int1Δ6 cells, whereas in int8Δ4 cells, the reduction of ALAS2 was relatively mild (90 versus 50% reduction at 1000 nM AraC treatment) (Figure 7D and E). Consequently, the percentage of benzidine-positive int1Δ6 cells was also significantly lower compared with int8Δ4 cells (∼0.5 versus 4% after 1000 nM AraC treatment) (Figure 7F). These results imply that the int-1-GATA site plays a more critical role in regulating *ALAS2* transcription. Moreover, the int1Δ6 deletion phenocopied int1Δ13-deleted erythroid cells (Figures 2 and 3), indicating that the int-1-GATA site is the core element of the 13 bp fragment and is principally responsible for the severe anemia and embryonic lethality of $Alas2^{\Delta 13/Y}$ embryos.

The int-1-GATA site deletion significantly reduces Pol II enrichment at both the intron 1 enhancer region and promoter

3C assays revealed that int-1-GATA or int-8-GATA site deletion alone in int1Δ6 and int8Δ4 cells sufficiently disrupted the formation of the long-range chromatin enhancer loop (Figure 8A and B). Surprisingly, ∼50% of ALAS2 expression was maintained in int8Δ4 cells (Figure 7D and bottom panel of E), indicating that the formation of the long-range enhancer loop, which ensures maximum ALAS2 activation, was sufficient but not necessary for ALAS2 expression. In contrast, ALAS2 expression was nearly eliminated in int1Δ6 cells (Figure 7D and upper panel of E), suggesting that int-1-GATA is indispensable for ALAS2 expression and its function extends beyond acting as an anchor for the long-range enhancer loop.

To understand how these two enhancer sites unequally modulate ALAS2 expression, we investigated the enrichment of other core components, in addition to GATA1 (Figure 7C), of the enhancer complex at the GATA sites in the intron 1 and 8 enhancer regions in the mutant cell lines. We observed that enrichment of TAL1, LDB1 and LMO2 was significantly decreased (Figure 8C–E) in these deleted cells. Surprisingly, in int8Δ4 cells, ∼50% of Pol II occupancy remained at the proximal promoter and the intron 1 enhancer region, but not at the intron 8 enhancer region. Conversely, int-1-GATA site deletion not only completely abolished Pol II enrichment at the intron 1 and 8 enhancer regions but also decreased its occupancy at the promoter region by ∼90% (Figure 8F). This result may partially explain why the int-1-GATA site is indispensable for ALAS2 expression.

To further support our hypothesis that the int-1-GATA site deletion significantly reduced Pol II enrichment at both the intron 1 enhancer region and the promoter region *in vivo*, we investigated the Pol II occupancy at the int-1-GATA site and promoter region in the yolk sac cells of $Alas2^{\Delta 13/Y}$ mice. We first confirmed that the int-1-GATA site deletion completely abolished GATA1 enrichment (Figure 8G). We also failed to detect the enrichment of Pol II at both the intron 1 enhancer and promoter regions (Figure 8H).

In summary, we propose a regulatory model in which GATA1 regulates ALAS2 in erythroid cells via the dual enhancer elements. To achieve full activation of the *ALAS2* gene transcription, the int-1-GATA1 site acts as an anchor to link the GATA sites at promoter and intron 8 to form a long-rang enhancer loop by recruiting Pol II, TAL1, LMO2 and LDB1 (Figures 4-6). When the int-8-GATA site is mutated or deleted, the enhancer loop is disturbed (Figure 8A and B). Nevertheless, weak enrichment of several components of the enhancer complex at the intron 1 enhancer region and partial Pol II enrichment at the proximal promoter is maintained (Figure 8C–F), which allows for the initiation of ALAS2 expression, albeit at low levels (∼50% ALAS2 transcription) (Figure 7D). This finding suggests that the enhancer loop is functional but not essential. However, the disruption or deletion of the intron 1 GATA site not only disturbs the enhancer loop but also significantly reduces enrichment of Pol II at the promoter and int-1-GATA sites simultaneously (Figure 8), resulting in insufficient ALAS2 expression (Figure 7D and upper panel of E) and, ultimately, the development of XLSA, which indicates the indispensable role of int-1-GATA site in ALAS2 regulation.

DISCUSSION

In this study, we assessed an XLSA pedigree harboring an A>G mutation in the GATA site of *ALAS2* intron 1. We discovered that deletion of the int-1-GATA site and its flanking fragment (13 bp) led to early embryonic lethality due to severe anemia resulting from the absence of ALAS2 expression, suggesting an indispensable role for this noncoding *cis*-regulatory element in ALAS2 expression *in vivo*. By generating int-1-GATA and int-8-GATA deletion cell lines using CRISPR/Cas9, we demonstrated that GATA1 activated ALAS2 expression by forming an enhancer loop via GATA sites at the promoter and introns 1 and 8 involving at least TAL1, LMO2, LDB1 and Pol II.

During the course of our study, two papers demonstrating similar mutations in the GATA1 binding site of the first intron of the *ALAS2* gene were published (9,10). Campagna *et al*. studied five XLSA families with mutations in the *ALAS2* intron 1 GATA site; three of these XLSA families carry an identical point mutation identified by us (A>G at g.55054635), whereas the other two families carry other point mutations in the int-1-GATA site (G>C and T>C at g.55054634 and g.55054636, respectively) (9). The mutation identified by Kaneko *et al*. in two XLSA Japanese families (10) is the same as the mutation described here (A>G at g.55054635). Taken together, these data clearly indicate that int-1-GATA mutations can result in XLSA and that int-1-GATA holds great promise as a screening site for the diagnosis of XLSA patients with unknown mutations.

In our study, we generated mice with a deletion in the int-1-GATA site and its flanking fragment ($Alas2^{\Delta 13/Y}$). Astonishingly, we discovered that this deletion led to an embryonic-lethal phenotype due to severe anemia between E11.5 and E12.5. Hence, this study is the first to demonstrate *in vivo* that the int-1-GATA site is an indispensable regulatory element and that embryos bearing this germline deletion cannot survive to birth. The phenotype of $Alas2^{\Delta 13/Y}$ mice nearly completely recapitulated that of *Alas2* null mutant knockout mice ($Alas2^{-/-}$), which are also embryonic lethal due to anemia (38). However, $Alas2^{\Delta 13/Y}$ embryos died at E11.5-E12.5, which is slightly later than

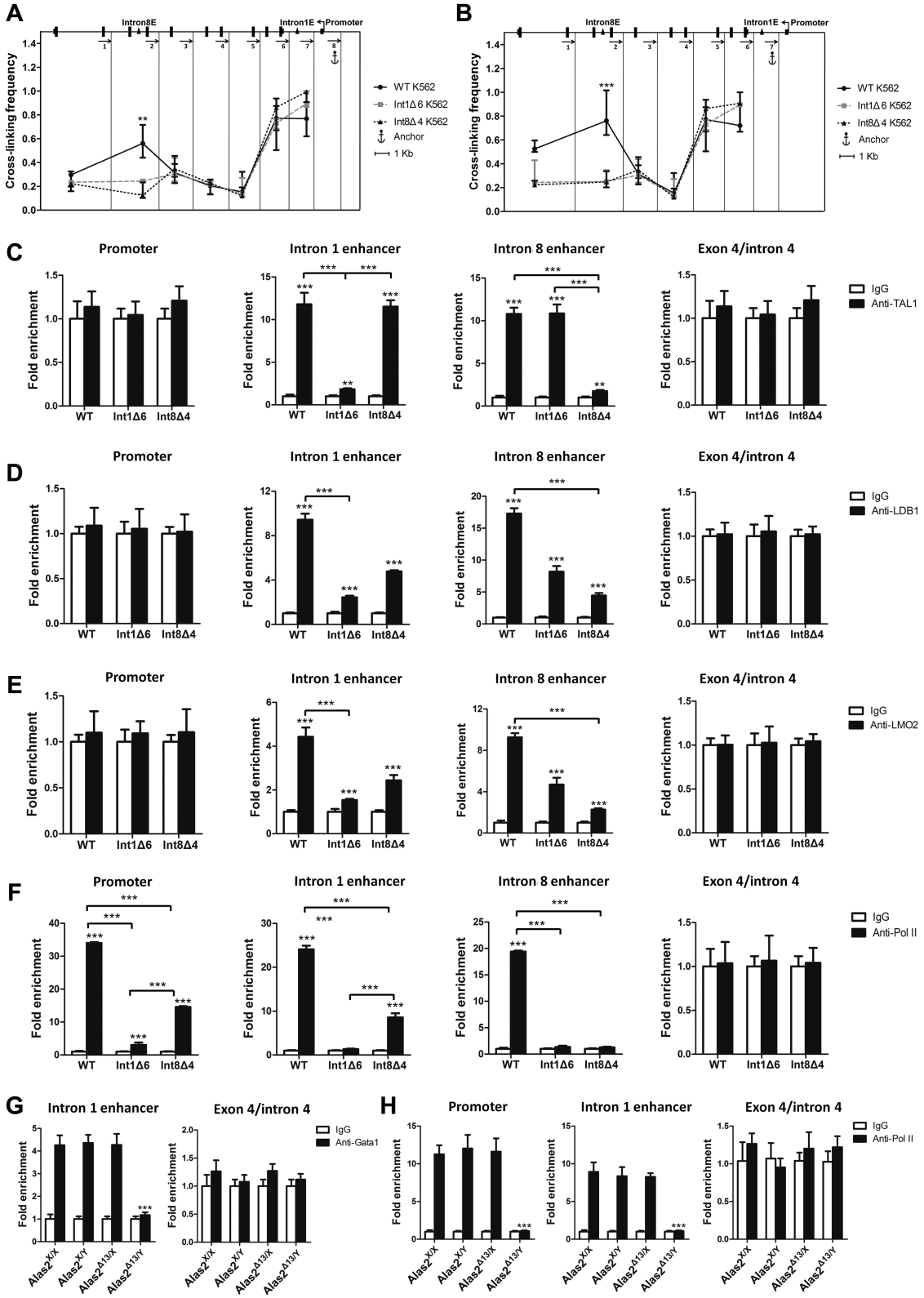


Figure 8. Deletion of the int-1-GATA site abolishes the enhancer loop and almost entirely removes Pol II enrichment at the intron 1 enhancer region and promoter. (**A** and **B**) The formation of a chromatin loop between the intron 8 GATA site and the intron 1 GATA site (A) or the promoter region (B) in int1Δ6, int8Δ4 and WT K562 cells was measured by 3C assays. (**C**–**F**) The binding of TAL1, LDB1, LMO2 and Pol II to the promoter and the intron 1 and 8 enhancer regions was measured by ChIP-qPCR in int1Δ6, int8Δ4 and WT K562 cells, respectively. (**G**) The occupancy of Gata1 at the intron 1 GATA site in yolk sac cells of *Alas2*$^{X/X}$, *Alas2*$^{X/Y}$, *Alas2*$^{\Delta 13/X}$, *Alas2*$^{\Delta 13/Y}$ mice at E11.5 by ChIP-qPCR. (**H**) The occupancy of Pol II at the intron 1 GATA site and promoter in yolk sac cells of *Alas2*$^{X/X}$, *Alas2*$^{X/Y}$, *Alas2*$^{\Delta 13/X}$, *Alas2*$^{\Delta 13/Y}$ mice at E11.5 by ChIP-qPCR. Normal rabbit IgG was employed as the control in all ChIP assays. The DNA sequence at the exon 4-intron 4 junction (exon 4/intron 4) acts as a negative control region for all ChIP-qPCR assays. All experiments were performed independently in triplicate. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

Alas2$^{-/-}$ embryos (E10.5–E11.5). This variation might be due to the different methods used for calculating embryonic development in different laboratories or to minute residual Alas2 expression in *Alas2*$^{\Delta 13/Y}$ embryos that enables them to survive slightly longer.

Our FACS and colony-forming assays revealed that the erythroid cells in the *Alas2*$^{\Delta 13/Y}$ embryos were arrested in an immature erythroblast stage, most likely at the proerythroblast stage, when hemoglobin production begins (39). Hemoglobin production needs to coordinate or balance the generation of heme and globin chains (40,41). Tanimura *et al*. revealed that heme deficiency affected a subset of GATA1 target genes that encode factors involved in erythroid cell homeostasis and red blood cell genesis from progenitor cells, indicating that heme is required to aid GATA1 in establishing the erythroblast transcriptome (41). Since heme biosynthesis is impaired in the erythroid progenitor cells of *Alas2*$^{\Delta 13/Y}$ embryos, which results in discoordination of hemoglobin production as well as disorders of GATA1-mediated erythroid cell homeostasis, these erythroid cells in the mutant embryos ultimately arrested around the proerythroblast stage.

Though the mice in which this GATA site is deleted are exclusively embryonic lethal, human patients who harbor an inherited single nucleotide point mutation within this int-1-GATA site are viable. The phenotype discrepancy between the int-1-GATA site deletion and mutation might depend on the amount of residual ALAS2 expression caused by either deletion or mutation. One possible interpretation is that the inherited single nucleotide point mutation in the int-1-GATA site can partially impair its enhancer function, resulting in subnormal but minimal ALAS2 expression that confers survival to these patients. Interestingly, heterozygous females carrying the single nucleotide mutations of this site in our XLSA pedigree suffered spontaneous abortion more frequently than homozygous WT females (data not shown), which could be a consequence of lower ALAS2 expression levels below the minimal requirement. The observation that an identical point mutation could cause variable ALAS2 expression levels suggests the possible involvement of other factors in the regulation of ALAS2 expression, such as epigenetic modifications; this possibility needs to be further addressed. In the future, the generation of mice harboring single nucleotide point mutations in the int-1-GATA site and the subsequent tracing of their development and measuring of their ALAS2 expression levels will further illustrate the pathogenesis of XLSA.

Prior to our study, several lines of evidence have indicated that *ALAS2* is a potential key downstream target of GATA1. For example, the expression levels of ALAS2 and other heme biosynthesis enzymes are significantly reduced in *GATA1* promoter-disrupted erythroid cells differentiated from ES cells *in vitro* (17). Moreover, the phenotypes of *Alas2*$^{\Delta 13/Y}$ and *Alas2*$^{-/-}$ mice partially resemble those of GATA1 germline deletion mice (*Gata1*$^{-/-}$), even though *Alas2*$^{\Delta 13/Y}$ and *Alas2*$^{-/-}$ embryos die slightly later than *Gata*1$^{-/-}$ embryos (14). Since abundant GATA1 expression begins as early as day 4, we speculate that GATA1 might activate other critical erythroid processes prior to initiating the heme biosynthesis process, which could partially explain why *Gata1*$^{-/-}$ germline null mutant mice die (14) earlier than *Alas2*$^{-/-}$ knockout mice (38).

Our mechanistic analysis revealed that the GATA1 protein forms an enhancer protein complex involving TAL1, LDB1, LMO2 and Pol II, but not FOG1, to efficiently regulate ALAS2 expression by occupying GATA sites in introns 1 and 8. These findings shed light on how GATA1 regulates *ALAS2* in erythroid cells. Consistent with our findings, ENCODE ChIP-Seq data have shown that in K562 cells, GATA1 and TAL1 are enriched at regulatory elements in introns 1 and 8 of *ALAS2* (42). In further support of our hypothesis of the formation of an enhancer protein complex, previous studies have shown that LDB1 knockdown in erythroid cell lines decreases ALAS2 expression (43). Moreover, Wakabayashi *et al*. also recently showed that GATA1, TAL1, LMO2 and LDB1 can form a protein complex to regulate efficient ALAS2 expression (44). In regard to the TAL1 binding capacity after the int-1-GATA site deletion, we agree with those authors that TAL1 enrichment was significantly reduced after this GATA site deletion. However, a slight discrepancy between their study and ours is that we still detected sparse TAL1 occupancy after GATA site deletion, whereas in their study, TAL1 enrichment was completely abolished (44). We speculate that the slight residual levels of TAL1 enrichment might act as an anchor for remnant levels of LDB1 and LMO2 occupancy since these two proteins are not DNA binding proteins. However, it is also possible that the experimental procedures applied by different laboratories resulted in this slight discrepancy.

Our *in vitro* results also demonstrated that the deletion of the int-1-GATA region nearly eliminated ALAS2 expression, which is in agreement with recently published data by Tanimura *et al*. and Wakabayashi *et al*.; those studies showed that CRISPR/Cas9-mediated deletion of the int-1-GATA site strongly reduces *ALAS2* transcription and heme biosynthesis in both murine proerythroblast-like cells (G1E-ER-GATA1 cells) (41) and human K562 cells (44), indicating that the int-1-GATA site is indispensable for ALAS2 expression. Moreover, consistent with Tanimura *et al*. study, we also detected the upregulation of BACH1 and the downregulation of HBA, HBB, TBCEL and SLC7A11 in int-1-GATA deleted K562 cells and in the yolk sac cells of *Alas2*$^{\Delta 13/Y}$ mice (Supplementary Figure S2A and B) (41).

Tanimura *et al*. and we both found that compared to the int-1-GATA site, the regulatory role of int-8-GATA site is relatively weak. We observed that int-8-GATA site deletion only impaired partial ALAS2 expression. Though Tanimura *et al*. observed a more significant reduction of Alas2 expression in the intron 1 and 8 double GATA site-deleted G1E-ER-GATA-1 cells than that in the cells containing only the int-1-GATA site deletion, int-1-GATA site plays a much more important role in regulating ALAS2 expression (41). This might explain why no intron 8 GATA site mutants have thus far been identified in XLSA patients. An individual who harbors the intron 8 GATA site mutation possibly produces lower but still sufficient levels of ALAS2 to meet the minimal physiological requirements. The different degrees of ALAS2 reduction after sole int-8-GATA site deletion between Tanimura *et al*. and our studies might result from the different cells lines of different species used

(they deleted GATA site at chrX: 150564815–150564820 in mouse G1E-ER-GATA1 cells (41); while we deleted GATA site at chrX: 55,041,626-55,041,631 in human K562 cells).

In summary, our study suggests that the int-1-GATA site should be examined in patients with XLSA in clinical settings when no known mutation is found in *ALAS2* exons. We believe that gaining more insight based on coding and non-coding regions of the genome will further promote the development of targeted medicine.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Biocytogen for assisting with the generation of the knockout mice and cell lines using TALEN and CRISPR/Cas9. We thank the members of the Zhu and Cheng labs for their insightful discussions during the course of this work.

FUNDING

Ministry of Science and Technology of China [2012CB966603 to X.Z.]; National Key Research and Development Program of China Stem Cell and Translational Research [2016YFA0100600 to T.C., 2016YFA0102300 to L.S.] National Nature Science Foundation of China [81421002 to T.C., X.Z., 81270651 to Y.G., 81400137 to Y.Z., 31471291 to L.S.]; PUMC Youth Fund [3332015185 to L.S.]. Funding for open access charge: Ministry of Science and Technology of China [2012CB966603 to X.Z]. Innovation Fund of Peking Union Medical College for Postgraduates [2005-1002-01-17 to J.Z.].
Conflict of interest statement. None declared.

REFERENCES

1. Bessis,M.C. and Jensen,W.N. (1965) Sideroblastic anaemia, mitochondria and erythroblastic iron. *Br. J. Haematol.*, **11**, 49–51.
2. Fleming,M.D. (2011) Congenital sideroblastic anemias: iron and heme lost in mitochondrial translation. *Hematology Am. Soc. Hematol. Educ. Program.*, **2011**, 525–531.
3. Bergmann,A.K., Campagna,D.R., McLoughlin,E.M., Agarwal,S., Fleming,M.D., Bottomley,S.S. and Neufeld,E.J. (2010) Systematic molecular genetic analysis of congenital sideroblastic anemia: evidence for genetic heterogeneity and identification of novel mutations. *Pediatr. Blood Cancer*, **54**, 273–278.
4. Yamamoto,M., Yew,N.S., Federspiel,M., Dodgson,J.B., Hayashi,N. and Engel,J.D. (1985) Isolation of recombinant cDNAs encoding chicken erythroid delta-aminolevulinate synthase. *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 3702–3706.
5. Cox,T.C., Bottomley,S.S., Wiley,J.S., Bawden,M.J., Matthews,C.S. and May,B.K. (1994) X-linked pyridoxine-responsive sideroblastic anemia due to a Thr388-to-Ser substitution in erythroid 5-aminolevulinate synthase. *N. Engl. J. Med.*, **330**, 675–679.
6. Ducamp,S., Kannengiesser,C., Touati,M., Garcon,L., Guerci-Bresler,A., Guichard,J.F., Vermylen,C., Dochir,J., Poirel,H.A., Fouyssac,F. *et al.* (2011) Sideroblastic anemia: molecular analysis of the ALAS2 gene in a series of 29 probands and functional studies of 10 missense mutations. *Hum. Mutat.*, **32**, 590–597.
7. Bishop,D.F., Tchaikovskii,V., Hoffbrand,A.V., Fraser,M.E. and Margolis,S. (2012) X-linked sideroblastic anemia due to carboxyl-terminal ALAS2 mutations that cause loss of binding to the beta-subunit of succinyl-CoA synthetase (SUCLA2). *J. Biol. Chem.*, **287**, 28943–28955.
8. Harigae,H. and Furuyama,K. (2010) Hereditary sideroblastic anemia: pathophysiology and gene mutations. *Int. J. Hematol.*, **92**, 425–431.
9. Campagna,D.R., de Bie,C.I., Schmitz-Abe,K., Sweeney,M., Sendamarai,A.K., Schmidt,P.J., Heeney,M.M., Yntema,H.G., Kannengiesser,C., Grandchamp,B. *et al.* (2014) X-linked sideroblastic anemia due to ALAS2 intron 1 enhancer element GATA-binding site mutations. *Am. J. Hematol.*, **89**, 315–319.
10. Kaneko,K., Furuyama,K., Fujiwara,T., Kobayashi,R., Ishida,H., Harigae,H. and Shibahara,S. (2014) Identification of a novel erythroid-specific enhancer for the ALAS2 gene and its loss-of-function mutation which is associated with congenital sideroblastic anemia. *Haematologica*, **99**, 252–261.
11. Tsai,S.F., Martin,D.I., Zon,L.I., D'Andrea,A.D., Wong,G.G. and Orkin,S.H. (1989) Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. *Nature*, **339**, 446–451.
12. Evans,T. and Felsenfeld,G. (1989) The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell*, **58**, 877–885.
13. Yamamoto,M., Ko,L.J., Leonard,M.W., Beug,H., Orkin,S.H. and Engel,J.D. (1990) Activity and tissue-specific expression of the transcription factor NF-E1 multigene family. *Genes Dev.*, **4**, 1650–1662.
14. Fujiwara,Y., Browne,C.P., Cunniff,K., Goff,S.C. and Orkin,S.H. (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 12355–12358.
15. Pevny,L., Simon,M.C., Robertson,E., Klein,W.H., Tsai,S.F., D'Agati,V., Orkin,S.H. and Costantini,F. (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, **349**, 257–260.
16. Bresnick,E.H., Katsumura,K.R., Lee,H.Y., Johnson,K.D. and Perkins,A.S. (2012) Master regulatory GATA transcription factors: mechanistic principles and emerging links to hematologic malignancies. *Nucleic Acids Res.*, **40**, 5819–5831.
17. Suwabe,N., Takahashi,S., Nakano,T. and Yamamoto,M. (1998) GATA-1 regulates growth and differentiation of definitive erythroid lineage cells during in vitro ES cell differentiation. *Blood*, **92**, 4108–4118.
18. Omichinski,J.G., Clore,G.M., Schaad,O., Felsenfeld,G., Trainor,C., Appella,E., Stahl,S.J. and Gronenborn,A.M. (1993) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science*, **261**, 438–446.
19. Martin,D.I. and Orkin,S.H. (1990) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes Dev.*, **4**, 1886–1898.
20. Tsang,A.P., Visvader,J.E., Turner,C.A., Fujiwara,Y., Yu,C., Weiss,M.J., Crossley,M. and Orkin,S.H. (1997) FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell*, **90**, 109–119.
21. Merika,M. and Orkin,S.H. (1995) Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Kruppel family proteins Sp1 and EKLF. *Mol. Cell. Biol.*, **15**, 2437–2447.
22. Stumpf,M., Waskow,C., Krotschel,M., van Essen,D., Rodriguez,P., Zhang,X., Guyot,B., Roeder,R.G. and Borggrefe,T. (2006) The mediator complex functions as a coactivator for GATA-1 in erythropoiesis via subunit Med1/TRAP220. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 18504–18509.
23. Wadman,I.A., Osada,H., Grutz,G.G., Agulnick,A.D., Westphal,H., Forster,A. and Rabbitts,T.H. (1997) 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.*, **16**, 3145–3157.
24. Mikkola,H.K., Klintman,J., Yang,H., Hock,H., Schlaeger,T.M., Fujiwara,Y. and Orkin,S.H. (2003) Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia SCL/tal-1 gene. *Nature*, **421**, 547–551.
25. Ema,M., Faloon,P., Zhang,W.J., Hirashima,M., Reid,T., Stanford,W.L., Orkin,S., Choi,K. and Rossant,J. (2003) Combinatorial effects of Flk1 and Tal1 on vascular and hematopoietic development in the mouse. *Genes Dev.*, **17**, 380–393.
26. Xu,Z., Meng,X., Cai,Y., Liang,H., Nagarajan,L. and Brandt,S.J. (2007) Single-stranded DNA-binding proteins regulate the abundance

of LIM domain and LIM domain-binding proteins. *Genes Dev.*, **21**, 942–955.
27. Deng,W., Lee,J., Wang,H., Miller,J., Reik,A., Gregory,P.D., Dean,A. and Blobel,G.A. (2012) Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, **149**, 1233–1244.
28. Love,P.E., Warzecha,C. and Li,L. (2014) Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends Genet.*, **30**, 1–9.
29. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
30. Morgan,K., Kharas,M., Dzierzak,E. and Gilliland,D.G. (2008) Isolation of early hematopoietic stem cells from murine yolk sac and AGM. *J. Vis. Exp.*, **16**, doi:10.3791/789.
31. Giarratana,M.C., Kobari,L., Lapillonne,H., Chalmers,D., Kiger,L., Cynober,T., Marden,M.C., Wajcman,H. and Douay,L. (2005) Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells. *Nat. Biotechnol.*, **23**, 69–74.
32. Shi,L., Cui,S., Engel,J.D. and Tanabe,O. (2013) Lysine-specific demethylase 1 is a therapeutic target for fetal hemoglobin induction. *Nat. Med.*, **19**, 291–294.
33. Li,X., Xu,Z., Du,W., Zhang,Z., Wei,Y., Wang,H., Zhu,Z., Qin,L., Wang,L., Niu,Q. *et al.* (2014) Aiolos promotes anchorage independence by silencing p66Shc transcription in cancer cells. *Cancer Cell*, **25**, 575–589.
34. Zhang,Y., Hu,T., Hua,C., Gu,J., Zhang,L., Hao,S., Liang,H., Wang,X., Wang,W., Xu,J. *et al.* (2014) Rictor is required for early B cell development in bone marrow. *PLoS One*, **9**, e103970.
35. Surinya,K.H., Cox,T.C. and May,B.K. (1998) Identification and characterization of a conserved erythroid-specific enhancer located in intron 8 of the human 5-aminolevulinate synthase 2 gene. *J. Biol. Chem.*, **273**, 16798–16809.
36. Ko,L.J. and Engel,J.D. (1993) DNA-binding specificities of the GATA transcription factor family. *Mol. Cell. Biol.*, **13**, 4011–4022.
37. Merika,M. and Orkin,S.H. (1993) DNA-binding specificity of GATA family transcription factors. *Mol. Cell. Biol.*, **13**, 3999–4010.
38. Nakajima,O., Takahashi,S., Harigae,H., Furuyama,K., Hayashi,N., Sassa,S. and Yamamoto,M. (1999) Heme deficiency in erythroid lineage causes differentiation arrest and cytoplasmic iron overload. *EMBO J.*, **18**, 6282–6289.
39. Stamatoyannopoulos,G., Majerus,P.W., Perlmutter,R.M. and Varmus,H. (2001) Human Hemoglobins: Sickle Hemoglobin and Other Mutants. In: Bunn,HF (ed). *The Molecular Basis of Blood Disease*. 3rd edn. W. B. Saunders Publisher, Philadelphia, pp. 227–274.
40. Bank,A. (1968) Hemoglobin synthesis in beta-thalassemia: the properties of the free alpha-chains. *J. Clin. Investig.*, **47**, 860–866.
41. Tanimura,N., Miller,E., Igarashi,K., Yang,D., Burstyn,J.N., Dewey,C.N. and Bresnick,E.H. (2016) Mechanism governing heme synthesis reveals a GATA factor/heme circuit that controls differentiation. *EMBO Rep.*, **17**, 249–265.
42. Consortium,E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
43. Li,L., Freudenberg,J., Cui,K., Dale,R., Song,S.H., Dean,A., Zhao,K., Jothi,R. and Love,P.E. (2013) Ldb1-nucleated transcription complexes function as primary mediators of global erythroid gene activation. *Blood*, **121**, 4575–4585.
44. Wakabayashi,A., Ulirsch,J.C., Ludwig,L.S., Fiorini,C., Yasuda,M., Choudhuri,A., McDonel,P., Zon,L.I. and Sankaran,V.G. (2016) Insight into GATA1 transcriptional activity through interrogation of cis elements disrupted in human erythroid disorders. *Proc. Natl. Acad. Sci. U.S.A.*, **133**, 4434–4439.