HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2022 February 02.

Activation of γ-globin gene expression by GATA1 and NF-Y in hereditary persistence of fetal hemoglobin

Phillip A. Doerfler[1], **Ruopeng Feng**[1], **Yichao Li**[1], **Lance E. Palmer**[1], **Shaina N. Porter**[2,3], **Henry W. Bell**[4], **Merlin Crossley**[4], **Shondra M. Pruett-Miller**[2,3], **Yong Cheng**[1,5], **Mitchell J. Weiss**[1]

[1]Department of Hematology, St. Jude Children's Research Hospital, Memphis, TN, USA

[2]Department of Cell & Molecular Biology, St. Jude Children's Research Hospital, Memphis, TN, USA

[3]Center for Advanced Genome Engineering, St. Jude Children's Research Hospital, Memphis, TN, USA

[4]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia

[5]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN, USA

Abstract

Hereditary persistence of fetal hemoglobin (HPFH) ameliorates β-hemoglobinopathies by inhibiting the developmental switch from γ-globin (*HBG1*/*HBG2*) to β-globin (*HBB*) gene expression. Some forms of HPFH are associated with γ-globin promoter variants that either disrupt binding motifs for transcriptional repressors or create new motifs for transcriptional activators. How these variants sustain γ-globin gene expression postnatally remains undefined. We mapped γ-globin promoter sequences functionally in erythroid cells harboring different HPFH variants. Those that disrupt a BCL11A repressor binding element induce γ-globin expression by facilitating the recruitment of transcription factors NF-Y to a nearby proximal CCAAT box and GATA1 to an upstream motif. The proximal CCAAT element becomes dispensable for HPFH variants that generate new binding motifs for activators NF-Y or KLF1, but GATA1 recruitment remains essential. Our findings define distinct mechanisms through which transcription factors and

their *cis*-regulatory elements activate γ-globin expression in different forms of HPFH, some of which are being recreated by therapeutic genome editing.

Four major β-like globin genes undergo developmental shifts in expression that are mediated by the formation of DNA loops between their proximal promoters and the locus control region, a powerful upstream enhancer[1,2]. The major β-hemoglobinopathies (β-thalassemia and sickle cell disease (SCD)), become symptomatic after birth as erythroid expression of the γ-globin genes switches to that of the adjacent β-globin gene. This switch is seldom absolute and residual fetal hemoglobin (HbF) levels, determined largely by genetic factors, can influence the severity of these disorders[3,4]. Common genetic variants that reduce *BCL11A* gene expression are associated with elevated red blood cell (RBC) HbF levels and milder β-hemoglobinopathy phenotypes[5–8]. Rare HPFH variants that cause more extreme increases in RBC HbF can eliminate entirely the pathophysiology of co-inherited SCD or β-thalassemia[9,10]. The *BCL11A* gene encodes a repressor protein that binds the γ-globin gene promoters to inhibit their transcription[11,12]. Consistent with this mechanism, multiple HPFH-associated variants disrupt a BCL11A binding motif (TGACC) at positions −114 to −117 relative to the γ-globin transcription start site. These variants include −117 G>A, −114 C>T, −114 C>A, −114 C>G[13–18], and a 13-bp deletion (13 , −102 to −114)[19]. Another variant located 3 base pairs away from the core BCL11A binding motif, −110 A>C, causes HPFH via unknown mechanisms[20,21]. The HPFH-associated BCL11A binding motif overlaps with a CCAAT box element (−111 to −115) and an identical sequence (TGACCAAT) exists at positions −91 to −84 of the γ-globin promoters. These tandem duplicated sequences, referred to as the −85 proximal and −115 distal CCAAT box regions, are conserved among primates. Other γ-globin promoter HPFH variants act by disrupting a GC-rich motif around position −200 to inhibit binding of transcriptional repressor ZBTB7A[11] or by creating new binding sites for the erythroid transcriptional activators KLF1[22], TAL1[23], or GATA1[24].

While it is now clear that BCL11A and its cognate DNA motif in the distal CCAAT box of the γ-globin promoter regulate post-natal gene silencing, how interference with this repressive mechanism by HPFH variants leads to transcriptional activation is not fully understood. The BCL11A motif overlaps with the distal CCAAT box named after a common DNA motif that recruits transcriptional activators to gene promoters. Most likely, the major CCAAT box-binding factor for globin gene transcription is the nuclear transcription factor-Y (NF-Y), a ubiquitous trimeric protein with chromatin opening activity[25–29]. Current evidence supports a mechanism whereby NF-Y activates γ-globin transcription through the proximal CCAAT box during normal fetal erythropoiesis and in some forms of HPFH. Coincident with the shift to adult erythropoiesis, BCL11A associated with the NuRD co-repressor complex binds the distal CCAAT box, eliminates NF-Y and inhibits transcription[11,12,30–32].

The γ-globin promoter also includes 2 GATA motifs separated by a conserved octamer. This sequence between positions −170 to −195 binds a single molecule of the erythroid transcription factor GATA1 in electrophoretic mobility assays (EMSA) and contributes to erythroid promoter activity in reporter assays[28]. A −175 T>C variant in the proximal GATA

motif causes HPFH by creating a *de novo* binding site for the erythroid transcription factor TAL1[23]. The roles of GATA1 in the transcriptional activation of γ-globin genes during normal development and HPFH are not fully understood.

To gain further insights into the positive regulation of γ-globin gene expression in HPFH, we mapped the promoter CCAAT box and bipartite GATA motif regions functionally at the nucleotide level via CRISPR/Cas9-mediated genome editing and base editing in an adult-like erythroid cell line and in primary erythroblasts. We show that gene activation caused by disruption of the BCL11A binding motif in the distal −115 distal CCAAT box occurs through the additive effects of recruiting NF-Y to the −85 proximal CCAAT box and GATA1 to the bipartite GATA motif. Additionally, we show that the previously unexplained HPFH variant −110 A>C causes *de novo* recruitment of NF-Y to the −115 distal CCAAT box. In contrast, NF-Y binding to the γ-globin promoter is not required at all for γ-globin induction by HPFH variants that create new binding sites for KLF1 (−198 T>C) or GATA1 (−113 A>G) although recruitment of GATA1 to the bipartite motif remains essential.

Results

HPFH variants induce HbF in a human erythroid cell line.

We created HPFH variants in HUDEP-2 cells, immortalized human erythroid progenitors that express mainly HbA (α2β2) but can be stimulated to produce HbF by the introduction of HPFH-like mutations[11,12,24,33–37]. The tandem duplicated *HBG1* and *HBG2* genes are nearly identical, which complicates mutational analysis. Therefore, we used CRISPR/Cas9-mediated non-homologous end joining to generate HUDEP-2 $^{εγδβ/GγAγ}$ cells, which lack the entire β-like globin gene cluster on one chromosome (εγδβ), and contain a single in-frame *HBG2-HBG1* (*GγAγ*) fusion gene on the other (Extended Data Fig. 1). This fusion gene, henceforth referred to as *HBG*, resembles a natural variant with normal developmental regulation[37–40]. HUDEP-2 $^{εγδβ/GγAγ}$ cells grow normally, express minimal HbF, and exhibit slightly accelerated erythroid maturation *in vitro* compared to wild-type HUDEP-2 cells (Extended Data Fig. 2). In-situ Hi-C results confirmed that the single modified β-like globin locus in HUDEP-2 $^{εγδβ/GγAγ}$ cells exhibits a normal chromatin structure (Extended Data Fig. 3). We introduced 5 different −115 CCAAT box HPFH variants and 2 control mutations separately into the *HBG* promoter of HUDEP-2 $^{εγδβ/GγAγ}$ cells by Cas9-mediated homology-directed repair (HDR) (Fig. 1a, b). All HPFH variants caused significantly increased HbF expression (Fig. 1c and Extended Data Fig. 4a) that was associated with chromatin opening detected by assay for transposase-accessible chromatin using sequencing (ATAC)-seq[41] analysis (Fig. 1d and Extended Data Fig. 4b).

GATA1 recruitment activates γ-globin expression.

Disruption of the motif that recruits BCL11A and its associated NuRD co-repressor complex likely results in chromatin alterations that accommodate the binding of transcriptional activators. We first investigated whether GATA1 binding to the −189 motif facilitates transcriptional activation by the HPFH variants (Fig. 2a). Chromatin immunoprecipitation revealed GATA1 binding to the γ-globin promoter in HUDEP-2 $^{εγδβ/GγAγ}$ cells with HPFH variants, but not in cells with the WT promoter (Fig. 2b). Similarly, GATA1

occupancy of the *HBG1* and *HBG2* promoters was stronger in fetal-like, HbF-expressing HUDEP-1 cells compared to HUDEP-2 cells[37] and in fetal liver proerythroblasts compared to proerythroblasts derived from adult CD34+ hematopoietic stem and progenitor cells (HSPCs; Extended Data Fig. 5). Thus, GATA1 binding to the γ-globin promoter is associated with transcriptional activation. The relatively low magnitude of the observed GATA1 ChIP-seq signal could reflect dynamic GATA1 occupancy (high on-off rates) and/or masking of the GATA1 antibody epitope by nearby chromatin-bound factors. We tested the −189 GATA motif functionally by introducing the mutation −186 C>T, which is predicted to disrupt GATA1 binding (Fig. 2a). This mutation caused reductions in the %HbF in all HPFH clones tested (Fig. 2c and Extended Data Fig. 6a) and in the %HbF immunostaining cells (F-cells) associated with the −113 A>G HPFH variant (Fig. 2d and Extended Data Fig. 6b). We performed ChIP-seq to compare GATA1 occupancy in the different mutant clones. As experimental spike-in normalization is not well-established for transcription factor ChIP-seq, we used S3norm to normalize sequencing depths and signal-to-noise ratios in silico[42]. In two biological replicate experiments, the −186 C>T mutation caused a reduction of GATA1 occupancy at the *HBG* promoter (Fig. 2d, e and Extended Data Fig. 6c). Together, these findings show that GATA1 binds the bipartite motif to activate γ-globin transcription in HPFH. The −186 C>T mutation did not increase HbF levels in HUDEP-2 $^{εγδβ/GγAγ}$ cells with a WT *HBG* promoter, supporting a role in activation rather than repression (Extended Data Fig. 7a, b) [29].

To test whether the bipartite −189 GATA motif stimulates γ-globin gene transcription in normal erythroid progenitors, we electroporated umbilical cord blood (UCB)-derived CD34+ HSPCs with the adenosine base editor ABE7.10[43,44] and targeting guide RNA (gRNA), followed by *in vitro* erythroid differentiation. Adenosine base editors create A>G mutations within a 5 nucleotide window specified by the targeting gRNA[43,44]. The overall editing frequency was approximately 58%, with most base pair alterations predicted to disrupt GATA1 binding to its consensus motif (Fig. 3a, b). Editing of the GATA motif in CD34+ HSPCs resulted in a significant reduction of HbF in pooled erythroid progeny (Fig. 3c). To assess this at a clonal level, we seeded base-edited HSPCs into methylcellulose with erythroid cytokines and analyzed burst forming unit-erythroid (BFU-E) colonies. In two experiments using UCB CD34+ cells from different donors, the HbF levels in individual BFU-E colonies correlated inversely with base editing frequencies in the core GATA motif (Fig. 3d, e). Specifically, HbF levels were reduced by 33% or 56% in colonies with 90% disrupted GATA1 motifs compared to those with 10% disrupted motifs (Fig. 3f). Thus, the −189 GATA motif participates in normal γ-globin gene activation.

NF-Y recruitment activates γ-globin expression.

Disruption of the −189 GATA motif caused an approximately 40% reduction of HbF expression in HUDEP-2 $^{εγδβ/GγAγ}$ cells with HPFH variants near the −115 distal CCAAT box (Fig. 2c), indicating that additional positive-acting, *cis*-regulatory elements exist. Previous reports suggest that the transcription factor NF-Y activates γ-globin expression via binding to the −85 proximal CCAAT box (Fig. 4a)[25,27,32,45,46]. In support, ChIP-seq analysis revealed NF-Y occupancy at the *HBG* promoter in HUDEP-2 $^{εγδβ/GγAγ}$ cells harboring the 13 or −110 A>C HPFH variants but not in cells with a WT promoter (Fig.

4b). A relatively weak signal for NF-Y binding was observed in cells with the −113 A>G variant. Standard ChIP-seq analysis cannot resolve NF-Y binding to the −85 proximal vs. −115 distal CCAAT boxes. The distal CCAAT motif is eliminated by the 13Δ and −113 A>G HPFH variants, suggesting that NF-Y occupies the proximal element. In contrast, the −110 A>C variant is predicted by motif occurrence analysis[47] to enhance NF-Y binding to the distal CCAAT box (Fig. 4c). By ChIP-seq, the −110 A>C variant was associated with approximately 5-fold greater signal for NF-Y binding compared to other HPFH variants (Fig. 4b). In replicate electrophoretic mobility shift analysis (EMSA) experiments, NF-Y binding to a radiolabeled distal CCAAT box probe with the −110 A>C variant was 50% and 25% greater than binding to the WT probe (Fig. 4d and Extended Data Fig. 8a). This finding was supported by EMSA competition studies between unlabeled WT or −110 A>C probes (Fig. 4e). Motif occurrence analysis also predicted that the −110 A>C substitution reduces the affinity of BCL11A for its cognate motif at the distal CCAAT box (Extended Data Fig. 8b). However, this possibility was not supported by competitive EMSA analysis using BCL11A zinc fingers 4–6 (Extended Data Fig. 8c). Together, these results suggest that the −110 A>C HPFH variant stimulates γ-globin expression by recruiting NF-Y to the −115 distal CCAAT box.

GATA1 and NF-Y cooperate in γ-globin gene activation.

We performed mutational analysis to analyze further the functional effects of NF-Y and GATA1 binding motifs on γ-globin gene activation in HPFH (Fig. 5a). In HUDEP-2 $^{εγδβ/GγAγ}$ cells with the 13Δ HPFH variant, which eliminates both BCL11A and NF-Y consensus binding motifs in the distal CCAAT box, deletion of the −85 proximal CCAAT box reduced HbF levels by approximately 60% (13Δ/−85Δ; Fig 5a, b). Most of the remaining HbF expression in double mutant clones was eradicated by disruption of the −189 GATA motif (−186T/13Δ/−85Δ). Consistent with these findings, ChIP-seq analysis revealed that deletion of the proximal CCAAT box (85Δ) or the −186 C>T mutation eliminated occupancy of NF-Y or GATA1, respectively, at the γ-globin promoter in HUDEP-2 $^{εγδβ/GγAγ}$ cells harboring the 13Δ HPFH variant (Fig. 5c). Thus, transcriptional activation in the 13Δ variant, and likely other HPFH variants that disrupt the BCL11A binding motif near the −115 distal CCAAT box, is achieved additively by GATA1 binding to the −189 GATA element and NF-Y binding to the −85 proximal CCAAT box.

The −85 CCAAT box is dispensable for some HPFH variants.

Remarkably, deletion of the −85 proximal CCAAT box (Δ85) had no effect on γ-globin induction by the −113 A>G or −110 A>C variants (Fig. 6a, b). The −113 A>G variant disrupts the distal CCAAT box NF-Y motif and creates a new GATA1 motif[24]. In this case, NF-Y dependency is most likely substituted for by GATA1 occupancy at the same region, which could explain reduced NF-Y binding observed at the γ-globin promoter in HUDEP-2 $^{εγδβ/GγAγ}$ cells harboring the −113 A>G variant (see Fig. 4b). In contrast, the −110 A>C variant enhances the affinity of the distal CCAAT box motif for NF-Y, which likely displaces BCL11A, possibly explaining the amplified ChIP-seq signal for NF-Y binding (see Fig. 4b).

Next, we generated the −198 T>C HPFH variant in HUDEP-2 $^{εγδβ/GγAγ}$ cells and observed significant induction of HbF (Fig. 6a, c). This variant creates a new binding site for the transcriptional activator KLF1, which displaces the repressor protein ZBTB7A[22]. Disruption of the −189 GATA binding motif (−186 C>T mutation) fully eliminated HbF induction by the −198 T>C HPFH variant (Fig. 6c), while disruption of the NF-Y binding motif at the distal CCAAT box caused a slight increase in HbF expression.

Discussion

Here we elucidate distinct mechanisms for γ-globin gene activation by different non-deletional HPFH variants that either disrupt promoter binding motifs for the transcriptional repressors BCL11A or ZBTB7A, or create new binding motifs for transcriptional activators (Fig. 7)[48,49]. Our findings show that GATA1 and NF-Y activate γ-globin expression, consolidating previous studies implicating these transcription factors and their respective *cis* elements in the expression of γ-globin and other hematopoietic genes[25–28,32,50–53]. Collective data suggest that during normal fetal life, GATA1 and NF-Y activate transcription cooperatively though their respective binding motifs in the γ-globin promoter. Recent studies indicate that NF-Y binds specifically to the −85 proximal CCAAT box[12,32]. Around birth, BCL11A accumulates and occupies its cognate motif near the −115 distal CCAAT box, displacing NF-Y either directly[32] and/or by establishing a closed chromatin state through recruitment of the NuRD co-repressor complex, which has histone deacetylase and nucleosome remodeling activities[54]. Concomitantly, GATA1 occupancy is eliminated, likely through chromatin effects imparted by the BCL11A-NuRD complex and/or ZBTB7A, which also binds the NuRD complex (Fig. 7a). Most non-deletional HPFH variants near the −115 distal CCAAT box disrupt the BCL11A binding motif, allowing GATA1 and NF-Y to remain chromatin-bound (Fig. 7b). It is unknown why BCL11A and NF-Y specifically occupy the distal and proximal CCAAT boxes, respectively, as each region harbors identical overlapping binding motifs for each transcription factor. Presumably, this selectivity is maintained *in vivo* by flanking DNA sequences, local histone modifications and regional occupancy of other DNA binding proteins[30,32,55–57].

Here we show that the HPFH variant −110 A>C stabilizes NF-Y binding to the −115 distal CCAAT box (Fig. 7c). It will be interesting to determine whether the affinity is further augmented by transcription factor SP2, which potentiates NF-Y binding to DNA, particularly at promoters with tandem CCAAT boxes[58,59]. It is possible that NF-Y also occupies the proximal CCAAT box adjacent to the −110 A>C variant, although mutational analysis indicates that this is dispensable for transcriptional activation. Regardless, binding of NF-Y to the distal CCAAT box displaces BCL11A and obviates the normal requirement for NF-Y occupancy at the proximal CCAAT box. Through an analogous mechanism, the HPFH variant −113 A>G creates a new distal CCAAT box binding site for GATA1[24], which displaces BCL11A, facilitating the recruitment of GATA1 to its upstream motif (Fig. 7d). In this case, promoter occupancy of NF-Y is no longer required for gene activation. Interestingly, disruption of the −189 GATA motif in −113 A>G HUDEP-2 $^{εγδβ/GγAγ}$ cells reduced both %HbF expression and %F-cells, converting a pancellular distribution to a heterocellular one (Fig. 2c, Extended Data Fig. 6b). While the associated mechanism is unknown, it is possible that loss of GATA1 binding creates epigenetic alterations that

enhance the ability of BCL11A to outcompete GATA1 at the modified −115 CCAAT box, increasing the probability of stochastic gene silencing. This may be potentiated by methylation of the adjacent cytosine (−114), which reduces the affinity for GATA1 binding[60]. Another HPFH variant, −198 T>C, creates a new binding site for KLF1[22], which displaces ZBTB7A to facilitate GATA1-dependent, NF-Y-independent transcriptional activation (Fig. 7e). Overall, mutational analysis of UCB CD34[+] cell-derived erythroblasts and HUDEP-2 $^{εγδβ/GγAγ}$ cells show that GATA1 binding to its −189 motif contributes to γ-globin transcription during normal development and in all forms of non-deletional HPFH tested, most likely by cooperating with NF-Y or alternative transcription factors to stabilize looping with the LCR[61]. Thus, it will be interesting to investigate how HPFH variants that create de novo transcription factor binding sites alter protein contacts within the loop and/or its position at the γ-globin promoter. Such studies may require new DNA proximity assays with higher resolution that exceed the current kilobase limits [62].

Overall, our findings support a general model in which γ-globin expression is regulated cooperatively by two pairs of closely spaced DNA binding motifs that recruit transcriptional activators and repressors: the proximal and distal CCAAT boxes, which bind NF-Y and BCL11A respectively, and the upstream GATA1 and ZBTB7A motifs. Competition between activators and repressors, both directly and indirectly through antagonistic epigenetic effects, is central to this model. In this regard, steric effects may prohibit simultaneous binding of GATA1 and ZBTB7A to their closely spaced motifs, similar to what has been proposed for BCL11A and NF-Y at the tandem CCAAT boxes[32]. Cross-regulation between the paired activator-repressor motifs is evidenced by our findings that disruption of the γ-globin promoter BCL11A binding motif facilitates GATA1 occupancy to its motif located approximately 60 nucleotides upstream. Another layer of control occurs at the level of gene expression, whereby the GATA1-regulated erythroid transcription factor KLF1 activates the *ZBTB7A* and *BCL11A* genes[63,64].

Insights into developmental regulation of globin gene switching and mechanisms of HPFH gained from our study have practical implications for autologous hematopoietic stem cell therapies intended to treat β-hemoglobinopathies. For example, targeted disruption of the BCL11A or ZBTB7A binding motifs in the γ-globin promoter by genome editing can induce RBC HbF to potentially therapeutic levels[33,35,36,65,66]. However, the size of genome editing-induced deletions is uncontrolled and unpredictable. Deletions as small as 10–30 base pairs originating from end resection of double-stranded DNA breaks targeting the BCL11A or ZBTB7A binding motifs could disrupt adjacent motifs for transcriptional activators NF-Y or GATA1, potentially resulting in heterocellular HbF induction. This problem could be avoided by judicious screening of genome editing nucleases and targeting gRNAs, or with base-editors, which introduce precise nucleotide alterations with minimal indel formation[67].

Methods

Cell culture

All cell culture was performed at 37°C with 5% CO_2 in a water jacketed incubator. HUDEP-1 and HUDEP-2 cells were maintained in StemSpan serum-free expansion medium

(SFEM; StemCell Technologies) supplemented with 1 μM dexamethasone, 1 μg/mL doxycycline, 50 ng/ml human SCF (R&D Systems), 3 U/mL EPO (Amgen), and 1% penicillin/streptomycin. Differentiation of HUDEP-2 cells[37] was conducted using a 2-phase protocol. Phase 1 (days 0–3): IMDM supplemented with 2% FBS, 3% human blood type AB serum (Atlanta Biologicals), 1% penicillin/streptomycin, 3 U/mL EPO, 10 μg/mL insulin, 3 U/mL heparin, 1 mg/mL holo-transferrin (Millipore Sigma), 1 μg/mL doxycycline, and 50 ng/mL human SCF. Phase 2 (days 4–7): phase 1 medium without SCF. Maturation of erythroid cells was monitored on days 0, 3, and 7 via flow cytometry for FITC-CD235a (BD Pharmigen; 1:100 dilution), BV421-CD49d (BioLegend; 1:20 dilution), and APC-Band3 (gift from Xiuli An, New York Blood Center; 1:20 dilution). Flow cytometry gating strategies are shown in Extended Data Figure 9a.

Cord blood human CD34+ cells were obtained from four de-identified healthy donors (Key Biologics, Lifeblood) and enriched by immunomagnetic bead selection using an AutoMACS instrument (Miltenyi Biotec). These deidentified samples were exempt from St. Jude Children's Research Hospital Institutional Review Board approval. Cryopreserved CD34+ cells were thawed and pre-stimulated for 48 hours in SFEM supplemented with 100 ng/mL SCF, FLT3-L, and TPO (R&D Systems) prior to electroporation. Cells were grown in culture in complete SFEM for 48 hours following electroporation then either seeded 500 cells/mL in human methylcellulose (H4230; StemCell Technologies) with 2 U/mL EPO (Amgen), 10 ng/mL SCF, and 1 ng/mL IL-3 (R&D Systems) or collected after an additional 48 hours for genomic DNA extraction to measure base editing frequency. Individual BFU-E colonies were picked after 14 days of culture. Erythroid differentiation was conducted using a 2-phase protocol. Phase 1 (days 0–5): IMDM (Thermo) supplemented with 20% FBS, 1% penicillin/streptomycin, 20 ng/mL SCF, 1 ng/mL IL-3 (R&D Systems), and 2 U/mL EPO (Amgen). Phase 2 (days 5–10): IMDM supplemented with 20% FBS, 1% penicillin/ streptomycin, 2 U/mL EPO, and 0.2 mg/mL holo-transferrin (Millipore Sigma).

COS-7 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS, Bovogen Biologicals), and 1% penicillin-streptomycin-glutamine (PSG, Gibco). Cells were lifted for passaging by incubation in 0.05% Trypsin-EDTA, (Gibco) at 37°C for 5 minutes. K562 cells were cultured in RPMI 1640 media (Gibco) supplemented with 10% (v/v) FBS and 1% PSG.

Genome editing

Purified recombinant Cas9 protein was obtained from Berkeley Macrolabs. Purified recombinant ABE7.10 protein was a kind gift of Mark Osborn (U. Minnesota) and David Liu (HHMI/Harvard). Chemically modified single guide RNAs (sgRNA) were synthesized by Synthego with 2′-O-methyl 3′-phosphorothioate modifications between the 3 terminal nucleotides at both the 5′ and 3′ ends. Ribonucleoprotein complexes (RNPs) were formed by incubating Cas9 (32 pmol/100,000 cells) or ABE7.10 (50 pmol/100,000 cells) with sgRNAs at 1:2 or 1:3 molar ratio, respectively.

Cells were washed in PBS, resuspended in the manufacturer provided buffers for cell lines or primary cells (Thermo Fisher Scientific), mixed with ribonucleoprotein complexes and

electroporated using program 12 (HUDEP) or program 24 (CD34$^+$) of a Neon Transfection System (Thermo Fisher Scientific).

For homology-directed repair, 5 μM single-stranded oligo DNA nucleotides (ssODN) harboring the desired mutation were added immediately prior to electroporation. 2′-O-methyl 3′-phosphorothioate modifications between the 2 terminal nucleotides at both the 5′ and 3′ ends were included in the ssODNs. Clonal cell lines were derived following single-cell sorting into 96- or 384-well plates using a SH800 cell sorter (Sony Biotechnology). Editing and base editing efficiency were determined using CRIS.py[35,68]. Sequences of sgRNAs and ssODNs are provided in Supplementary Table 1 and primers are provided in Supplementary Table 2.

HbF quantification

Undifferentiated HUDEP cells were fixed with 0.05% glutaraldehyde, permeabilized with 0.1% Triton X-100, and stained with anti-human HbF-APC (1:20 dilution; ThermoFisher) for flow cytometry. The flow cytometry gating strategy is shown in Extended Data Figure 9b. Data were collected and analyzed using BD FACSDiva (v9) and FlowJo (v10) software. HUDEP cells differentiated for 7 days or single BFU-E colonies were lysed with hemolysate reagent (Helena Laboratories) and analyzed using ion-exchange columns on a Prominence HPLC System (LabSolutions Software v.5.81 SP1, Shimadzu Corporation). Proteins eluted from the column were identified at 220 and 415 nm with a diode array detector. The relative amounts were calculated from the area under the 415 nm peak and normalized based on the DMSO control. %HbF = [HbF/(HbA + HbF)] × 100.

Fluorescence in situ *hybridization (FISH)*

A 5.2 kb probe targeted to the region between the *HBG1* and *HBG2* promoters was labeled with a red-dUTP (AF594; Molecular Probes; chr11:5271371–5275869) and purified BAC DNA from chromosome 11 was labeled with a green-dUTP (AF488; Molecular Probes; hg19 chr11:5147629–5265447) by nick translation. The probes were hybridized to interphase and metaphase cells using routine cytogenetic methods in a solution containing 50% formamide, 10% dextran sulfate, and 2X SCC. The cells were then stained with 4, 6-diamidino-2-phenylindole (DAPI) and analyzed for signals representing the potentially deleted region (red) and chr11 (green).

Chromatin Immunoprecipitation (ChIP)

ChIP was performed as previously described[69] with the following modifications. Briefly, 2×10^7 HUDEP-2 $^{εγδβ/GγAγ}$ cells were used for each immunoprecipitation. Cells were cross-linked with 1% formaldehyde for 10 minutes on a rotary shaker at room temperature. The reaction was quenched for 5 minutes at room temperature with a final concentration of 125 mM glycine. Cells were lysed and the chromatin sonicated on ice using a Branson 250 micro-tip sonicator with the power settings of 100% duty cycle, 10 second pulses for 2 minutes, with 90 seconds on ice between pulses. The sonicated chromatin was pre-cleared overnight at 4°C using protein A/G agarose beads (Thermo: 20334 and 20399). Immunoprecipitations were performed using 10 μg of antibodies against GATA1 (Abcam: ab11852) and NF-YB (Santa Cruz Biotechnology: sc-376546x). Following elution, DNA-

protein complexes were treated with RNase for 30 minutes at 37°C followed by proteinase K treatment for 30 minutes at 45°C and overnight at 60°C with shaking. DNA was purified using a Qiagen MinElute kit.

ChIP-seq analysis

DNA libraries were prepared using NEBNext Ultra II DNA Library Prep (NEB: E7645) or KAPA HyperPrep (Roche: 07962363001) kits for Illumina sequencing. All fastq files generated in HUDEP-2 $\varepsilon\gamma\delta\beta/G\gamma A\gamma$ cells were mapped to a modified GRCh37/hg19 that masks the reference genome between the IVSII gRNA sequences within *HBG1* and *HBG2* (chr11:5269886–5275238). The masked reference was generated using bedtools maskfasta (v2.25.0)[70]. ChIP-seq analysis was performed using the HemTools pipeline chip_seq_pair (see "Code Availability" below). Fastq files were mapped using BWA mem (v0.7.16a)[71]. Duplicated reads were marked and removed using samtools (v0.17)[72]. Genome signal tracks (.bw files) were generated using deepTools bamCoverage (v3.2.0)[73]. ChIP-seq peaks were called using MACS2 (v2.1.1)[74] with "-f BAMPE". Normalized signal tracks were generated using S3norm (v2)[42]. De-duplicated bw files were used as input for conversion to bed files using bigWigAverageOverBed (v4) along with a bed file containing coordinates representing 50 bp bins across the human hg19 genome. The resulting bed file was converted to a bedgraph file as input to the S3norm pipeline using default options except for the '-r mean' option. Input signal files were included as a sample file (in addition to serving as the control signal). With the normalized read counts (S3norm_rc_bedgraph folder), the control normalized counts were subtracted from the normalized sample counts. The resulting bedgraph files were converted to bigwig files using bedGraphToBigWig.

MACS2-called peaks were merged for all the samples in one comparison using bedtools (v2.25.0). Read count matrix was extracted from S3norm normalized read counts using bigWigAverageOverBed.

ATAC-seq and analysis

ATAC-seq[75] was performed using 60,000 HUDEP-2 $\varepsilon\gamma\delta\beta/G\gamma A\gamma$ cells with the desired variant in the −115 distal CCAAT box. Cells were lysed and nuclei were resuspended using the Illumina Tagment DNA TDE1 Enzyme and Buffer Kits (Illumina, 20034197). Following purification, libraries were amplified using NEBnext PCR master mix and custom Nextera PCR primers 1 and 2 for 5 cycles. The degree of library amplification to reduce GC and size bias was determined using qPCR using SYBR green reagents (Thermo: S7567). A total of 13–15 cycles were performed and libraries were purified using a QIAGEN PCR purification kit.

ATAC-seq analysis was performed using the HemTools pipeline atac_seq (see "Code Availability" below). Raw reads were trimmed to remove Tn5 adaptor sequence using skewer (v0.2.2)[76] and were then mapped to hg19 using BWA mem (v0.7.16a). Duplicated and multi-mapped reads were removed using samtools (v0.17). ATAC-seq peaks were called using MACS2 (v2.1.1) with the following parameters "macs2 callpeak --nomodel --shift −100 --extsize 200". BigWiggle files were generated using DeepTools bamCoverage (v3.2.0) with "--centerReads".

Hi-C and analysis

In situ Hi-C was performed as previously described[77] with the following modifications. Briefly, 5×10^6 HUDEP-2 or HUDEP-2 $^{\epsilon\gamma\delta\beta/G\gamma A\gamma}$ cells were cross-linked with 1% formaldehyde for 10 minutes on a rotary shaker at room temperature. The reaction was quenched for 5 minutes at room temperature with a final concentration of 200 mM glycine. Cells were lysed and the chromatin digested with *Mbo*I (NEB: R0147). Following biotin fill-in, proximity ligation, and reverse cross-linking, the DNA was sonicated using a Covaris M220 sonicator with the following settings: 50W peak incident power, 20% duty factor, 200 cycles/burst, 90 seconds. The sheared DNA was purified using AMPure XP beads (Beckman Coulter: A63881). Following biotin pull-down, DNA libraries were prepared using NEBNext Ultra II DNA Library Prep (NEB: E7645) for Illumina sequencing.

Hi-C analysis was performed using the HemTools pipeline hicpro_batch (see "Code Availability" below). HiC-Pro (v2.11.1)[78] was used with default parameters; for global read mapping: --very-sensitive -L 30 --score-min L,−0.6,−0.2 --end-to-end --reorder; for local read mapping: --very-sensitive -L 20 --score-min L,−0.6,−0.2 --end-to-end --reorder; cutoffs for minimal and maximal values are not defined for FRAG_SIZE, INSERT_SIZE, and CIS_DIST; for read pair filtering: RM_SINGLETON = 1, RM_MULTI = 1, RM_DUP = 1. Then all Hi-C data were down sampled to 200 million valid read pairs based on .allValidPairs files. HiC-Pro was used for iced matrix generation with default parameters of BIN_SIZE = 10000, MAX_ITER = 100, FILTER_LOW_COUNT_PERC = 0.02, FILTER_HIGH_COUNT_PERC = 0, EPS = 0.1. HicFindTADs from HiCExplorer (v3.5.1)[79] was used for topologically associating domain (TAD) calling with default parameters after converting HiC-Pro iced matrix to H5 format (hicConvertFormat). Genome-wide TAD correlations between HUDEP-2 and HUDEP-2 $^{\epsilon\gamma\delta\beta/G\gamma A\gamma}$ samples were evaluated based on the union and the intersect of the called TAD boundaries.

In Silico *Analysis of TF Binding Affinity*

Known BCL11A and NF-Y motifs were downloaded from the Homer Motif Database (http://homer.ucsd.edu/homer/motif/motifDatabase.html). We performed FIMO[47] (from MEME suite, v5.1.0) motif scanning on the wild-type sequence TGACCAATAGCC and all individual nucleotide mismatches. FIMO provides a *P*-value computation based on the nucleotide frequency in position weight matrices and the observed DNA sequences, which can be used as an affinity score for TF binding. Changes in percent binding were normalized relative to the *P*-values derived from the wild-type sequence.

Electrophoretic Mobility Shift Assay

Oligonucleotides used as radiolabeled probes are listed in Supplementary Tables 3-5. The sense strand for each probe was labeled with ^{32}P from $\gamma–^{32}$P ATP (Perkin Elmer) using T4 PNK (NEB), annealed with the antisense strand by slow cooling from 100°C to room temperature, then purified using quick spin columns (Roche). Unlabeled probes for cold competition assays were annealed by slow cooling from 100°C to room temperature. K562 cell nuclear extracts were used to evaluate NF-Y binding. To assess BCL11A binding, we used nuclear extract from COS-7 cells engineered to express BCL11A zinc fingers 4 to 6. Empty extract from COS-7 cells without the BCL11A expression construct was used to

identify background bands caused by endogenous protein binding. Antibodies for NF-YA (200 ng per lane; Santa Cruz Biotechnology: G-2 sc-17753), and BCL11A (1 μg per lane; Novus Biologicals: NB600–261) were used for supershift studies. For cold competition assays, annealed unlabeled probe was added to the sample and incubated for 10 minutes at room temperature before addition of the labeled probe. Complexed samples were loaded on 6% native polyacrylamide gel in TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA). Electrophoresis was performed at 4°C and 250 V for 1 hour and 40 minutes, and vacuum dried before exposing a FUJIFILM BAS Cassette2 phosphor screen overnight. Imaging was performed on a GE Typhoon FLA 9500. Quantification of images was performed using Image Lab Software (Bio-Rad, v6.0.1).

COS-7 cell transfections and nuclear extraction

COS-7 cells were used for transient over-expression of transcription factors. Cells were transfected in 100 mm plates with 5 μg of mammalian expression plasmid using FuGENE® 6 (Promega) according to the manufacturer's instructions. Mammalian expression plasmids used are listed in Supplementary Table 6. Transfected cells were incubated at 37°C for 48 hours before harvest. Nuclear extractions were performed as previously described[80] with the following modifications. K562 cells were collected by centrifuging at 300 x g for 5 minutes and washed in PBS. COS-7 cells were washed in PBS before harvesting by scraping and centrifugation. Cells were resuspended in 10X packed cell volume (PCV) of complete hypotonic lysis buffer (10 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 5 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF), 10 μg/mL aprotinin, 10 μg/μL leupeptin). The cells were incubated on ice for 10 minutes then thoroughly vortexed before pelleting in a quick spin centrifuge and discarding the supernatant. The pellet was resuspended in 2–3x PCV of complete extraction buffer (20 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 0.42 M NaCl, 0.2 mM EDTA, 25% glycerol) with 5 mM DTT, 1 mM PMSF, 10 μg/mL aprotinin, and 10 mg/μL leupeptin and incubated on ice for 20–30 minutes. The suspension was then centrifuged at 16,000 x g for 3 minutes at 4°C and the supernatant recovered.

Statistical Analysis

The minimum, maximum, median, and interquartile ranges are shown for graphs containing boxplots. The mean and standard deviation are shown for graphs containing bar plots. An uncorrected two-tailed unpaired t-test was performed to assess statistical significance between 2 groups. Ordinary on-way ANOVA was used to assess statistical significances for >2 groups. ANOVA post hoc corrections for multiple comparisons using statistical hypothesis testing (Dunnett when comparing a sample mean to a control mean or Tukey when comparing the mean of each sample with the mean of every other sample) were performed where indicated. Linear regression and Pearson's correlation was used to measure the relationship between two linear variables without multiple testing correction. All analyses were performed using GraphPad Prism 9.

Data Availability

Data sets used in this study are listed in Supplementary Table 7. Raw and processed sequencing data generated in this study are available from the NCBI Gene Expression Omnibus (GSE152338). Source data are provided with this paper.
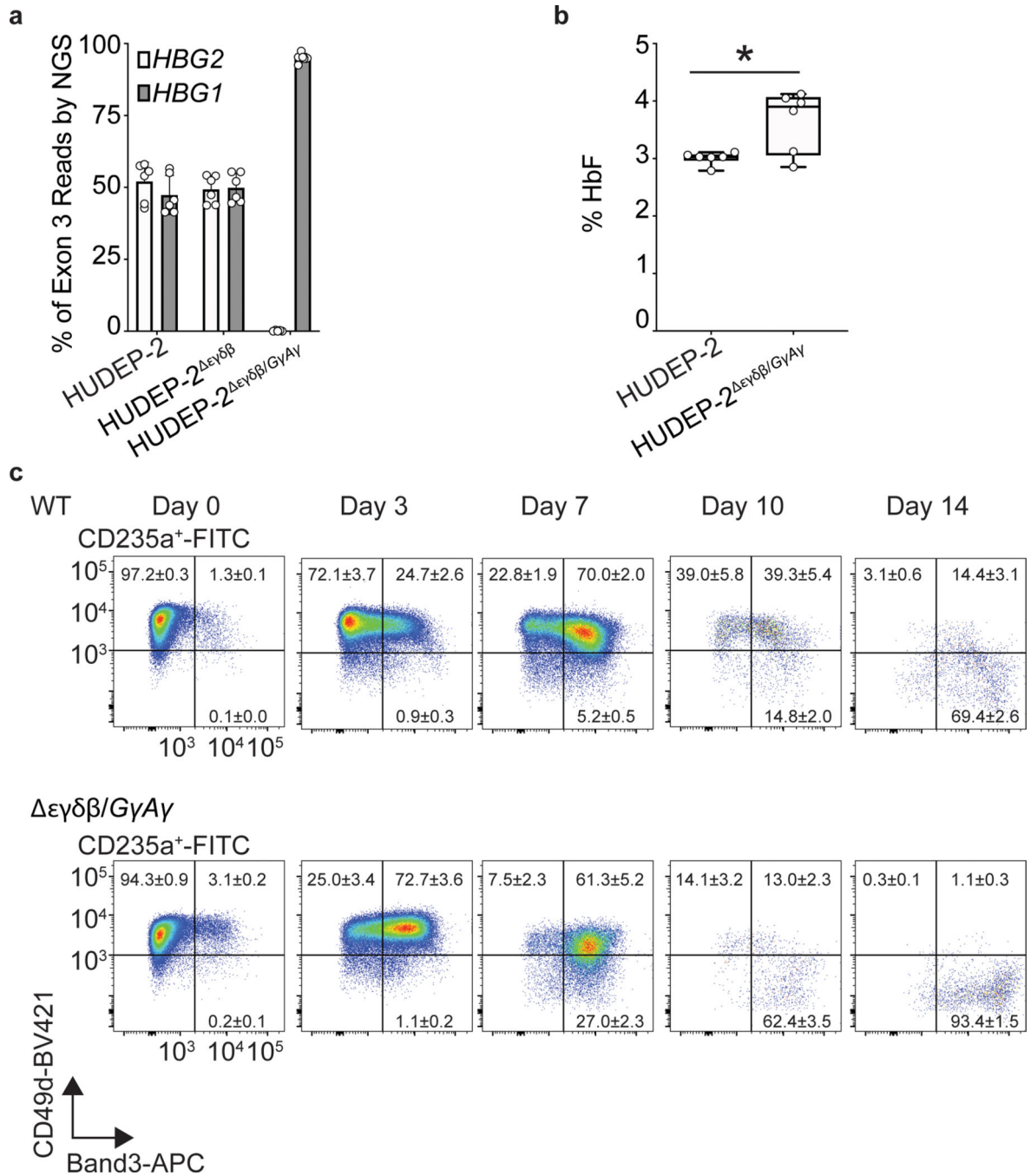
Code Availability

The code used to perform ATAC-seq (HemTools atac_seq), ChIP-seq (HemTools chip_seq_pair), and Hi-C analyses (hicpro_batch.py) is available at https://github.com/YichaoOU/HemTools and at https://doi.org/10.5281/zenodo.4783657. Pipeline documentation is available at: https://hemtools.readthedocs.io/en/latest/. The code used to perform motif analysis is available at https://github.com/YichaoOU/HPFH_code and at https://doi.org/10.5281/zenodo.4784805.

Extended Data

a

b

c
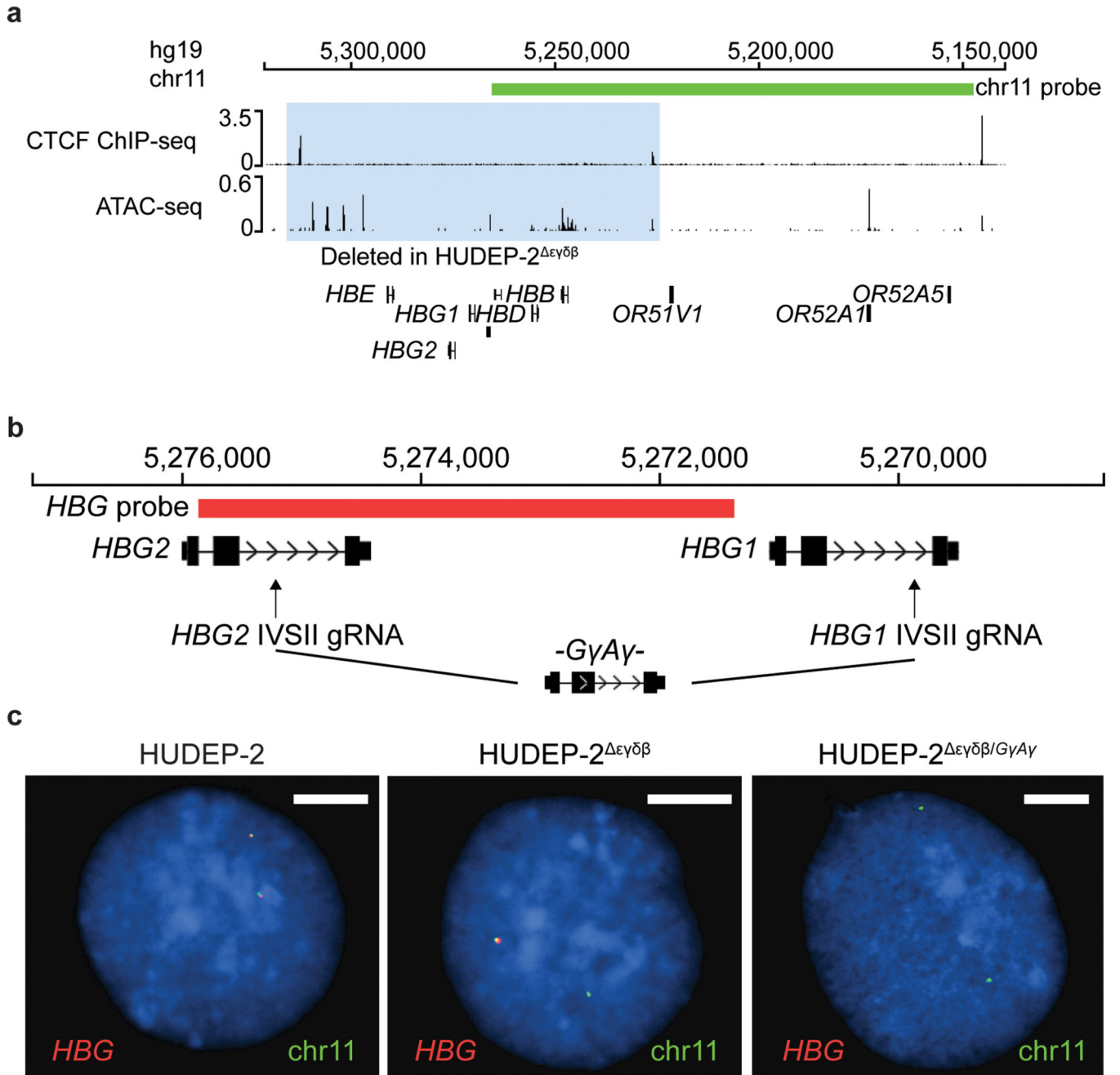
Extended Data Fig. 1. Derivation of HUDEP-2 cells containing a single γ-globin gene.
Genome browser view of deletions introduced into HUDEP-2 cells to generate the
HUDEP-2 $^{εγδβ/GγAγ}$ line, which contains a single γ-globin gene. The region of the
β-like globin gene cluster that was deleted on one chromosome is shown in blue.
Chromatin immunoprecipitation (ChIP-seq) analysis for CTCF occupancy and ATAC-seq
analysis of HUDEP-2 cells were derived from publicly available data. **b**, Generation of
a single *HBG2-HBG1* fusion gene on the remaining β-like globin gene locus. Positions

of the gRNAs targeting intron 2 of *HBG1* and *HBG2* are in shown with arrows. **c,** Fluorescence *in situ* hybridization analysis of wild-type (WT) HUDEP-2, HUDEP-2 $^{\epsilon\gamma\delta\beta}$, and HUDEP-2 $^{\epsilon\gamma\delta\beta/G\gamma A\gamma}$ score correlation comparing the frequency of the union (top) and intersection (bottom) between HUDEP-2 and HUDEP-2 $^{\epsilon\gamma\delta\beta/G\gamma A\gamma}$ cells. The TAD scores identify the degree of separation between the left and right boundaries based on the Hi-C interaction matrix. A TAD will be called at local minima. The union indicates the left and right boundaries across HUDEP-2 and HUDEP-2 $^{\epsilon\gamma\delta\beta/G\gamma A\gamma}$ cells. The intersection represents the fraction of shared boundaries between the HUDEP-2 and HUDEP-2 $^{\epsilon\gamma\delta\beta/G\gamma A\gamma}$ TAD sets. The Spearman correlation coefficients (ρ) are shown. Results were generated from merged reads derived from two independent experiments.

Extended Data Fig. 2. Characterization of the HUDEP-2 $^{εγδβ/G γA γ}$ cell line

a, Next generation sequencing (NGS) analysis of the indicated HUDEP-2 lines showing percentages of reads corresponding to *HBG1* or *HBG2* exon 3. (mean ± SD; $n = 6$ independent clones for each genotype). **b**, %HbF in WT and HUDEP-2 $^{εγδβ/G γA γ}$ clones, measured by ion-exchange high-performance liquid chromatography (IE-HPLC) after 7 days of erythroid differentiation. Box and whisker plots show minimum, maximum, median, and interquartile ranges. $n = 6$ independent clones for each genotype. $*p = 0.0156$ uncorrected two-tailed unpaired t-test. **c**, Kinetics of erythroid maturation of WT and

HUDEP-2 $^{\varepsilon\gamma\delta\beta/G\gamma A\gamma}$ cells determined by flow cytometry for CD49d and Band3 in the CD235a$^+$ population at the indicated timepoints after culture in erythroid differentiation medium. Mean ± SD is shown in each quadrant. $n = 6$ independent clones analyzed for each genotype.
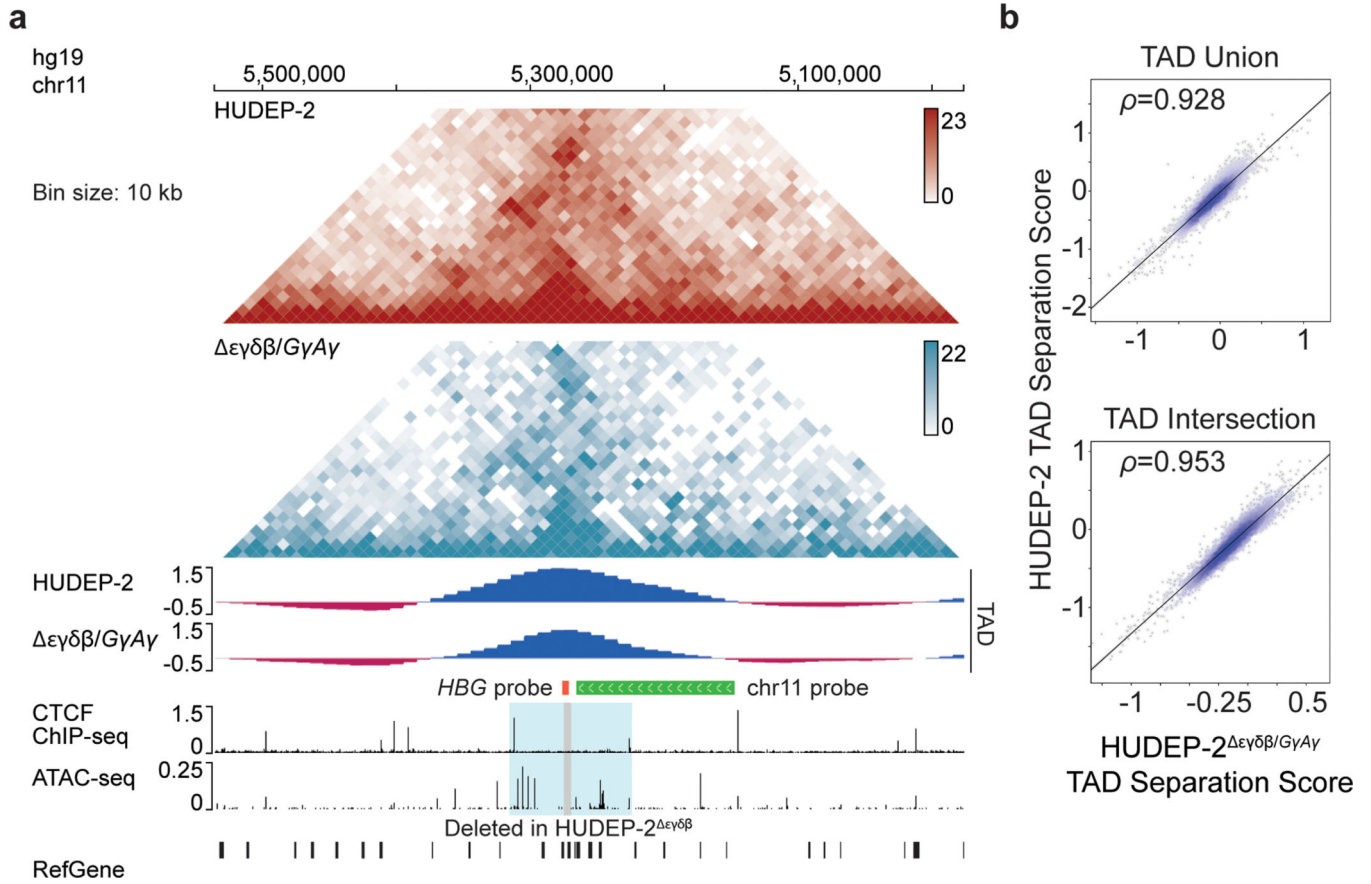
Extended Data Fig. 3. Hi-C analysis showing chromatin structure of the extended β-globin locus in HUDEP-2 cells containing a single γ-globin gene

a, Heat maps comparing chromatin interactions of the extended β-like globin locus in WT HUDEP-2 cells (red) and in HUDEP-2 $^{\varepsilon\gamma\delta\beta/G\gamma A\gamma}$ cells, which contain a single, modified β-like globin locus (blue). Tracks below show transcriptionally open or closed compartments as positive (blue) or negative (magenta) according to Hi-C analysis. CTCF ChIP-seq analysis and ATAC seq analysis are shown for WT HUDEP-2 cells. The 91.5 kb deletion of the extended β-globin locus ($\varepsilon\gamma\delta\beta$) is shown as a blue rectangle; the 5.4 kb deletion generating a single *HBG2-HBG1* fusion gene ($G\gamma A\gamma$) is designated in grey. Genes are designated as black vertical bars in the bottom track. **b**, The topologically associated domain (TAD) separation score correlation comparing the frequency of the union (top) and intersection (bottom) between HUDEP-2 and HUDEP-2 $^{\varepsilon\gamma\delta\beta/G\gamma A\gamma}$ cells. The TAD scores identify the degree of separation between the left and right boundaries based on the Hi-C interaction matrix. A TAD will be called at local minima. The union indicates the left and right boundaries across HUDEP-2 and HUDEP-2 $^{\varepsilon\gamma\delta\beta/G\gamma A\gamma}$ cells. The intersection represents the fraction of shared boundaries between the HUDEP-2 and

HUDEP-2 $^{\varepsilon\gamma\delta\beta/G\gamma A\gamma}$ TAD sets. The Spearman correlation coefficients (ρ) are shown. Results were generated from merged reads derived from two independent experiments.

a

b

Extended Data Fig. 4. Effects of HPFH variants on HbF expression in HUDEP-2 cells.
a, HPLC tracings showing hemoglobin analysis of WT HUDEP-2, HUDEP-2 $^{\varepsilon\gamma\delta\beta/G\gamma A\gamma}$ and HUDEP-2 $^{\varepsilon\gamma\delta\beta/G\gamma A\gamma}$ cells with the −110 A>C HPFH variant after 7 days of erythroid differentiation. **b**, ATAC-seq tracks showing open chromatin at the β-like globin gene cluster in single clones with distal CCAAT box HPFH variants −117 G>A and −114 C>A

and control mutations −110 A>G and −110 A>T. The shaded area highlighting the *HBG* promoter is shown in higher resolution on the right. The reference genes are shown at the bottom and the dotted lines indicate the region deleted to create the single in-frame *HBG* fusion gene.

a

b
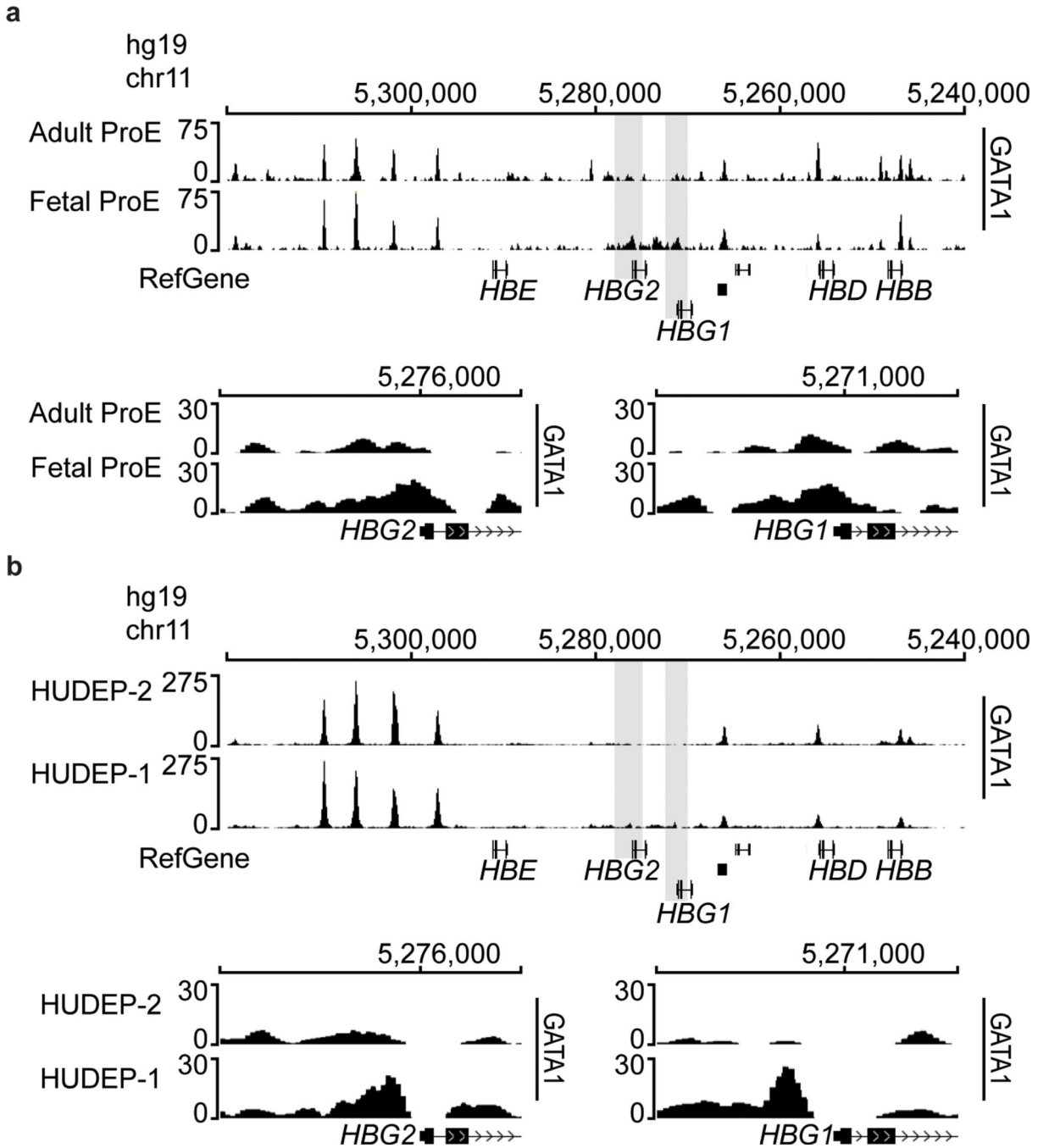
Extended Data Fig. 5. GATA1 occupancy at the γ-globin promoters is associated with HbF expression.

a, ChIP-seq analysis showing GATA1 occupancy at the β-like globin gene cluster in primary fetal and adult proerythroblasts[81]. The shaded areas highlighting the *HBG1 and*

HBG2 promoters are shown in higher resolution below. **b**, ChIP seq analysis for GATA1 in HUDEP-1 and HUDEP-2 cells, which express predominantly γ-globin and β-globin respectively, shown as described for panel a. GATA1 occupancy in HUDEP-2 cells was derived from publicly available data[82].
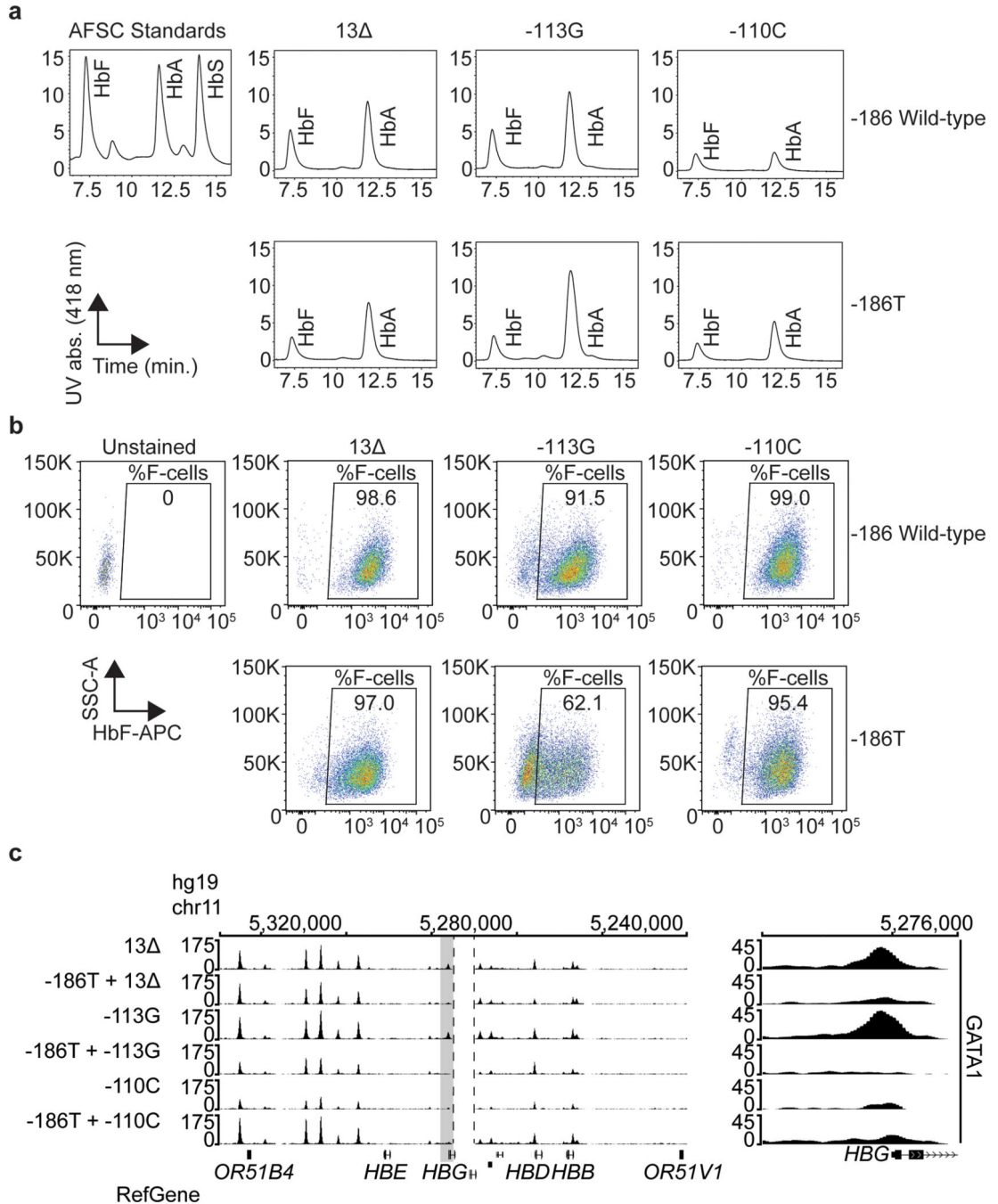
Extended Data Fig. 6. Disruption of the bipartite GATA motif via −186 C>T impairs HbF expression associated with HPFH variants

a, Representative ion-exchange high-performance liquid chromatography (HPLC) traces showing reduced fetal hemoglobin (HbF) peak intensity in HPFH clones without (top)

and with −186 C>T (bottom). Cells were grown in culture for 7 days under erythroid differentiation conditions. **b**, Representative F-cell staining flow cytometry plots in undifferentiated HPFH clones without (top) and with −186 C>T (bottom). **c**, Replicate ChIP-seq analysis showing GATA1 occupancy at the β-like globin gene cluster in clones harboring distal CCAAT box HPFH variants ± the −186 C>T GATA motif mutation (related to Fig. 2d). The shaded area highlighting the *HBG* promoter is shown in higher resolution on the right. The reference genes are shown at the bottom and the dotted lines indicate the region deleted to create the single in-frame *HBG* fusion gene.

a

b
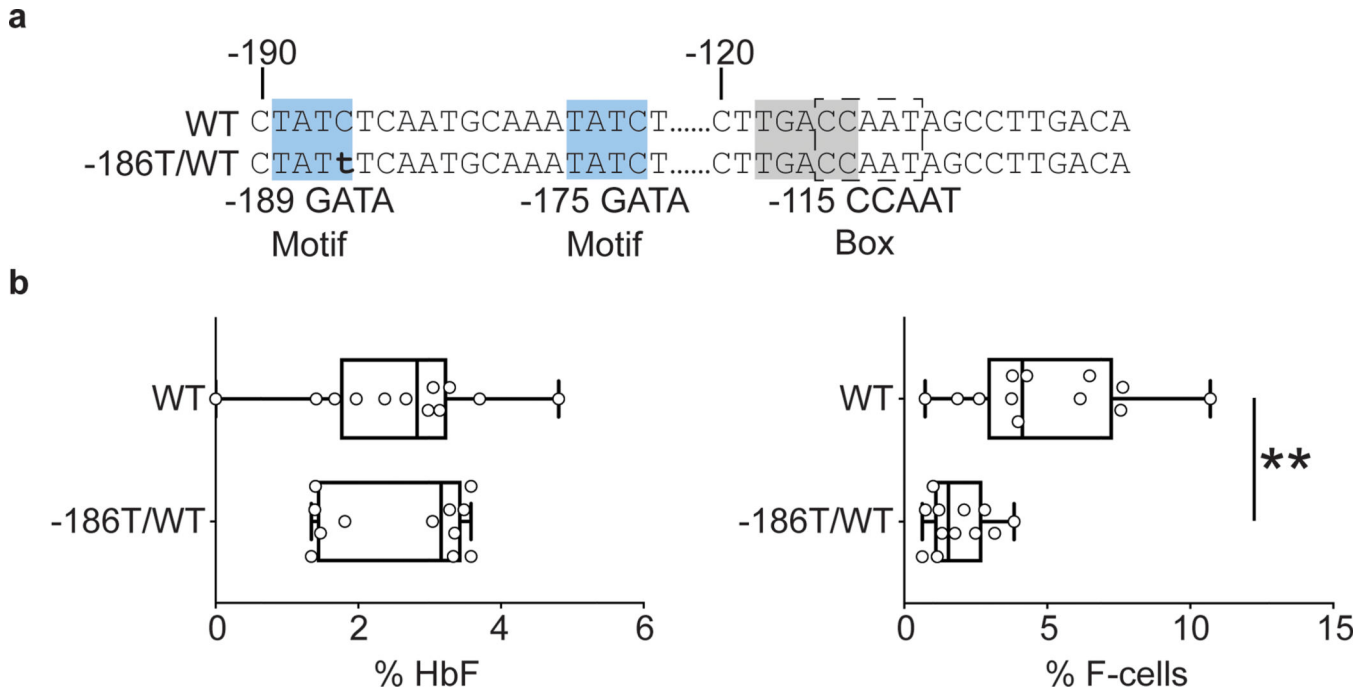
Extended Data Fig. 7. Disruption of the - 189 GATA motif in HUDEP-2 $^{εγδβ/GγAγ}$ **cells does not induce γ-globin expression**

a, Sequence of the *HBG* promoter showing the bipartite GATA motif (blue), BCL11A binding motif (grey) and the distal CCAAT box (dotted rectangle; hg19 – chr11:5,276,112–5,276,201). The −186 C>T mutation (lower case bold), disrupts GATA1 binding. **b**, %HbF (left) and %F-cells (right) after 7 days of erythroid differentiation in HUDEP-2 $^{εγδβ/GγAγ}$ cells ± the −186 C>T mutation (lower case bold). Each dot represents an individual clone ($n = 12$ per genotype). Box and whisker plots show minimum, maximum, median, and interquartile ranges. **$p = 0.0017$, uncorrected two-tailed unpaired t-test. An uncorrected two-tailed unpaired t-test indicated no significant effect of the −186T mutation on %HbF in WT CCAAT box clones, $p > 0.9999$.
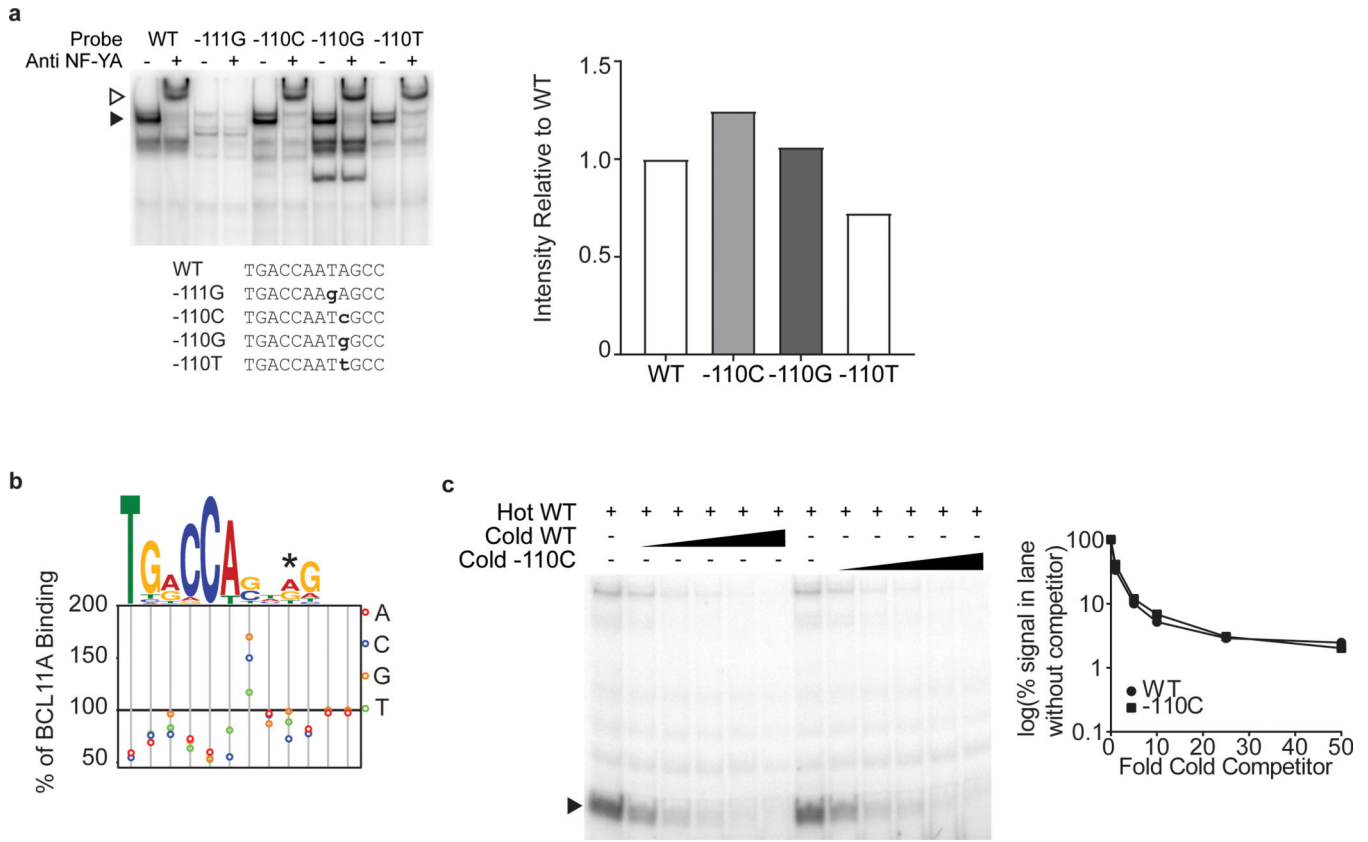
Extended Data Fig. 8. −110 A>C at the distal CCAAT box enhances NF-Y binding

a, Electrophoretic mobility shift assay (EMSA) for NF-Y binding to WT or mutant γ-globin promoter distal CCAAT box oligonucleotides in K562 cell nuclear extracts. Mutations are indicated in lower case bold. Bound probe is indicated by the closed triangle and supershift product of the NF-Y:probe complex is indicated by the open triangle. Graph on the right shows densitometry analysis of NF-Y band intensity relative to WT signal. **b**, Motif analysis showing the predicted effects of single nucleotide alterations on BCL11A binding to the −115 distal CCAAT box. The −110 A>C HPFH variant (asterisk) is predicted to decrease BCL11A affinity for the motif. **c**, Competitive EMSA assay for BCL11A binding to distal CCAAT box probes. The autoradiogram shows competition of cold WT or −110 A>C probes (1X, 5X, 10X, 25X, and 50X molar excess) with radiolabeled WT probe for binding to BCL11A zinc fingers 4–6 expressed in COS-7 cells. Bound probe is indicated by a closed triangle. The graph shows densitometry analysis of this band after incubation with cold competitor oligonucleotides, normalized to intensity with no competitor.

a

b
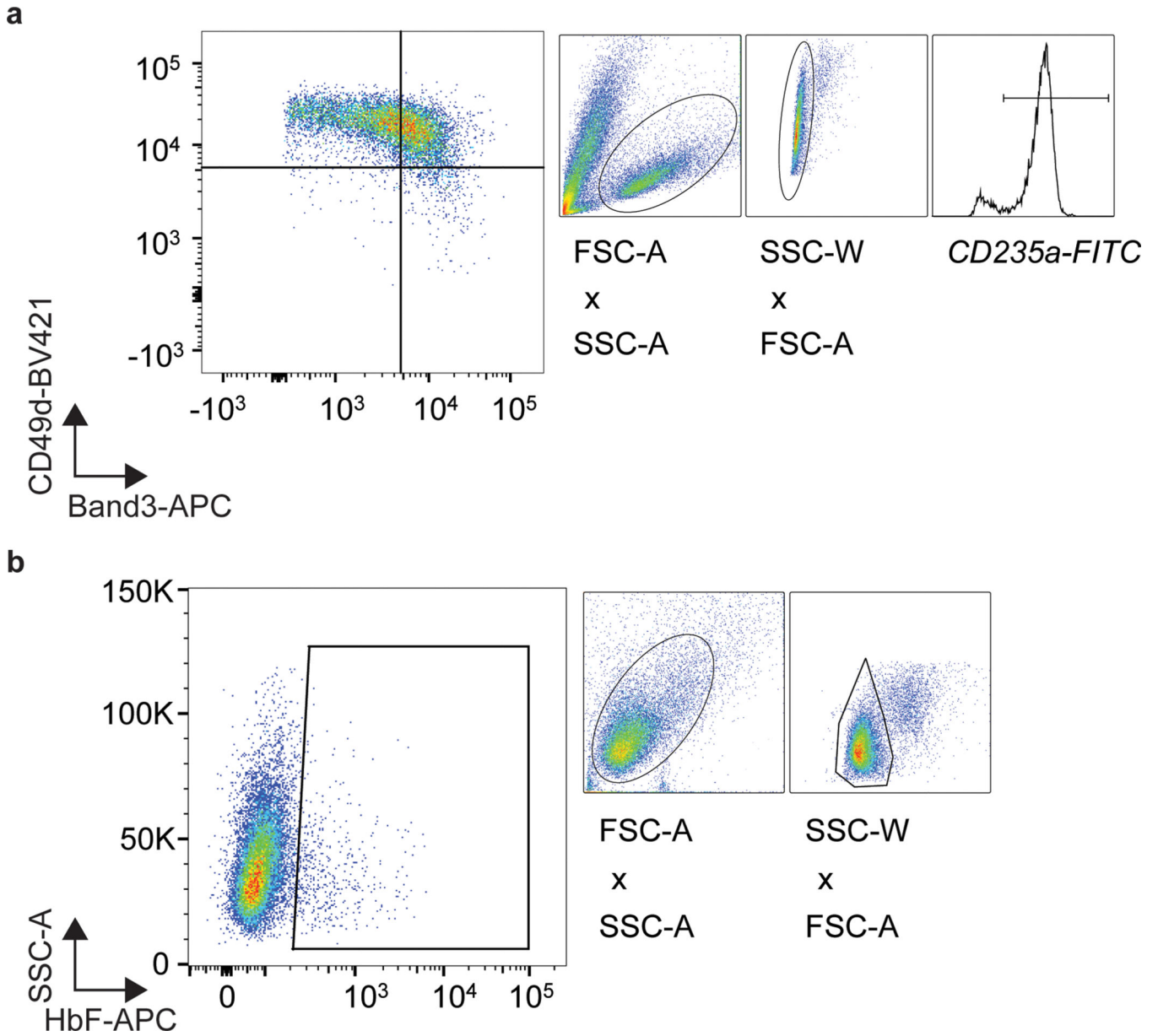
Extended Data Fig. 9. Flow cytometry gating strategies

a, Gating strategy for monitoring the differentiation of HUDEP-2 cells (see Extended Data Fig. 2c). **b**, Gating strategy for F-cell determination in undifferentiated HUDEP-2 cells (see Extended Data Fig. 6b). Antibodies used are listed in the methods section.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Dean A. On a chromosome far, far away: LCRs and gene expression. Trends Genet. 22, 38–45 (2006). [PubMed: 16309780]

2. Palstra R, de Laat W. & Grosveld F. Beta-globin regulation and long-range interactions. Adv. Genet. 61, 107–42 (2008). [PubMed: 18282504]

3. Platt OS et al. Mortality in sickle cell disease. Life expectancy and risk factors for early death. N. Engl. J. Med 330, 1639–44 (1994). [PubMed: 7993409]

4. Musallam KM et al. Fetal hemoglobin levels and morbidity in untransfused patients with β-thalassemia intermedia. Blood 119, 364–7 (2012). [PubMed: 22096240]

5. Lettre G. et al. DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease. Proc. Natl. Acad. Sci 105, 11869–11874 (2008). [PubMed: 18667698]

6. Uda M. et al. Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc. Natl. Acad. Sci 105, 1620–1625 (2008). [PubMed: 18245381]

7. Wonkam A. et al. Association of variants at BCL11A and HBS1L-MYB with hemoglobin F and hospitalization rates among sickle cell patients in Cameroon. PLoS One 9, e92506 (2014). [PubMed: 24667352]

8. Lettre G. & Bauer DE Fetal haemoglobin in sickle-cell disease: From genetic epidemiology to new therapeutic strategies. Lancet 387, 2554–64 (2016). [PubMed: 27353686]

9. Natta CL, Niazi GA, Ford S. & Bank A. Balanced globin chain synthesis in hereditary persistence of fetal hemoglobin. J. Clin. Invest 54, 433–8 (1974). [PubMed: 4847253]

10. Stamatoyannopoulos G, Wood WG, Papayannopoulou T. & Nute PE A new form of hereditary persistence of fetal hemoglobin in blacks and its association with sickle cell trait. Blood 46, 683–92 (1975). [PubMed: 1174703]

11. Martyn GE et al. Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. Nat. Genet 50, 498–503 (2018). [PubMed: 29610478]

12. Liu N. et al. Direct promoter repression by BCL11A controls the fetal to adult hemoglobin switch. Cell 173, 430–442.e17 (2018). [PubMed: 29606353]

13. Fessas P. & Stamatoyannopoulos G. Hereditary persistence of fetal hemoglobin in Greece: A study and a comparison. Blood 24, 223–40 (1964). [PubMed: 14214133]

14. Collins FS et al. A point mutation in the A gamma-globin gene promoter in Greek hereditary persistence of fetal haemoglobin. Nature 313, 325–6 (1985). [PubMed: 2578620]

15. Oner R, Kutlar F, Gu LH & Huisman TH The Georgia type of nondeletional hereditary persistence of fetal hemoglobin has a C---T mutation at nucleotide-114 of the A gamma-globin gene. Blood 77, 1124–5 (1991). [PubMed: 1704803]

16. Fucharoen S, Shimizu K. & Fukumaki Y. A novel C-T transition within the distal CCAAT motif of the G gamma-globin gene in the Japanese HPFH: Implication of factor binding in elevated fetal globin expression. Nucleic Acids Res. 18, 5245–53 (1990). [PubMed: 1698280]

17. Motum PI, Deng ZM, Huong L. & Trent RJ The Australian type of nondeletional G gamma-HPFH has a C-->G substitution at nucleotide −114 of the G gamma gene. Br. J. Haematol 86, 219–21 (1994). [PubMed: 7516698]

18. Zertal-Zidani S. et al. A novel C-->A transversion within the distal CCAAT motif of the Ggamma-globin gene in the Algerian Ggammabeta+-hereditary persistence of fetal hemoglobin. Hemoglobin 23, 159–69 (1999). [PubMed: 10335983]

19. Gilman JG et al. Distal CCAAT box deletion in the A gamma globin gene of two black adolescents with elevated fetal A gamma globin. Nucleic Acids Res. 16, 10635–42 (1988). [PubMed: 2462713]

20. Indrak K. et al. Compound heterozygosity for a beta zero-thalassemia (frameshift codons 38/39; -C) and a nondeletional Swiss type of HPFH (A----C at NT −110, G gamma) in a Czechoslovakian family. Ann. Hematol 63, 111–5 (1991). [PubMed: 1716997]

21. Amato A. et al. Interpreting elevated fetal hemoglobin in pathology and health at the basic laboratory level: New and known γ-gene mutations associated with hereditary persistence of fetal hemoglobin. Int. J. Lab. Hematol 36, 13–19 (2014). [PubMed: 23621512]

22. Wienert B. et al. KLF1 drives the expression of fetal hemoglobin in British HPFH. Blood 130, 803–807 (2017). [PubMed: 28659276]

23. Wienert B. et al. Editing the genome to introduce a beneficial naturally occurring mutation associated with increased fetal globin. Nat. Commun 6, 7085 (2015). [PubMed: 25971621]

24. Martyn GE et al. A natural regulatory mutation in the proximal promoter elevates fetal globin expression by creating a de novo GATA1 site. Blood 133, 852–856 (2019). [PubMed: 30617196]

25. Liberati C, Ronchi A, Lievens P, Ottolenghi S. & Mantovani R. NF-Y organizes the gamma-globin CCAAT boxes region. J. Biol. Chem 273, 16880–16889 (1998). [PubMed: 9642249]

26. Duan Z, Stamatoyannopoulos G. & Li Q. Role of NF-Y in in vivo regulation of the gamma-globin gene. Mol. Cell. Biol 21, 3083–95 (2001). [PubMed: 11287613]

27. Martyn GE, Quinlan KGR & Crossley M. The regulation of human globin promoters by CCAAT box elements and the recruitment of NF-Y. Biochim. Biophys. Acta 1860, 525–536 (2017).

28. Martin DI, Tsai SF & Orkin SH Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor. Nature 338, 435–8 (1989). [PubMed: 2467208]

29. Vernimmen D. & Bickmore WA The hierarchy of transcriptional activation: from enhancer to promoter. Trends Genet. 31, 696–708 (2015). [PubMed: 26599498]

30. Xu J. et al. Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. Proc. Natl. Acad. Sci. U. S. A 110, 6518–23 (2013). [PubMed: 23576758]

31. Sankaran VG et al. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science (80-.). 322, 1839–42 (2008).

32. Liu N. et al. Transcription factor competition at the γ-globin promoters controls hemoglobin switching. Nat. Genet (2021) doi:10.1038/s41588-021-00798-y.

33. Traxler EA et al. A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. Nat. Med 22, 987–990 (2016). [PubMed: 27525524]

34. Masuda T. et al. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. Science (80-.). 351, 285–9 (2016).

35. Métais J-Y et al. Genome editing of HBG1 and HBG2 to induce fetal hemoglobin. Blood Adv. 3, 3379–3392 (2019). [PubMed: 31698466]

36. Weber L. et al. Editing a γ-globin repressor binding site restores fetal hemoglobin synthesis and corrects the sickle cell disease phenotype. Sci. Adv 6, eaay9392 (2020). [PubMed: 32917636]

37. Kurita R. et al. Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. PLoS One 8, e59890 (2013). [PubMed: 23533656]

38. Huisman TH et al. The occurrence of different levels of G gamma chain and of the A gamma T variant of fetal hemoglobin in newborn babies from several countries. Am. J. Hematol 14, 133–48 (1983). [PubMed: 6188375]

39. Sukumaran PK et al. Gamma thalassemia resulting from the deletion of a gamma-globin gene. Nucleic Acids Res. 11, 4635–43 (1983). [PubMed: 6306596]

40. Zeng YT, Huang SZ, Nakatsuji T. & Huisman TH -G gamma A gamma-thalassemia and gamma-chain variants in Chinese newborn babies. Am. J. Hematol 18, 235–42 (1985). [PubMed: 2579547]

41. Buenrostro JD, Giresi PG, Zaba LC, Chang HY & Greenleaf WJ Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–8 (2013). [PubMed: 24097267]

42. Xiang G. et al. S3norm: simultaneous normalization of sequencing depth and signal-to-noise ratio in epigenomic data. Nucleic Acids Res. 48, e43 (2020). [PubMed: 32086521]

43. Gaudelli NM et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 551, 464–471 (2017). [PubMed: 29160308]

44. Koblan LW et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat. Biotechnol 36, 843–846 (2018). [PubMed: 29813047]

45. Zhu X. et al. NF-Y recruits both transcription activator and repressor to modulate tissue- and developmental stage-specific expression of human γ-globin gene. PLoS One 7, e47175 (2012). [PubMed: 23071749]

46. Superti-Furga G, Barberis A, Schaffner G. & Busslinger M. The −117 mutation in Greek HPFH affects the binding of three nuclear factors to the CCAAT region of the gamma-globin gene. EMBO J. 7, 3099–107 (1988). [PubMed: 3181130]

47. Grant CE, Bailey TL & Noble WS FIMO: Scanning for occurrences of a given motif. Bioinformatics 27, 1017–8 (2011). [PubMed: 21330290]

48. Forget BG Molecular basis of hereditary persistence of fetal hemoglobin. Ann. N. Y. Acad. Sci 850, 38–44 (1998). [PubMed: 9668525]

49. Wienert B, Martyn GE, Funnell APW, Quinlan KGR & Crossley M. Wake-up sleepy gene: Reactivating fetal globin for β-hemoglobinopathies. Trends Genet. 34, 927–940 (2018). [PubMed: 30287096]

50. Fang X, Han H, Stamatoyannopoulos G. & Li Q. Developmentally specific role of the CCAAT box in regulation of human gamma-globin gene expression. J. Biol. Chem 279, 5444–5449 (2004). [PubMed: 14645237]

51. Ronchi A. et al. Role of the duplicated CCAAT box region in gamma-globin gene regulation and hereditary persistence of fetal haemoglobin. EMBO J. 15, 143–9 (1996). [PubMed: 8598197]

52. Katsube T. & Fukumaki Y. A role for the distal CCAAT box of the gamma-globin gene in Hb switching. J. Biochem 117, 68–76 (1995). [PubMed: 7539789]

53. Huang D-Y et al. GATA-1 and NF-Y cooperate to mediate erythroid-specific transcription of Gfi-1B gene. Nucleic Acids Res. 32, 3935–46 (2004). [PubMed: 15280509]

54. Torchy MP, Hamiche A. & Klaholz BP Structure and function insights into the NuRD chromatin remodeling complex. Cell. Mol. Life Sci 72, 2491–507 (2015). [PubMed: 25796366]

55. Yang Y. et al. Structural insights into the recognition of γ-globin gene promoter by BCL11A. Cell Res. 29, 960–963 (2019). [PubMed: 31467406]

56. Jolma A. et al. DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature 527, 384–8 (2015). [PubMed: 26550823]

57. Pique-Regi R. et al. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. 21, 447–55 (2011). [PubMed: 21106904]

58. Vökel S. et al. Zinc finger independent genome-wide binding of SP2 potentiates recruitment of histone-fold protein NF-Y distinguishing it from SP1 and SP3. PLoS Genet. 11, 1–25 (2015).

59. Suske G. NF-Y and SP transcription factors — New insights in a long-standing liaison. Biochim. Biophys. Acta 1860, 590–597 (2017).

60. Yang L. et al. Methylation of a CGATA element inhibits binding and regulation by GATA-1. Nat. Commun 11, 2560 (2020). [PubMed: 32444652]

61. Vakoc CR et al. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. Mol. Cell 17, 453–462 (2005). [PubMed: 15694345]

62. Downes DJ et al. High-resolution targeted 3C interrogation of cis-regulatory element organization at genome-wide scale. Nat. Commun 12, 531 (2021). [PubMed: 33483495]

63. Norton LJ et al. KLF1 directly activates expression of the novel fetal globin repressor ZBTB7A/LRF in erythroid cells. Blood Adv. 1, 685–692 (2017). [PubMed: 29296711]

64. Zhou D, Liu K, Sun C-W, Pawlik KM & Townes TM KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching. Nat. Genet 42, 742–744 (2010). [PubMed: 20676097]

65. Humbert O. et al. Therapeutically relevant engraftment of a CRISPR-Cas9-edited HSC-enriched population with HbF reactivation in nonhuman primates. Sci. Transl. Med 11, eaaw3768 (2019). [PubMed: 31366580]

66. Lux CT et al. TALEN-mediated gene editing of HBG in human hematopoietic stem cells leads to therapeutic fetal hemoglobin induction. Mol. Ther. Methods Clin. Dev 12, 175–183 (2019). [PubMed: 30705922]

67. Anzalone AV, Koblan LW & Liu DR Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat. Biotechnol 38, 824–844 (2020). [PubMed: 32572269]

68. Connelly JP & Pruett-Miller SM CRIS.py: A versatile and high-throughput analysis program for CRISPR-based genome editing. Sci. Rep 9, 4194 (2019). [PubMed: 30862905]

69. Landt SG et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 22, 1813–31 (2012). [PubMed: 22955991]

70. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–2 (2010). [PubMed: 20110278]

71. Li H. & Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–60 (2009). [PubMed: 19451168]

72. Li H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–9 (2009). [PubMed: 19505943]

73. Ramírez F, Dündar F, Diehl S, Grüning BA & Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–91 (2014). [PubMed: 24799436]

74. Zhang Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008). [PubMed: 18798982]

75. Buenrostro JD, Giresi PG, Zaba LC, Chang HY & Greenleaf WJ Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10, 1213–1218 (2013). [PubMed: 24097267]

76. Jiang H, Lei R, Ding S-W & Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15, 182 (2014). [PubMed: 24925680]

77. Rao SSP et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680 (2014). [PubMed: 25497547]

78. Servant N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 16, 259 (2015). [PubMed: 26619908]

79. Ramírez F. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat. Commun 9, 189 (2018). [PubMed: 29335486]

80. Andrews NC & Faller DV A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic Acids Res. 19, 2499 (1991). [PubMed: 2041787]

81. Xu J. et al. Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. Dev. Cell 23, 796–811 (2012). [PubMed: 23041383]

82. Cheng L. et al. Single-nucleotide-level mapping of DNA regulatory elements that control fetal hemoglobin expression. Nat. Genet 1–58 (2021) doi:10.1038/s41588-021-00861-8. [PubMed: 33414547]
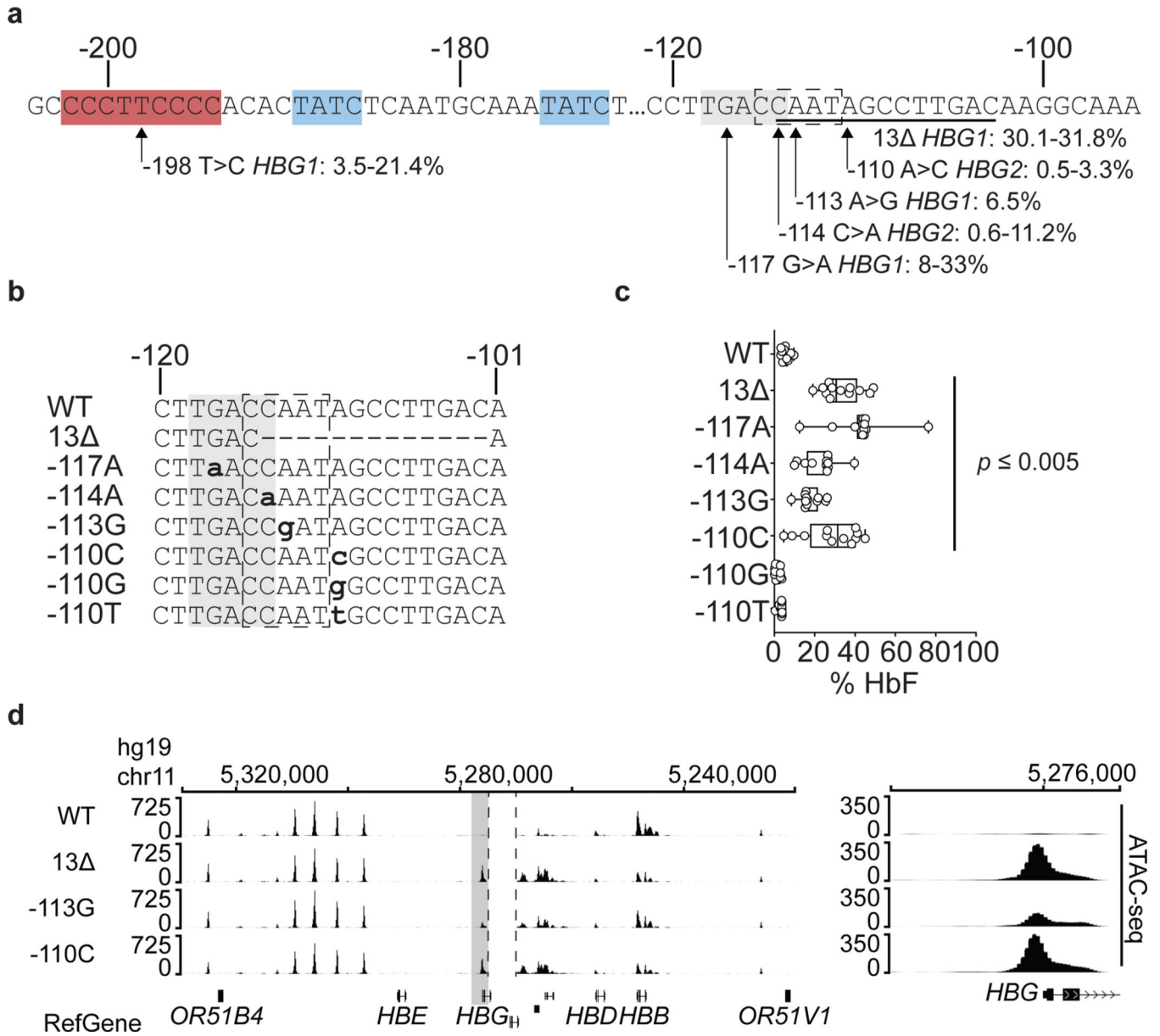
Figure 1. HPFH variants in the γ-globin promoter distal CCAAT box facilitate recruitment of GATA1 to an upstream motif.

a, The *HBG1/2* promoter sequence with transcriptional start at position +1 (hg19 – chr11:5,276,105–5,276,215). HPFH variants examined in this study and their associated %HbF range in RBCs of heterozygous individuals are shown[21,27,49]. Transcription factor binding motifs for ZBTB7A (red), BCL11A (grey) and GATA1 (blue), are shown. The −115 distal CCAAT box is indicated by a dashed rectangle. **b**, Distal CCAAT box HPFH variants were generated in HUDEP-2 $^{εγδβ/GγAγ}$ cells harboring a single wild-type γ-globin (*HBG*) gene. Nucleotide substitutions are shown in bold lower case; the 13Δ HPFH deletion is dashed. −198 T>C generates a *de novo* KLF1 binding motif; 13Δ, −117 G>A and −114 C>A disrupt the BCL11A binding motif; −113 A>G generates a *de novo* GATA1 binding motif; −110 A>C generates a *de novo* NF-Y motif (this study); −110 A>G and −110

A>T represent inert controls. **c**, Fetal hemoglobin (HbF) levels measured by ion-exchange high performance liquid chromatography (IE-HPLC) in WT and mutant clones grown after erythroid differentiation for 7 days. Each dot represents an individual clone ($n = 12$ per genotype). Box and whisker plots show minimum, maximum, median, and interquartile ranges of independent clones. Multiplicity adjusted p-values of each variant versus WT by ordinary one-way ANOVA with Dunnett's multiple comparisons test: 13   , −117A, −114A, −110C ($p < 0.0001$); −113G ($p = 0.005$). **d**, ATAC-seq analysis at the β-like globin gene cluster in WT HUDEP-2 $^{εγδβ/GγAγ}$ cells and selected mutant clones. Vertical dotted lines indicate the region deleted to generate an in-frame *HBG* fusion gene. The shaded area highlighting the single *HBG* promoter is shown in higher resolution on the right. Reference genes are shown at the bottom.
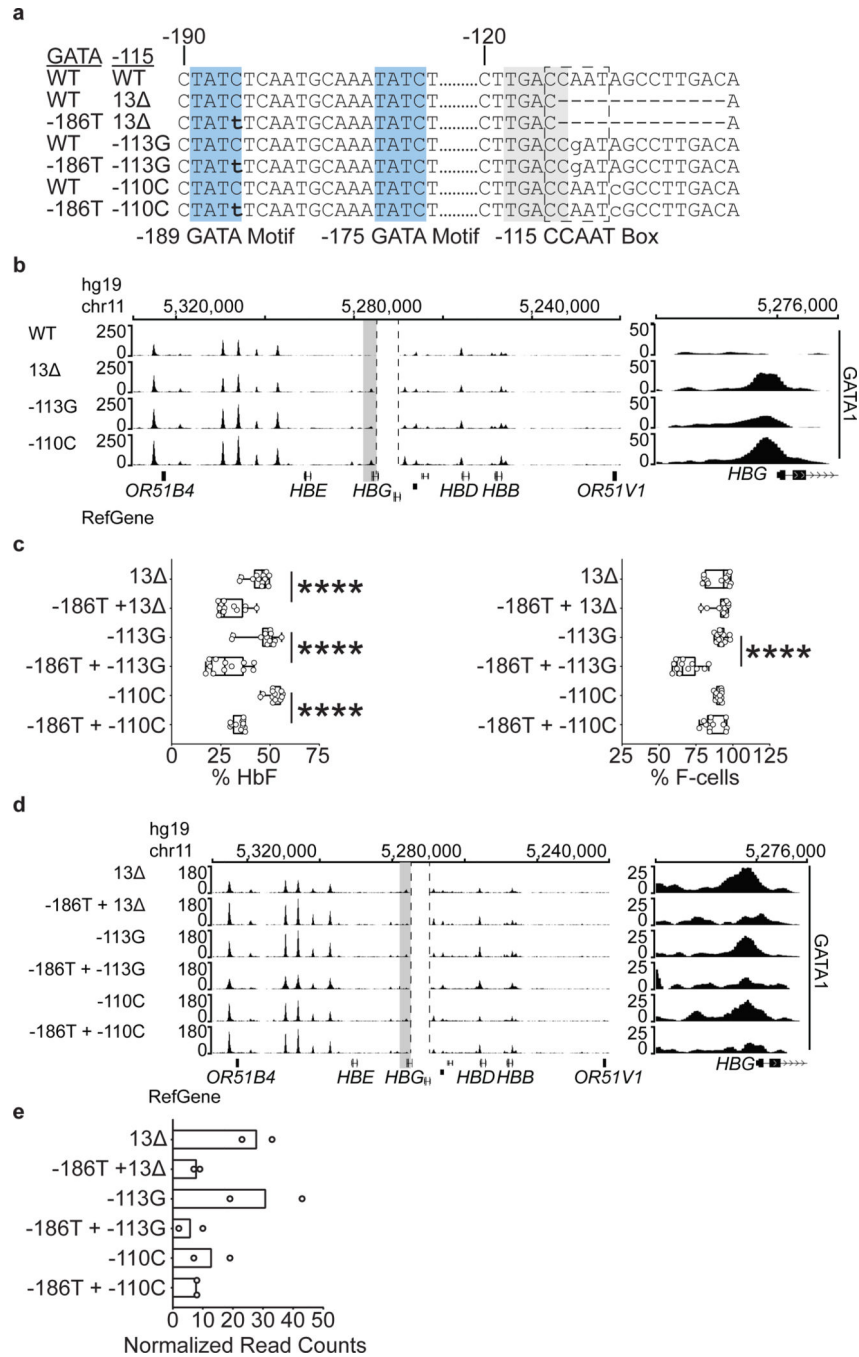
Figure 2. The −189 GATA motif facilitates γ-globin gene activation in HPFH.

a, The *HBG* promoters showing bipartite GATA motifs (blue), the −115 distal CCAAT box (dotted rectangle), and the BCL11A binding motif (grey; hg19 – chr11:5,276,112–5,276,201). **b,** GATA1 ChIP-seq analysis at the β-like globin gene cluster in WT HUDEP-2 $^{εγδβ/GγAγ}$ cells and selected mutant clones. **c,** Graph on the left shows %HbF in HUDEP-2 $^{εγδβ/GγAγ}$ clones harboring HPFH ± −186 C>T GATA motif mutations after 7 days of erythroid differentiation. Graph on the right shows, %HbF-immunostaining "F-cells", measured prior to differentiation. Each dot represents an individual clone (*n* = 12

per genotype). Box and whisker plots show minimum, maximum, median, and interquartile ranges of independent clones. ****$p < 0.0001$, uncorrected two-tailed unpaired t-test. **d,** ChIP-seq analysis of GATA1 occupancy at the β-like globin gene cluster in clones harboring distal CCAAT box HPFH variants ± −186 C>T GATA motif mutations. **e**, GATA1 ChIP-seq signals at the *HBG* promoter between clones harboring HPFH variants ± GATA motif −186 C>T mutations in two biological replicate experiments using S3norm to adjust for differences in sequencing depth and signal-to-noise ratios.
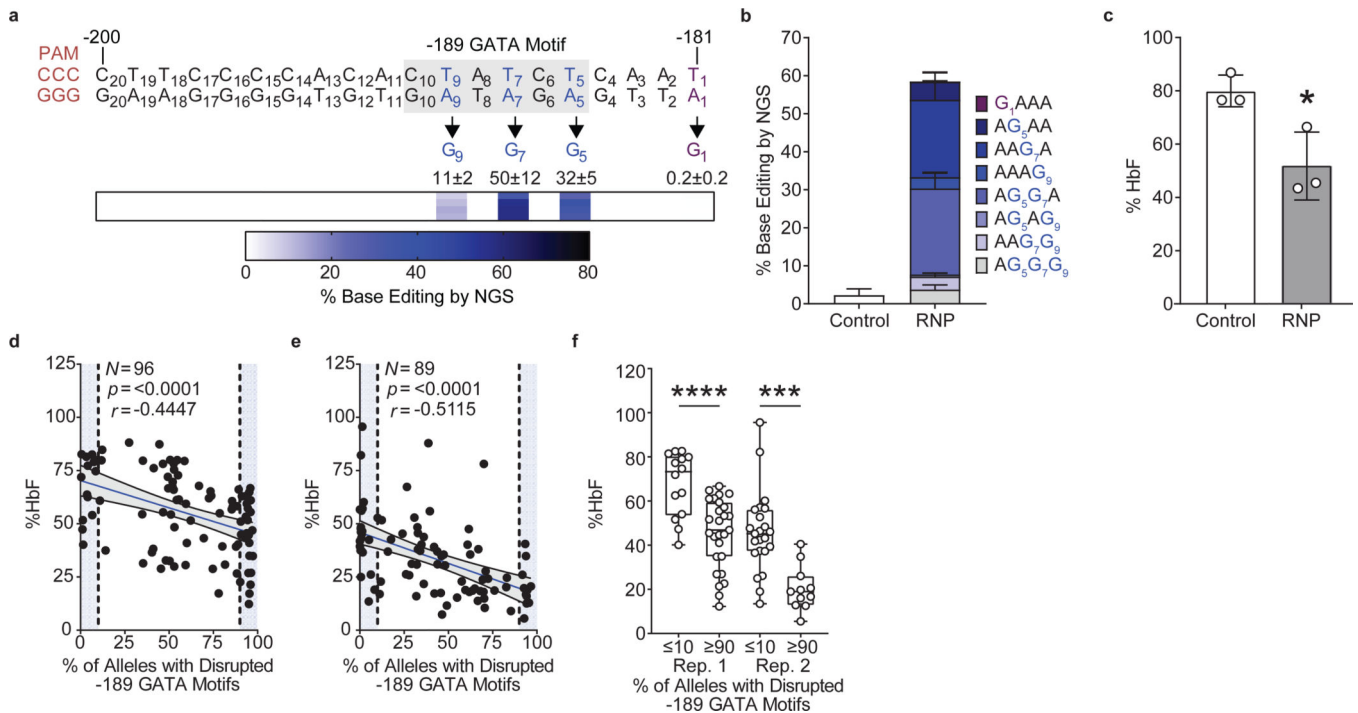
Figure 3. Disruption of the −189 GATA motif inhibits HbF expression in primary erythroblasts.
Normal donor umbilical cord blood (UCB) CD34+ cells were electroporated with ribonucleoprotein (RNP) containing the adenosine base editor ABE7.10 and gRNA targeting the −189 GATA motif. Edited cells were maintained in liquid culture or seeded into methylcellulose medium with erythroid cytokines after 48 hours. Control cells were not electroporated. **a**, Sequence of the gRNA target recognition sequence with the protospacer adjacent motif (PAM) in red. The −189 GATA motif is shaded grey (hg19 − chr11:5,276,192–5,276,214). Potential edits within or outside of the WGATAR motif are shown in blue or purple, respectively. Editing frequencies of individual adenosines, measured by next generation sequencing (NGS) after 96 hours, are shown for each position and color-coded according to the heat map (mean ± SD; $n = 4$ biological replicates across two independent experiments). **b**, Frequencies of mutant genotypes after base editing (mean ± SD; $n = 4$ biological replicates across two independent experiments). **c**, %HbF in bulk-edited or unedited control populations after 10 days of *in vitro* erythroid differentiation (mean ± SD; $n = 3$ biological replicates). *$p = 0.0255$, uncorrected two-tailed unpaired t-test **d,** %HbF in 14-day-old burst forming unit erythroid (BFU-E) colonies versus number of *HBG1* and *HBG2* alleles with mutations in the core −189 GATA motif (positions A_7 and/or A_9). Each dot represents a BFU-E colony from the same UCB cells analyzed in panel **c**. Linear regression analysis and two-tailed Pearson's correlation coefficient are shown. No adjustments for multiple corrections were made. **e**, BFU-E colony analysis performed as shown in panel **d** using UCB cells from a different donor. The mean is indicated by the blue line with the 95% confidence interval shaded between the black curves for **d** and **e**. **f**, HbF expression in BFU-E colonies with 10% or 90% editing of the −189 GATA motif, indicated by the shaded regions in panels **d** (Rep. 1) and **e** (Rep. 2). $n = (14)$ 10% edited Rep. 1 colonies, (27) 90% edited Rep. 1 colonies, (23) 10% edited Rep. 2 colonies, and

(11) 90% edited Rep. 2 colonies. ***$p = 0.0001$, ****$p < 0.0001$, uncorrected two-tailed unpaired t-test.
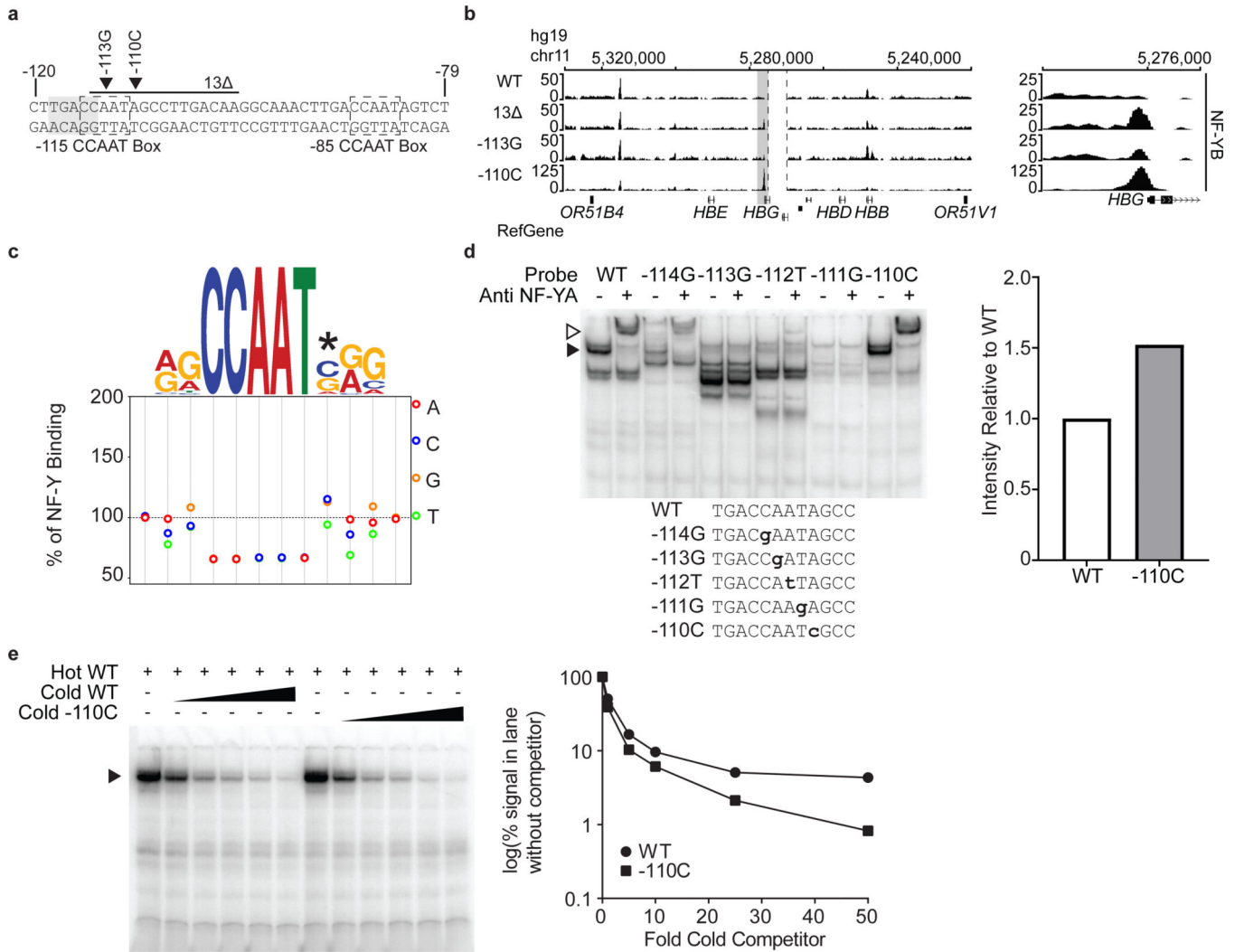
Figure 4. Distal CCAAT box HPFH variants recruit NF-Y to the γ-globin promoter.
a, Sequence of the *HBG* promoter showing the −85 proximal and −115 distal CCAAT boxes
(hg19 – chr11:5,276,090–5,276,131). HPFH nucleotide substitutions are indicated by filled
triangles. The 13Δ HPFH deletion is shown as a black line. **b**, ChIP-seq analysis showing
NF-YB occupancy in HUDEP-2 $^{εγδβ/GγAγ}$ clones harboring distal CCAAT box HPFH
variants. **c**, Motif analysis showing the predicted effects of single nucleotide alterations on
NF-Y binding to the −115 distal CCAAT box. The −110 A>C HPFH variant (asterisk) is
predicted to increase NF-Y affinity for the motif. **d,** Electrophoretic mobility shift assay
(EMSA) for NF-Y binding to WT or mutant oligonucleotides representing the γ-globin
promoter distal CCAAT box using K562 cell nuclear extracts. Mutations are indicated in
lower case bold. Bound probe is indicated by the closed triangle and supershift product
of the NF-Y:probe complex by the open triangle. Graph on the right shows densitometry
analysis of NF-Y band intensity for the −110 A>C probe relative to WT. **e**, Competitive
EMSA assay for NF-Y binding to distal CCAAT box probes. The autoradiogram shows
competition of cold WT or −110 A>C probes (1X, 5X, 10X, 25X, and 50X molar excess)
with radiolabeled WT probe for binding to NF-Y in K562 nuclear extracts. Bound probe

is indicated by a closed triangle. Graph on the right shows densitometry analysis of band intensities normalized to intensity of the band with no added competitor.
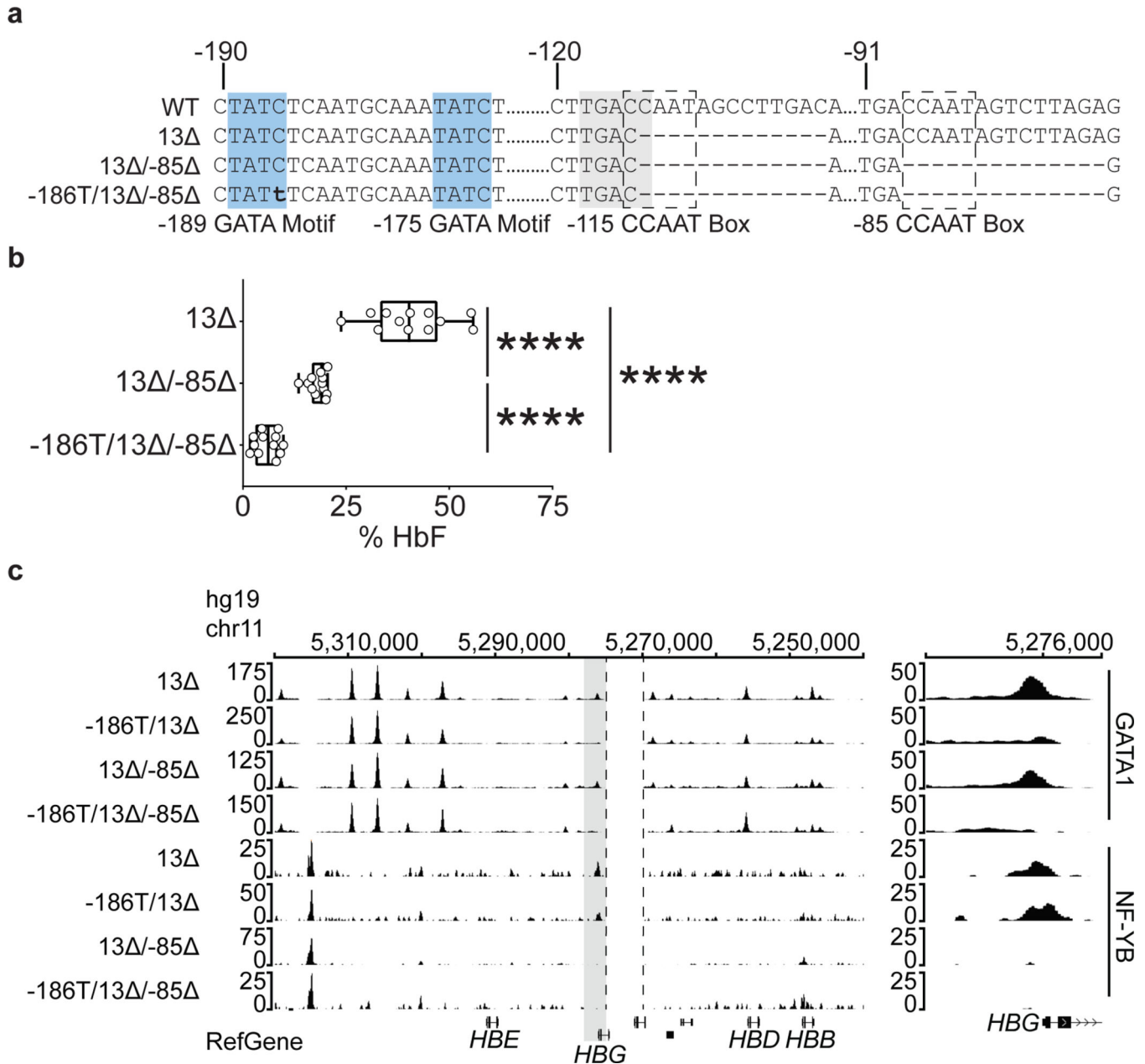
a

b

c

Figure 5. GATA1 and NF-Y cooperate to activate γ-globin gene expression.
a, The γ-globin promoter showing transcription factor binding motifs and mutations analyzed according to designations described for Figure 2a (hg19 – chr11:5,276,085–5,276,201). **b,** %HbF in clones with the indicated mutations, measured after 7 days of erythroid differentiation. Box and whisker plots show minimum, maximum, median, and interquartile ranges. Each dot represents an individual clone ($n = 12$ per genotype). Ordinary one-way ANOVA with Tukey's multiple comparisons test shows the multiplicity adjusted p-values between genotypes. ****$p < 0.0001$. **c**, ChIP-seq analysis showing GATA-1 and NFY occupancy at the β-like globin gene cluster in HUDEP-2 $^{εγδβ/GγAγ}$ clones with the indicated mutations.
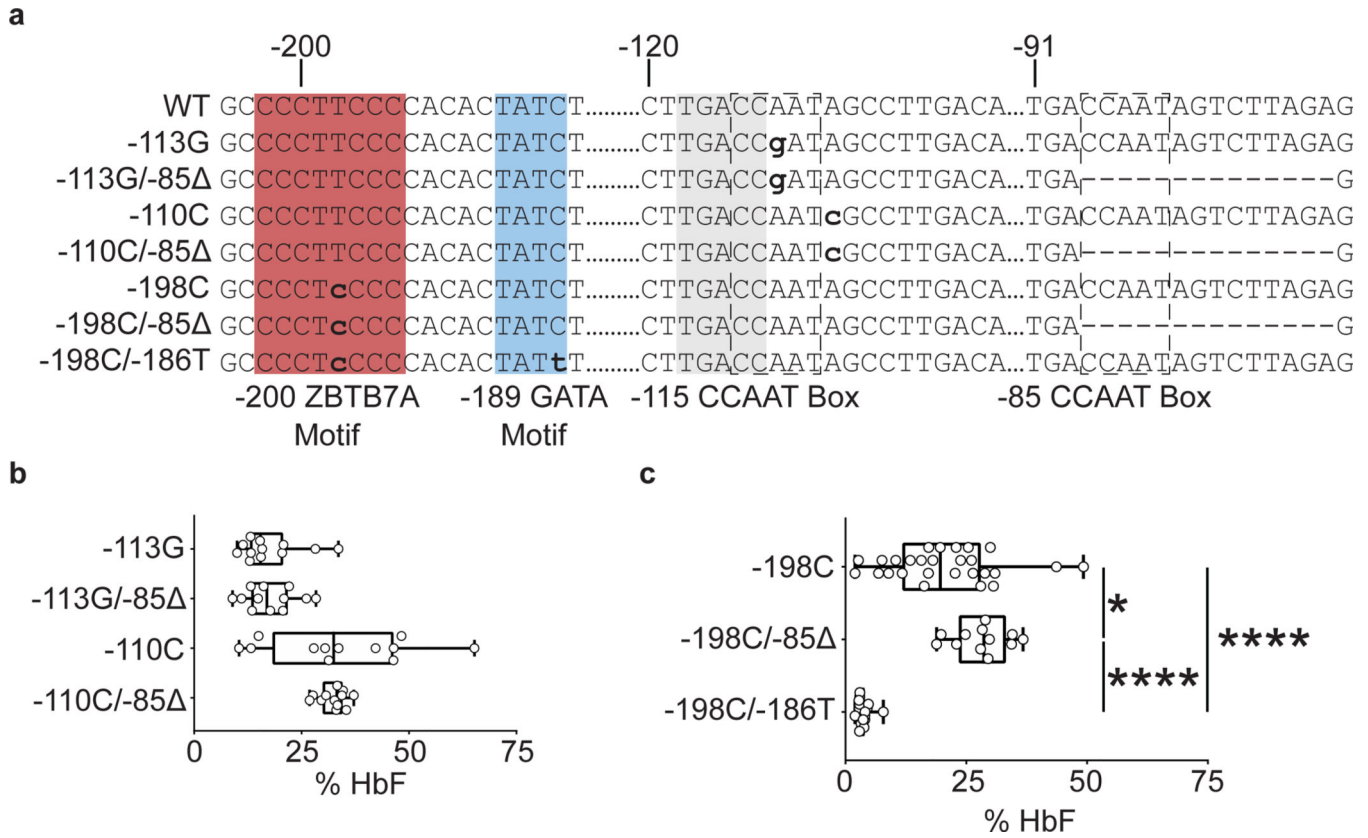
Figure 6. NF-Y binding to the −85 proximal CCAAT box is dispensable for HPFH mutations that create de novo transcription factor binding sites.

a, The γ-globin promoter showing transcription factor binding motifs and mutations analyzed according to designations described for Figure 1a (hg19 – chr11:5,276,085–5,276,215. HPFH variants that create de novo transcription factor binding sites include −198 T>C (KLF1)[22], −113 A>G (GATA1)[24], and −110 A>C (NF-Y, this report). **b,** %HbF in clones with the indicated mutations, measured after 7 days of erythroid differentiation. Box and whisker plots show minimum, maximum, median, and interquartile ranges. Each dot represents an individual clone ($n = 12$ per genotype). A two-tailed unpaired t-test indicated no significant effect of the 85 mutation on either HPFH variant. **c,** %HbF in clones with the indicated mutations, analyzed as described for panel **b**. $n = (27)$ −198C clones, (17) −198C + −186T clones, and (12) −198C + −85 clones). Ordinary one-way ANOVA with Tukey's multiple comparisons test shows the multiplicity adjusted p-values between genotypes. $*p = 0.0351$; $****p < 0.0001$.
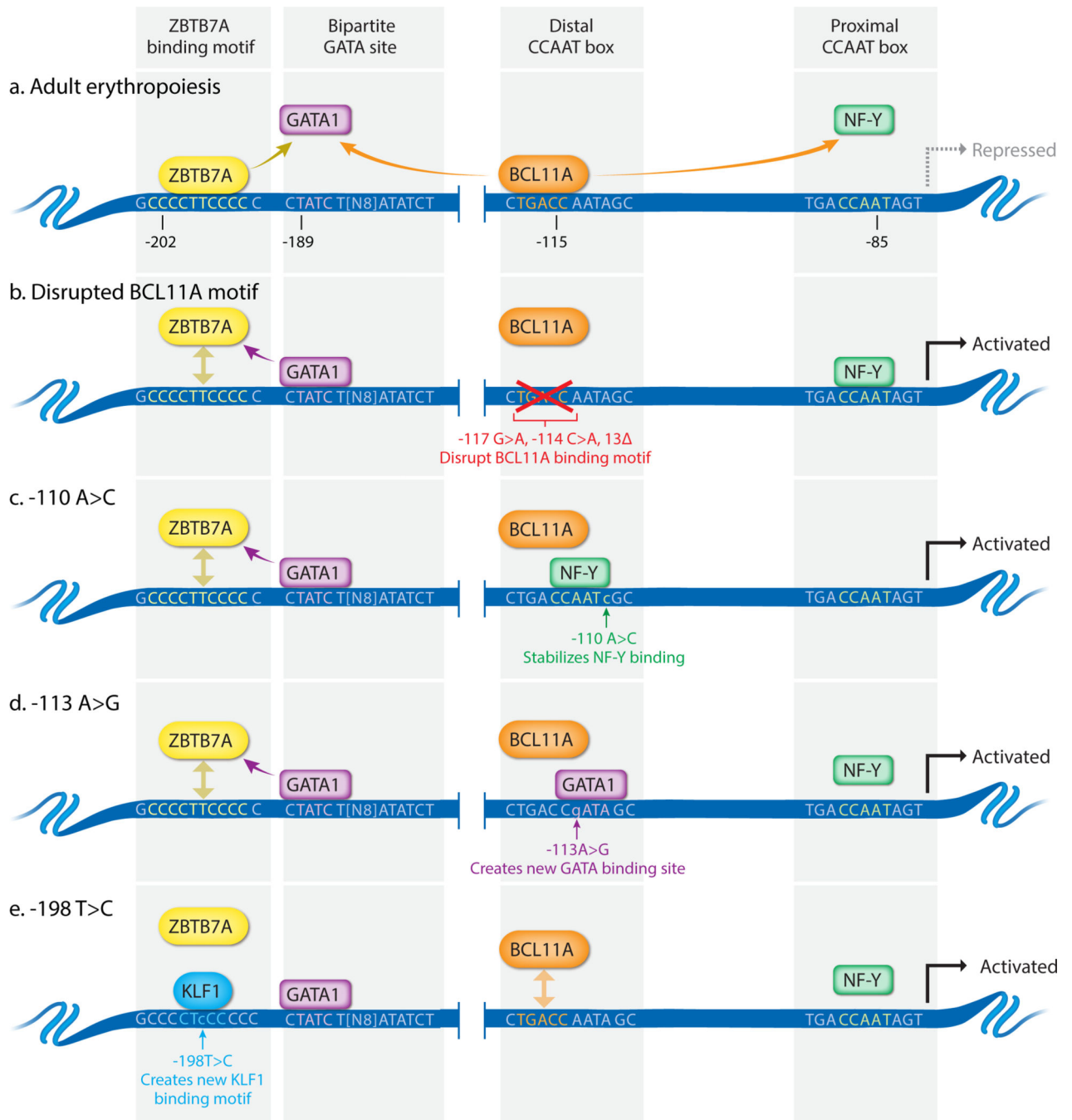
Figure 7. Competition between transcriptional repressors and activators for the γ-globin promoter in HPFH.

a, In adult-stage erythroid cells, ZBTB7A, BCL11A and their associated NuRD repressor complex (not shown) bind their indicated motifs and inhibit the recruitment of transcriptional activators GATA1 and NF-Y via steric effects and/or by establishing epigenetic modifications that inhibit chromatin occupancy (yellow and orange arrows). **b,** Numerous HPFH mutations disrupt the BCL11A binding motif, leading to GATA1 and NF-Y chromatin occupancy and transcriptional activation. Double arrows indicate that ZBTB7A

may still bind its motif to exert a partial inhibitory effect. **c,** The HPFH variant −110 A>C stabilizes ectopic binding of NF-Y to the distal CCAAT box, which activates transcription by displacing BCL11A and promoting GATA1 occupancy. **d,** The −113 A>G HPFH variant creates a new GATA1 binding site at the distal CCAAT box. GATA-1 displaces BCL11A to activate transcription, in part by facilitating GATA1 binding to the upstream −189 motif. **e,** The HPFH variant −198 T>C creates a new binding motif for KLF1, which displaces ZBTB7A and activates GATA1-dependent transcription. BCL11A may still bind its motif, resulting in partial gene silencing, as indicated by the double arrow. Binding of GATA1 to the −189 motif is required for normal fetal γ-globin expression and for all HPFH variants tested. In contrast, NF-Y binding to the proximal CCAAT box is dispensable for transcriptional activation by the −110 A>C, −113 A>G, and −198 T>C HPFH variants (panels c-e).