nature genetics

ARTICLES

https://doi.org/10.1038/s41588-021-00798-y

Check for updates

Transcription factor competition at the γ-globin promoters controls hemoglobin switching

Nan Liu[1,7], Shuqian Xu[1,2,7], Qiuming Yao[1,3], Qian Zhu[4], Yan Kai[4], Jonathan Y. Hsu[3], Phraew Sakon[1], Luca Pinello[3], Guo-Cheng Yuan[4,6], Daniel E. Bauer[1] and Stuart H. Orkin[1,5]✉

BCL11A, the major regulator of fetal hemoglobin (HbF, $\alpha_2\gamma_2$) level, represses γ-globin expression through direct promoter binding in adult erythroid cells in a switch to adult hemoglobin (HbA, $\alpha_2\beta_2$). To uncover how BCL11A initiates repression, we used CRISPR–Cas9, dCas9, dCas9-KRAB and dCas9-VP64 screens to dissect the γ-globin promoters and identified an activator element near the BCL11A-binding site. Using CUT&RUN and base editing, we demonstrate that a proximal CCAAT box is occupied by the activator NF-Y. BCL11A competes with NF-Y binding through steric hindrance to initiate repression. Occupancy of NF-Y is rapidly established following BCL11A depletion, and precedes γ-globin derepression and locus control region (LCR)–globin loop formation. Our findings reveal that the switch from fetal to adult globin gene expression within the >50-kb β-globin gene cluster is initiated by competition between a stage-selective repressor and a ubiquitous activating factor within a remarkably discrete region of the γ-globin promoters.

Human hemoglobin expression undergoes two switches during development: embryonic to fetal and fetal to adult. The second wave initiates in utero and is completed postnatally when the predominant form switches from HbF ($\alpha_2\gamma_2$) in fetal liver to HbA ($\alpha_2\beta_2$) in bone marrow. HbF comprises <2% of total hemoglobin in adults. Increasing the level of HbF ameliorates β-hemoglobinopathies, β-thalassemia and sickle cell disease.

Sequential expression of β-like globin loci in development—ε-embryonic to γ-fetal to β-adult—is reflected in successive looping between the LCR[1] and each globin promoter[2–4]. Transgenic mouse experiments incorporating individual human γ-globin[5], β-globin[6] or the β-globin locus lacking the LCR[7,8] revealed that developmental specificity resides in and about the globin genes rather than the LCR, which enhances transcription. In humans, naturally occurring mutations or deletions in the γ-globin promoters cause continued HbF expression in adults, known as hereditary persistence of fetal hemoglobin (HPFH). The mutations fall in distinct clusters, consistent with repressor binding at these elements. BCL11A and LRF/ZBTB7A, established HbF repressors[9–11], bind directly (relative to γ-globin transcriptional start site (TSS)) to HPFH clusters −115 and −200, respectively[12,13]. We and others have found that BCL11A selectively acts at a distal TGACCA motif (−118 to −113) rather than at a proximal duplicate in the γ-promoters[12,13]. Cas9-mediated mutagenesis of the motif impairs BCL11A binding and γ-repression[12–14]. These findings established local control through *cis* elements as a primary determinant for hemoglobin switching.

Contributions of distal elements of the β-globin locus in HbF repression are unclear. A boundary element encompassing the *HBBP1* gene[15], the ncRNA gene *BGLT3* (ref. [16]) and sequences upstream of the *HBD* gene[17] have been implicated in repression. However, previous work argues against a repressive element upstream of *HBD*[18].

Here we investigate how BCL11A initiates γ-globin gene repression and counteracts transcription activation. Our findings lead to a parsimonious model of hemoglobin switching.

Results

dCas9 placement at BCL11A motif in γ-promoter reduces HbF. We performed CRISPR–Cas9 dense perturbation throughout the β-globin cluster to identify regulatory elements (Methods and Extended Data Fig. 1a). We employed adult-type (low-HbF) HUDEP-2 cells[19] stably expressing either (1) Cas9 to mutate target sequences, (2) inactive Cas9 (dCas9) to bind target sequences but not introduce DNA breaks or (3) transcription-activating dCas9-VP64 or repressive dCas9-KRAB. Following the introduction of pooled, densely spaced guide RNAs (9,293 gRNAs targeting 106 kb of the cluster, one gRNA per ~11 base pairs (bp)), high-HbF cells were isolated to assess the enrichment or depletion of individual gRNAs[20]. In principle, enriched gRNAs may target repressive elements whereas depleted gRNAs may pinpoint activating sequences. We deconvoluted gRNA enrichment scores to identify elements controlling γ-globin expression (Fig. 1a). Based on Cas9-mediated mutagenesis, HbF-repressive sequences resided at *HBB* and *HBD* coding sequences (Extended Data Fig. 1b), and noncoding sequences proximal to *HBG1* (Extended Data Fig. 1c), *HBG2* and around *HBBP1*. We suspect that induction of HbF by targeting of these sequences reflects the secondary effects of Cas9 editing and subsequent DNA repair. In addition, gRNAs targeting the duplicated *HBG1* and *HBG2* genes frequently result in 4.9-kb deletions[14], which may lead to remodeling of local chromatin or removal of repressive elements. Authentic *cis*-acting elements may therefore be obscured. Targeting *HBG1* and *HBG2* coding sequences led to reduced HbF levels, as expected. We did not detect distal HbF-repressive elements, as proposed to reside between

[1]Cancer and Blood Disorders Center, Dana-Farber Cancer Institute and Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. [2]Department of Hematology, Qilu Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China. [3]Molecular Pathology Unit & Center for Cancer Research, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [4]Department of Pediatric Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA. [5]Howard Hughes Medical Institute, Boston, MA, USA. [6]Present address: Department of Genetics and Genomic Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [7]These authors contributed equally: Nan Liu, Shuqian Xu. ✉e-mail: stuart_orkin@dfci.harvard.edu

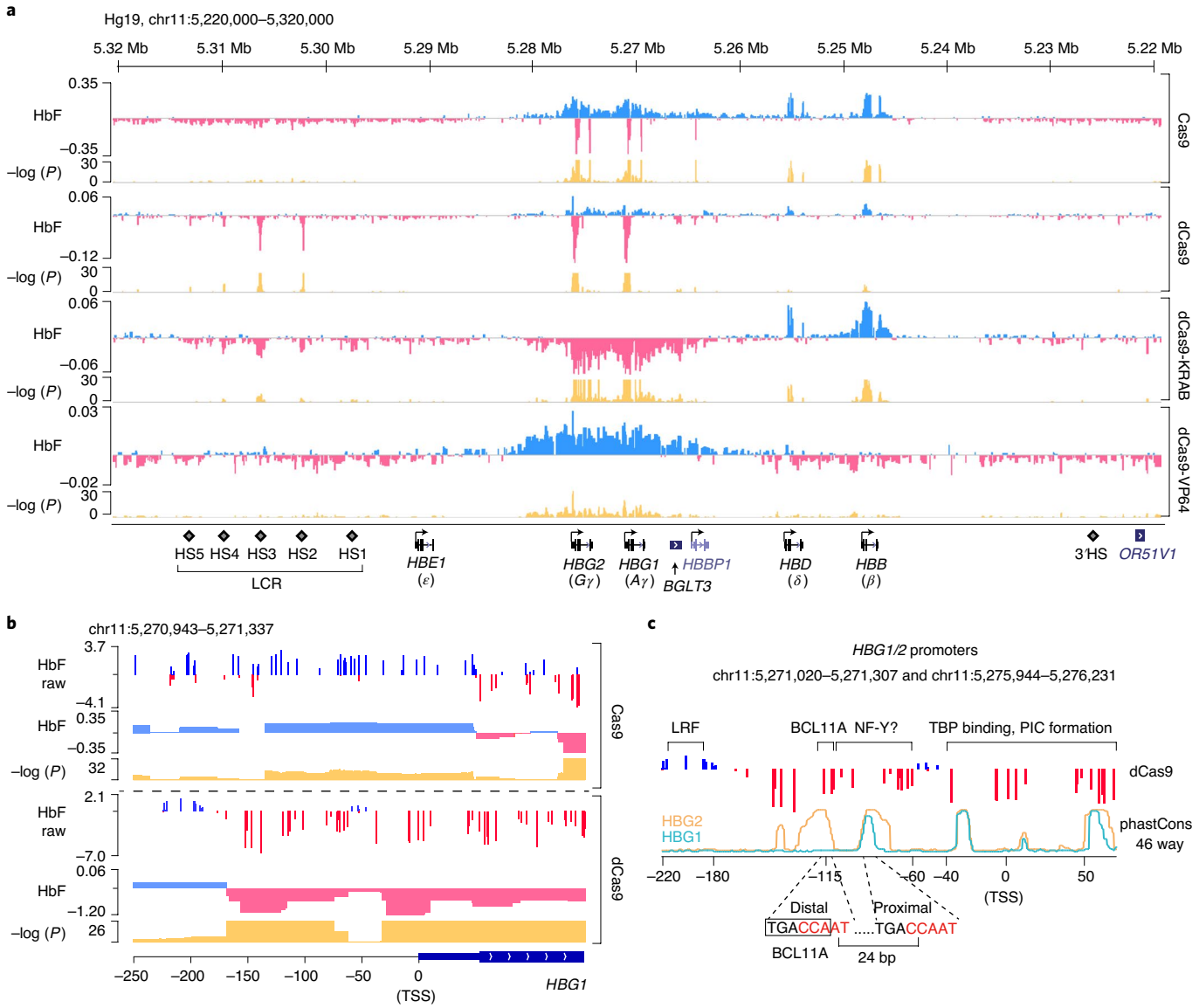


Fig. 1 | dCas9 dense perturbation reveals an activator element at the γ-globin promoters. a, Dense perturbation of genomic sequences around the β-globin gene cluster by pooled gRNA library in Cas9-, dCas9-, dCas9-KRAB- or dCas9-VP64-expressing HUDEP-2 cells. gRNA enrichment in HbF-high cells was used to deconvolute underlying genomic regulatory signal (with HbF tracks displaying beta-coefficient). Corresponding *P* values are shown on a $-\log_{10}$ scale. **b**, Enlarged view of dense perturbation results at the γ-globin (*HBG1*) promoter and first exon. HbF raw scores were calculated as enrichment of reads for each sgRNA in HbF-high compared to the unsorted population at the end of erythroid maturation (shown as $\log_2$(fold change)). HbF tracks depict the deconvoluted regulatory signal as beta-coefficient with associated *P* values below. Minor ticks indicate a value of zero. The result at the *HBG2* promoter was the same as the sequences sharing 99.3% identity. **c**, Schematic structure of γ-globin promoters. The binding sites of LRF/ZBTB7A, BCL11A and TBP are indicated. Sequence conservation of *HBG1* and *HBG2* promoters across 46 vertebrates (phastCons46way) are shown as cyan and orange lines, respectively. Two CCAAT boxes that are potential binding sites of NF-Y are highlighted in red. The distal TGACCA motif through which BCL11A represses γ-globin expression is delineated by a rectangle. Statistical tests of beta-coefficients were performed empirically through bootstrapping and two-tailed tests. Multiple hypothesis testing was accounted for with the Benjamini–Hochberg procedure.

HBD and *HBBP1* (ref. [21]). The effects of dCas9-VP64 (activator) or dCas9-KRAB (repressor) were as expected. Targeting VP64 or KRAB to the body and flanking regions of the γ-globin gene led to induction or reduction of HbF, respectively (Fig. 1a).

An unexpected observation in the dCas9 screen caught our attention. Although dCas9 is inactive for DNA breakage, its targeting to chromatin may interfere with binding of endogenous regulators[22]. The majority of gRNAs in the dCas9 screen were neither enriched nor depleted, suggesting that dCas9 binding at numerous sites did not perturb γ-globin transcription (Fig. 1a). Expression changes were observed when dCas9 was targeted to HS2, HS3 and γ-globin promoters. Within HS2 and HS3, the most depleted gRNAs mapped to composite GATA1–TAL1 motifs (Extended Data Fig. 1d), suggesting that eviction of GATA1 and/or TAL1 at these motifs impairs LCR activity. Within the γ-promoters, however, enriched and depleted gRNAs mapped to discrete regions (Fig. 1b,c). Targeting of dCas9 (or Cas9) to approximately −200 bp, the site at which LRF/ZBTB7A binds[9,13], increased HbF, in agreement with factor displacement. In contrast, targeting of dCas9 to approximately −115 bp, where BCL11A is normally bound to a TGACCA motif[12,13], reduced

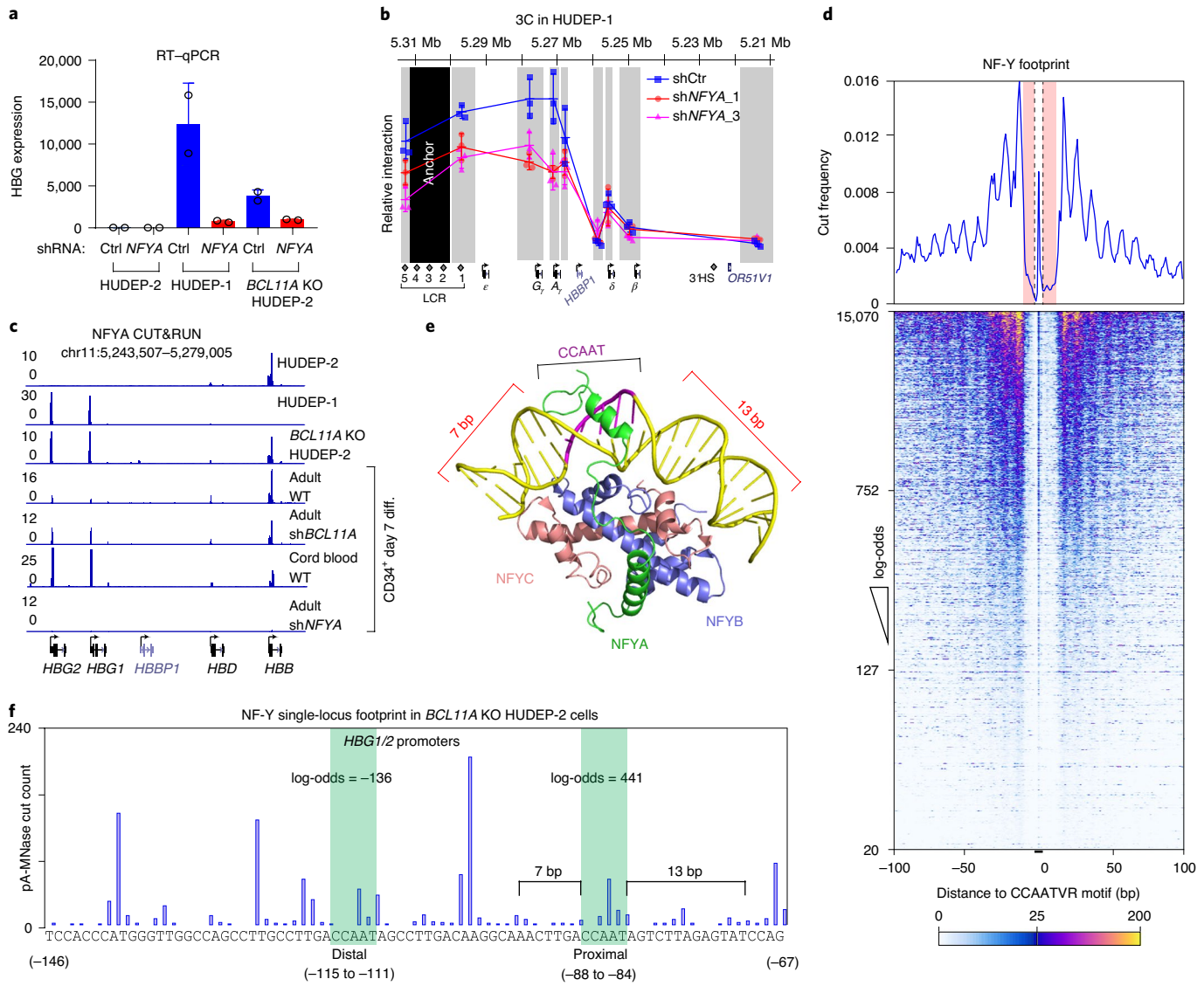


Fig. 2 | NF-Y activates γ-globin through direct binding to the proximal CCAAT. a, RT–qPCR analysis of γ-globin expression in HUDEP-2, HUDEP-1 and *BCL11A* KO HUDEP-2 cells with and without *NFYA* knockdown. The result is shown as mean (s.d.) of two technical replicates and is representative of two biological replicates. *HPRT1* was used throughout as an endogenous control to normalize between samples. ctrl, control. **b**, 3C–qPCR in HUDEP-1 cells with or without *NFYA* knockdown for evaluation of LCR–globin interaction. The EcoRI fragment encompassing HS2-4 of the LCR was used as anchor point. Each gray box indicates a restriction fragment. The result is shown as mean (s.d.) of three technical replicates. **c**, NFYA CUT&RUN in HUDEP-2, HUDEP-1 and *BCL11A* KO HUDEP-2 cells, and in primary human CD34⁺ cells at day 7 of erythroid differentiation (the four bottom tracks are: adult, adult with *BCL11A* knockdown, fetal and adult with *NFYA* knockdown, respectively). CUT&RUN tracks in HUDEP cells are representative of multiple biological replicates. CUT&RUN in CD34⁺ cells with CRISPR editing yielded similar results (Fig. 4). WT, wild type. **d**, Motif footprint analysis of NFYA CUT&RUN. Top: average cut probability of each base surrounding and within CCAATVR motifs was plotted. The core CCAAT motif lies within the dashed lines. Motif flanking regions (7 bp upstream and 11 bp downstream) are shaded red and were protected from nuclease digestion. Bottom: heat map of NF-Y footprints at all NF-Y peaks with log-odds > 20. Each row represents one NF-Y-binding site, ranked by log-odds value. **e**, Structure of NF-Y–DNA complex adapted from ref. [26] and generated with PyMOL. Note that DNA bending is induced by NF-Y binding, and flanking sequences are wrapped around NF-Y through the histone-fold domains of NFYB and NFYC. NFYA is responsible for motif recognition. **f**, Single-locus cut profile at the γ-globin promoters, generated using NFYA CUT&RUN in *BCL11A* KO HUDEP-2 cells. The proximal, but not the distal, CCAAT motif revealed a footprint of NF-Y. The log-odds of NF-Y binding are labeled.

rather than increased HbF expression. As expected, targeting of Cas9 with these gRNAs increased HbF expression.

The effect of positioning dCas9 at the BCL11A-binding site seemed paradoxical. The further reduction in HbF expression was striking, given the low basal HbF level in HUDEP-2 cells (Fig. 1b and Extended Data Fig. 1e). Moreover, dCas9 targeting to multiple positions from −150 to −60 bp led to a similar reduction in HbF. We hypothesized that dCas9 exerted this effect through displacement of an activator.

NF-Y activates γ-globin expression. The cognate BCL11A-binding site is duplicated in the γ-promoters and overlaps CCAAT boxes, a conserved activating motif present in ~30% of promoters (Fig. 1c). Moreover, the CCAAT box often co-occurs with other transcription factor (TF) motifs with precise spatial positioning, suggesting an architectural role in activation[23–25].

NF-Y, a ubiquitous protein complex composed of subunits NFYA, NFYB and NFYC, is a major effector that recognizes CCAAT. NFYA confers sequence specificity, whereas NFYB and NFYC form a

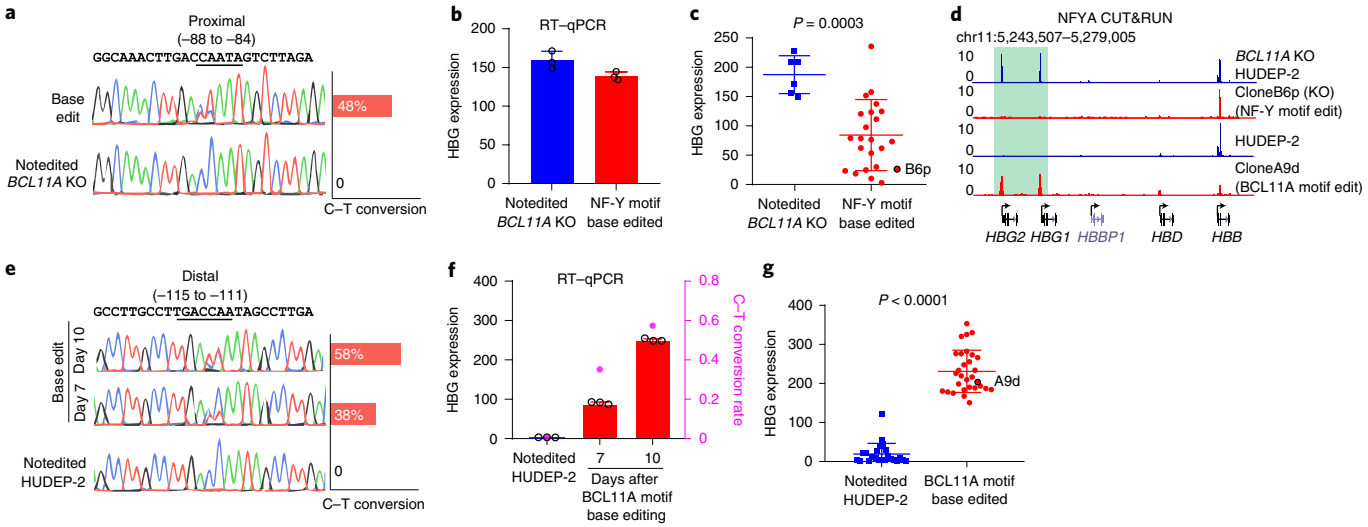


Fig. 3 | Base editing of the NF-Y motif reduces γ-globin expression. a, Sanger sequencing confirmed base editing of the NF-Y motif in the γ-globin promoters. Quantification of editing efficiency is shown as a bar graph on the right. Editing was performed in *BCL11A* KO HUDEP-2 cells. **b**, RT–qPCR showing γ-globin expression level in bulk cells after NF-Y motif editing. The result is shown as mean (s.d.) of three technical replicates. **c**, RT–qPCR analysis of γ-globin expression in multiple independent clones derived from NF-Y motif base editing. Data are shown as mean (s.d.) of multiple independent clones. Not edited, $n=6$; base edited, $n=27$. Two-tailed, unpaired *t*-test, $t=4.010$, d.f = 32. Clone B6p (circled) was used for CUT&RUN analysis (**d**). **d**, NFYA CUT&RUN in *BCL11A* KO HUDEP-2, NF-Y motif edit clone B6p (in *BCL11A* KO background), HUDEP-2 and BCL11A motif edit clone A9d (wild-type background). **e**, Sanger sequencing confirmed base editing of the BCL11A motif in γ-globin promoters. Quantification of editing frequency is shown as a bar graph on the right. Editing was performed in wild-type HUDEP-2 cells. **f**, RT–qPCR showing γ-globin expression level in bulk cells after BCL11A motif editing. Cells collected at different time points showed different degrees of editing. The result is shown as mean (s.d.) of three technical replicates. C–T conversion rates are shown on the right *y* axis. **g**, RT–qPCR analysis of γ-globin expression in multiple independent clones derived from BCL11A motif base editing. Data are shown as mean (s.d.) of multiple independent clones. Not edited, $n=23$; base edited, $n=30$. Two-tailed, unpaired *t*-test, $t=17.11$, d.f = 51. Clone A9d was used for CUT&RUN analysis (**d**).

nucleosome-like structure with their histone-fold domains[26]. NF-Y disrupts nucleosomal structure[27], displays pioneer factor activity[28] and is essential for recruitment of TATA box-binding protein (TBP) and PolII to promoters[28–32].

Although NF-Y has been implicated in globin gene transcription[33–36], its mode of action remained unclear. The paradoxical effect of dCas9 binding at the BCL11A-binding site suggested that dCas9 might prevent binding of an activator, such as NF-Y, in either of two ways. NF-Y might rely on the distal CCAAT box for binding, which is prevented by dCas9 occupancy. Alternatively, dCas9 might displace NF-Y binding at the proximal CCAAT box, which lies within the vicinity of, but not adjacent to, the functional BCL11A-binding site.

We first explored the involvement of NF-Y in γ-globin transcription; short-hairpin RNA-mediated knockdown of *NFYA* reduced γ-globin expression in *BCL11A* knockout (KO) HUDEP-2 cells, and in HUDEP-1 cells[19], where HbF predominates (Fig. 2a, Extended Data Fig. 2a and Source Data). Similar results were obtained in primary human CD34$^+$-cell-derived erythroid precursors (see below). Chromatin immunoprecipitation sequencing (ChIP–seq) of NFYA revealed occupancy of NF-Y at the γ-globin promoters (Extended Data Fig. 2b,c). These data confirmed that NF-Y regulates γ-globin expression.

High-level expression of individual genes in the β-cluster requires looping of the LCR to the respective gene in a developmental-specific manner[4]. We previously showed that knockout of *BCL11A* in HUDEP-2 cells shifts LCR interaction from the β- to the γ-globin gene, as revealed by chromosome conformation capture (3C) analysis[12,37]. Preferential interaction of the LCR with the γ-globin gene in HUDEP-1 and *BCL11A* KO HUDEP-2 cells was markedly reduced following *NFYA* knockdown (Fig. 2b and Extended Data Fig. 2d). Similar results were obtained in CD34$^+$-cell-derived erythroid precursors (see below). These results suggest that NF-Y is required for chromosomal looping between the LCR and the γ-globin genes to achieve high-level expression.

NF-Y binds the proximal CCAAT box in γ-globin promoters. We mapped NF-Y binding within the γ-globin promoters by CUT&RUN, because this nuclease-based method maps TF chromatin occupancy at higher resolution than ChIP–seq[12,38] (Extended Data Fig. 3a). NFYA CUT&RUN in HUDEP-2 cells mapped 11,900 peaks genome wide, 6,357 of which overlapped with NFYA ChIP–seq peaks (Extended Data Fig. 3a,b). The large number of CUT&RUN peaks is due to the lack of an immunoprecipitation step, resulting in indirect peaks caused by protein A–micrococcal nuclease cutting at proximal regions[12,38]. Peaks that reflect direct TF binding are identified by the presence of a TF footprint (see below). shRNA-mediated *NFYA* knockdown led to a global reduction of CUT&RUN signals (Extended Data Fig. 3a), and de novo motif discovery within CUT&RUN peaks identified CCAATVR as the most highly enriched motif (Extended Data Fig. 3c)[39]. We observed NF-Y peaks at γ-globin promoters in HUDEP-1 cells and at the β-globin promoter in HUDEP-2 cells, and at both promoters in *BCL11A* KO cells (Fig. 2c), a pattern consistent with ChIP–seq and expression of the respective genes. Similar results were obtained in CD34$^+$ cells. NF-Y binds strongly at the γ-globin promoters in cord blood CD34$^+$ and *BCL11A* knockdown adult CD34$^+$-derived erythroid cells, but less strongly in adult CD34$^+$-derived erythroid cells (Fig. 2c).

Previous in vitro EMSA experiments indicated that NF-Y binds either of the two CCAAT motifs of the γ-promoter[40,41]. However, the profile of the NF-Y CUT&RUN peak at the γ-globin promoters was biased toward the proximal motif (Extended Data Fig. 3d). As reported previously, the BCL11A CUT&RUN peak is biased to the distal TGACCA BCL11A motif[12] (Extended Data Fig. 3d). These

observations initially raised the possibility that NF-Y might act at the proximal CCAAT site rather than at the distal BCL11A site.

To determine the CCAAT motifs bound by NF-Y in vivo, we performed digital footprinting using CUT&RUN data. In principle, DNA sequences bound by a TF are protected from nuclease digestion while flanking regions are cut more frequently. We generated an average NF-Y footprint profile for all CCAATVR motifs using CUT&RUNTools[39], and observed a strong but atypical footprint (Fig. 2d). Discrete cleavage was observed within the core motif, yet flanking sequences were protected from digestion. This pattern is compatible with the structure of the NF-Y–DNA complex[26]. NF-Y binding induces a kink within the CCAAT motif at the adenine residues (Fig. 2e), the preferred substrates for micrococcal nuclease[42]. Moreover, the NF-Y–DNA complex forms a nucleosomal structure that protects the upstream 7-bp and downstream 13-bp DNA of the motif[26]. The computationally derived CUT&RUN footprint faithfully reflects the native conformation of NF-Y bound to DNA.

The high signal-to-noise ratio of CUT&RUN allows single-locus footprinting[39]. We first plotted the single-locus cut profile of the promoter of an established NF-Y target gene, Cyclin B1 (*CCNB1*)[43]. Duplicated CCAAT motifs, separated by 27 bp, reside in the *CCNB1* promoter and exhibit footprints similar to those determined for all CCAATVR motifs (Extended Data Fig. 3e). The calculated log-odds of NF-Y binding at the two CCAAT motifs were 809 and 526, values suggesting a high probability of direct binding. Similar results were obtained for duplicated CCAAT motifs in the *CDK1* promoter, another NF-Y target (Extended Data Fig. 3e)[44]. These observations are consistent with a report that CCAAT boxes separated by 27 bp (32 bp end to end) allow synergistic binding of two NF-Y complexes[41].

Following validation of NF-Y footprinting by CUT&RUN, we plotted the single-locus cut profile at the γ-promoters in *BCL11A* KO HUDEP-2 cells. In contrast with findings at the *CCNB1* and *CDK1* promoters, the cut profile revealed a characteristic NF-Y footprint at the proximal, but not distal, CCAAT motif (Fig. 2f). The calculated log-odds of binding at proximal and distal CCAAT boxes were 441 and −136, respectively. Similar results were obtained in HUDEP-1 and CD34$^+$ cells (Extended Data Fig. 3f). These data demonstrated that NF-Y selectively binds the proximal CCAAT motif of the γ-globin promoters in native chromatin.

Base editing of proximal NF-Y motif impairs γ-expression. To correlate NF-Y binding at the proximal CCAAT motif and γ-globin expression, we used base editing to convert cytidine to thymidine residues[45]. We employed Target-AID-NG[46], which recognizes NG protospacer-adjacent motif (PAM) sequences, because no gRNAs with the NGG PAM sequence are available at the site. Base editing was conducted in *BCL11A* KO HUDEP-2 cells (Fig. 2a,c). To achieve adequate expression of the base editor, we engineered a split-intein[47] ligated Target-AID-NG (Extended Data Fig. 4a, Source Data and Supplementary Note). Separate expression of Cas9NG-Intein-N and Intein-C-AID resulted in high levels of each protein component. Split-intein-mediated protein ligation in vivo generated full-length Target-AID-NG at a level exceeding that from a vector expressing intact Target-AID-NG (Extended Data Fig. 4a).

The proximal CCAAT NF-Y motif was edited by the split-intein Target-AID-NG (Fig. 3a) at a conversion rate of 48% on day 11. In contrast to bulk-edited cells (Fig. 3b), independent, single-cell clones with base edits exhibited reduced γ-globin expression (Fig. 3c).

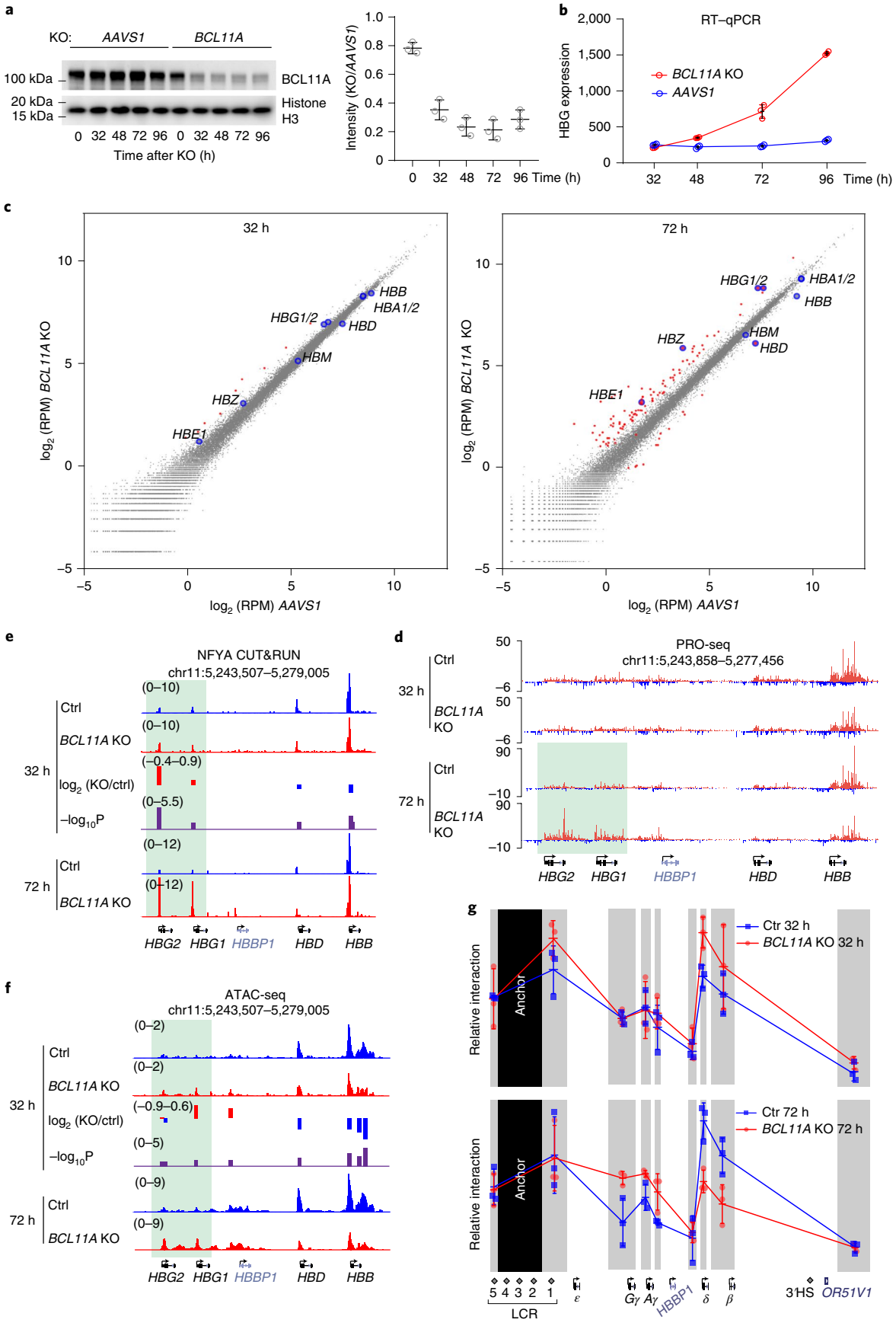
To assess NF-Y occupancy, we performed NF-Y CUT&RUN in nine independent clones. NF-Y binding was diminished at the γ-promoters in all clones (Fig. 3d and Extended Data Fig. 4b). BCL11A CUT&RUN was not conducted, since the clones were generated in a BCL11A null background. Taken together, our results indicated that base editing of the proximal CCAAT motif, the site of NF-Y binding in vivo, impaired NF-Y occupancy and γ-globin gene expression.

Other proteins have been reported to bind CCAAT sequences based on in vitro assays. These include the protein families CCAAT enhancer-binding protein (C/EBP), CCAAT displacement protein (CDP/CUX1) and nuclear factor I (NF1). Subsequent ChIP–seq experiments determined different binding motifs for each protein family, none of which match the γ-globin CCAAT box. We knocked out each protein in *BCL11A* KO HUDEP-2 cells but none revealed a contribution to γ-globin activation approximating that of NF-Y (Extended Data Fig. 4c and Source Data). Therefore, we conclude that NF-Y is the predominant factor acting at the proximal CCAAT box of γ-globin promoters.

Base editing of distal BCL11A motif promotes NF-Y binding. We previously reported that BCL11A functions through the distal TGACCA motif, rather than the proximal TGACCA motif that also overlaps a CCAAT sequence[12]. In contrast, the data presented here indicate that NF-Y acts through the proximal rather than the distal CCAAT box. Therefore, the functionally relevant BCL11A- and NF-Y-binding sites are encompassed within 35 bp of the promoter. We next explored how mutation of the BCL11A-binding site affects NF-Y occupancy. The unexpected finding in the dCas9 screen (Fig. 1) that positioning of dCas9 at the BCL11A motif represses γ-globin led us to ask whether the distal CCAAT also supports activation and is required for NF-Y function. We base edited the distal motif (TGACCAAT to TGATTAAT). Because Target-AID-NG failed to edit the distal motif in HUDEP-2 cells, we used a codon-optimized cytidine base editor[48] and achieved 38 and 58% C–T conversion on days 7 and 10, respectively (Fig. 3e). We observed marked derepression of γ-globin expression in bulk-edited cells, and the level of derepression correlated with the extent of C–T conversion, supporting the role of the distal motif as a repressor-binding site (Fig. 3f and Extended Data Fig. 4d). Edits of a nearby cytidine 10 bp

Fig. 4 | NF-Y rapidly activates γ-globin after acute depletion of BCL11A. a, Left, immunoblot showing the level of BCL11A following CRISPR–Cas9-mediated acute *BCL11A* KO at different time points in adult primary human CD34$^+$ cells undergoing erythroid differentiation (cropped). Right, quantification of BCL11A depletion using ImageJ. The result is representative of two biological replicates. Control cells were edited with *AAVS1* sgRNA. **b**, RT–qPCR analysis of γ-globin expression level at different time points after acute depletion of BCL11A. Data are shown as mean (s.d.) of three technical replicates. **c**, Scatter plot of Pro-seq data at 32 h (left) and 72 h (right) after acute depletion of BCL11A. The *x* axis represents $\log_2$ (reads per million (RPM)) of each gene in control experiments (*AAVS1*), and the *y* axis represents $\log_2$ (RPM) of each gene in *BCL11A* KO experiments. Each dot represents one gene. Globin genes are enlarged and highlighted by blue circles. Genes that are differentially transcribed between control and KO ($\log_2$ (KO/control) <−1 or >1) are highlighted in red. The results are shown as means of two biological replicates. **d**, Pro-seq tracks at the β-globin locus. Transcripts of positive and negative strands are denoted by different colors. γ-globin remained repressed in *BCL11A* KO at 32 h. Derepression became evident at 72 h, as highlighted in green. **e**,**f**, NFYA CUT&RUN (**e**) and ATAC-seq (**f**) in CD34$^+$ cells undergoing erythroid differentiation after 32 or 72 h of BCL11A acute depletion. Quantification of KO/control (ctrl) and corresponding *P* values are reported by MAnorm. **g**, 3C–qPCR in CD34$^+$ cells undergoing erythroid differentiation after 32 (upper) or 72 (lower) h of BCL11A acute depletion. The EcoRI fragment encompassing HS2-4 of the LCR was used as the anchor point to evaluate LCR–globin interaction. The result is shown as mean (s.d.) of three technical replicates and is representative of two biological replicates.

distant with a different gRNA failed to perturb γ-globin expression, indicating that the effects are not attributable to base-editor binding or base editing by itself (Extended Data Fig. 4e). We sorted bulk-edited cells according to intracellular HbF levels on day 10 and compared C–T conversion in high- and low-HbF populations. The distal motif was 87% edited in HbF-high cells but only 7.4% in



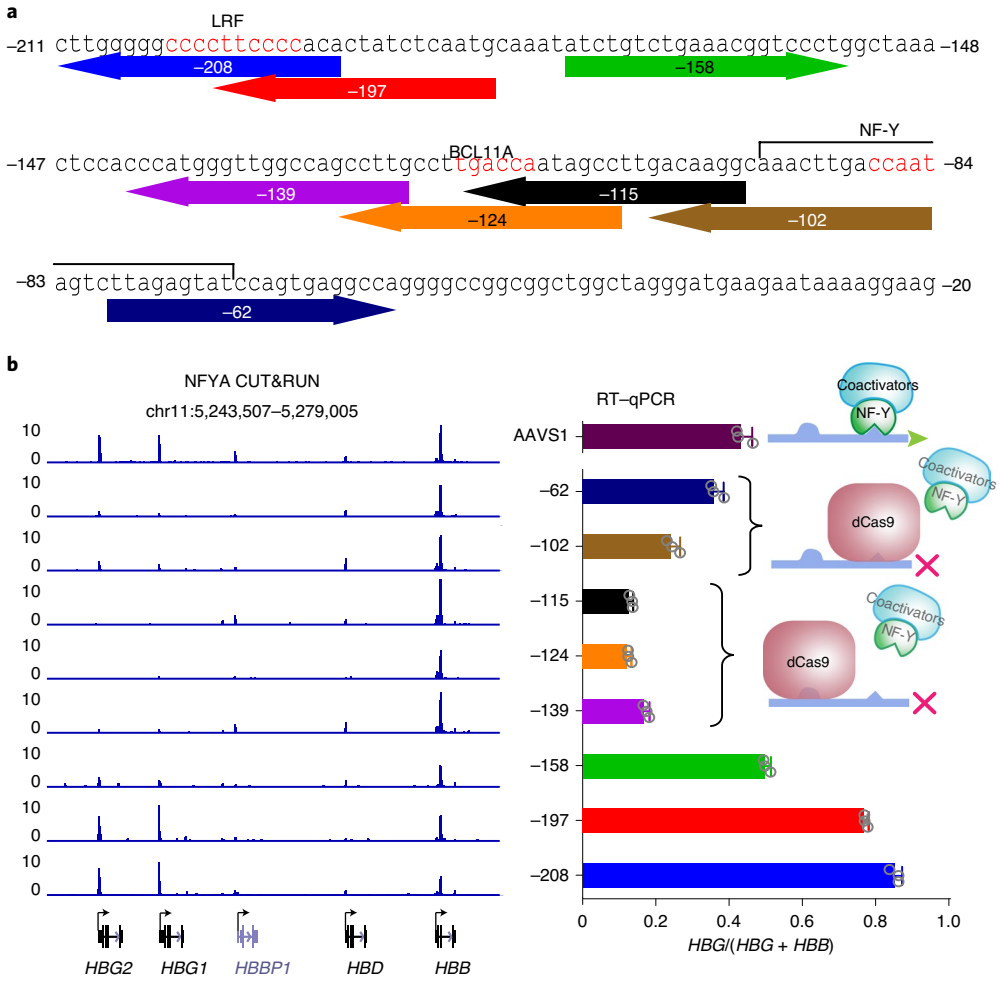


Fig. 5 | NF-Y binding is affected by steric hindrance at the γ-globin promoters. a, The sequence of γ-globin promoters. The eight sgRNAs used for the dCas9 disruption experiment are shown below the sequence, and color coded. The numbers (−208 and so on) represent the distances between the TSS and positions of the 17th nucleotide of each sgRNA. The positions of LRF, BCL11A and NF-Y motifs are labeled in red. The flanking sequences of the NF-Y motif are also involved in NF-Y binding. **b**, Right, RT–qPCR analysis of the percentage of γ-globin in *BCL11A* KO HUDEP-2 cells expressing dCas9 with different sgRNAs. Data are shown as mean (s.d.) of three technical replicates and are representative of two biological replicates. Left, NFYA CUT&RUN in these cells. The cartoon indicates the deduced protein binding at the promoters.

HbF-low cells (Extended Data Fig. 4f). Single clones were isolated, and γ-globin expression was assessed by quantitative PCR with reverse transcription (RT–qPCR) and flow cytometry. All clones with C–T conversion at the BCL11A motif exhibited increased γ-globin expression (Fig. 3g and Extended Data Fig. 4g), confirming repression through the distal motif. We observed strong NF-Y binding in BCL11A-motif-edited clones (Fig. 3d and Extended Data Fig. 4b). Single-locus footprinting revealed protection of the proximal CCAAT motif, providing additional evidence for NF-Y action at the proximal motif (Extended Data Fig. 4h). Thus, mutation of the distal BCL11A motif was accompanied by NF-Y occupancy at the proximal CCAAT motif and robust γ-globin expression. These results support base editing of the distal BCL11A motif for reactivation of HbF as treatment for sickle cell disease or β-thalassemia[49].

BCL11A depletion leads to NF-Y binding and γ-transcription. Knockout and editing experiments provide a static view at the endpoint of a genetic pertubation, and the resulting interpretation may be confounded by secondary effects. To explore the dynamics of NF-Y binding and downstream events following acute depletion of BCL11A, we knocked out BCL11A in CD34$^+$ cells undergoing erythroid differentiation by nucleofection of the Cas9–sgRNA ribonucleoprotein complex (RNP) (Extended Data Fig. 5a). BCL11A protein decreased to 35% at 32 h after RNP delivery (Fig. 4a and Source Data), while γ-globin transcripts remained unchanged at 32 h and increased thereafter (Fig. 4b). Because RT–qPCR quantifies the total level of old and newly synthesized RNA, it does not directly measure transcriptional dynamics. To compare γ-globin transcription rates at different times, we analyzed nascent transcripts using precision nuclear run-on sequencing (Pro–seq)[50]. Two replicates were highly correlated (Extended Data Fig. 5b). We observed promoter pausing of PolII, indicating a successful Pro–seq experiment (Extended Data Fig. 5c). Acute depletion of BCL11A did not lead to global changes in transcription: only 9 and 110 genes were transcribed differentially at 32 and 72 h, respectively (Fig. 4c). Consistent with RT–qPCR, γ-globin transcription remained unchanged at 32 h but was increased at 72 h in *BCL11A* KO (Fig. 4c,d and Extended Data Fig. 5d). We chose the times 32 and 72 h for detailed analysis, the former representing a time at which changes at globin promoters are initiated whereas the latter approximates an endpoint.

We compared chromatin accessibility, TF binding and LCR–globin looping by assay for transposase-accessible chromatin using

sequencing (ATAC-seq), CUT&RUN and 3C–qPCR, respectively. At 72 h, NF-Y and TBP exhibited strong occupancy at the γ-promoters in *BCL11A* KO (Fig. 4e and Extended Data Fig. 5e). Chromatin accessibility increased (Fig. 4f) and contacts between the LCR and γ-globin genes were more frequent, whereas those with the β-globin gene were reduced (Fig. 4g, bottom). Combined depletion of NF-Y and BCL11A impaired γ-globin expression (Extended Data Fig. 5f,g and Source Data) and reduced LCR–γ-globin interaction at 72 h (Extended Data Fig. 5h). These results confirmed the opposing roles of BCL11A and NF-Y. Data obtained at 32 h after *BCL11A* KO provided insights into the order of events. At this time, we observed increased NF-Y and TBP binding at the γ-promoters (Fig. 4e and Extended Data Fig. 5e), with a fold increase of 1.9 ($P=3.2\times10^{-6}$) and 1.6 ($P=7.9\times10^{-5}$), respectively, as quantified by the MAnorm algorithm[51]. Chromatin accessibility (Fig. 4f) showed a small but statistically insignificant increase. LCR–γ-globin gene interaction (Fig. 4g, top) and γ-globin transcription (Fig. 4b–d) exhibited negligible changes. These data reveal that NF-Y binds rapidly to the γ-globin promoters following BCL11A depletion to open up local chromatin, which precedes the formation of enhancer–promoter contacts and transcriptional activation, consistent with NF-Y pioneer activity[25,27,28].

BCL11A displaces NF-Y by steric hindrance. The above findings suggested a dynamic model for how BCL11A initiates γ-globin repression. We posit that binding of BCL11A (or dCas9) at the distal TGACCA sequence constitutes a steric barrier to NF-Y binding at the proximal CCAAT motif.

To test this model, we explored how recruitment of dCas9 at different positions affects γ-globin expression and NF-Y binding. We transduced eight gRNAs of the dCas9 screen individually into dCas9-expressing cells (Fig. 5a). To detect changes in NF-Y binding following dCas9 recruitment, we employed *BCL11A* KO HUDEP-2 cells rather than wild-type HUDEP-2 cells, in which NF-Y binding and γ-globin expression are low. As validation for subsequent experiments, we confirmed that dCas9 recruitment to sequences at −62 to −139 led to reduced γ-globin expression, and that dCas9 recruitment to −197 or −208 led to an increase in γ-globin due to displacement of ZBTB7A[9] (Fig. 5b). CUT&RUN revealed that NF-Y occupancy was reduced, although only partially, following placement of dCas9 at −102 or −62, which overlaps the NF-Y motif or downstream flanking sequence, respectively, suggesting that dCas9 displaces NF-Y following recruitment to an NF-Y-binding site. These findings are consistent with the modest reduction of γ-globin expression. Recruitment of dCas9 to the distal BCL11A motif (gRNAs −115 and −124) or even further (gRNA −139) perturbs NF-Y binding at the proximal CCAAT motif and reduces γ-globin expression. Therefore, dCas9 targeted to the BCL11A motif impairs NF-Y binding to its motif 24 bp downstream.

Discussion

Here we reveal that γ-globin repression in adult cells is controlled by competitive binding of BCL11A and NF-Y within the γ-globin promoters. This finding was inspired by a paradoxical observation that dCas9 displacement of BCL11A binding at the γ-globin promoters further represses, rather than activates, expression. Protein binding (either BCL11A or dCas9) at the BCL11A motif presents a steric barrier to the activator NF-Y, whose relevant binding motif lies 24 bp downstream. Competitive binding between BCL11A and NF-Y determines the activity of the γ-globin promoter and transcription output.

We propose a TF competition model for initiation of hemoglobin switching (Fig. 6). In adult cells, BCL11A expression rises with erythroid commitment and peaks during erythroblast maturation, at which time BCL11A binding is detectable at the γ-globin promoters[12]. BCL11A prevents NF-Y binding through steric hindrance to initiate

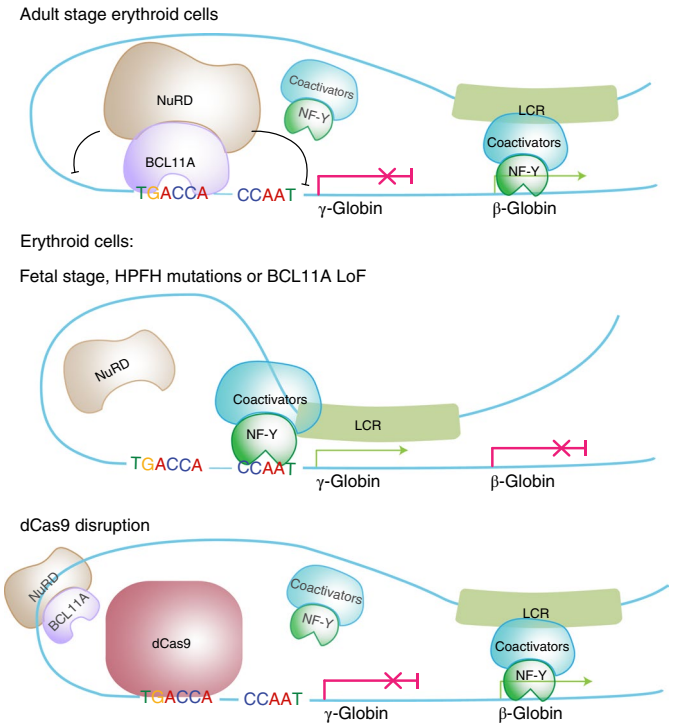


Fig. 6 | A simplified model for hemoglobin switching. Competitive binding between NF-Y and BCL11A controls hemoglobin switching. In fetal stage erythroid cells, or in cells with HPFH mutations or BCL11A loss of function (LoF), NF-Y binds to the γ-globin promoters and activates expression. In adult stage erythroid cells, BCL11A prevents NF-Y binding and represses γ-globin in concert with NuRD. LRF/ZBTB7A independently recruits NuRD and represses γ-globin through binding to the ~200-bp region of the γ-globin promoters (not illustrated in the model). dCas9 binding at the BCL11A motif is sufficient to disrupt NF-Y binding and repress γ-globin expression. When γ-globin is silenced, NF-Y may bind to β-globin and regulate its expression. Other known positive regulators of β-globin, including GATA1, KLF1 and LDB1, are not shown.

repression, followed by involvement of corepressors. BCL11A interacts with the NuRD corepressor complex[52], which possesses nucleosome remodeling and histone deacetylation activities. NuRD recruitment promotes chromatin compaction and the repressive state of the γ-globin promoters. NuRD components are 100 times more abundant than coactivators in erythroid cells[53], and this cofactor imbalance ensures the competitive advantage of BCL11A–NuRD. The roles of BCL11A and NuRD can be bypassed by placing dCas9 at the BCL11A motif, which displaces NF-Y. In fetal progenitors, BCL11A protein is expressed at a low level[10,54]. NF-Y occupies the γ-globin promoters and modifies chromatin through recruitment of cofactors (for example, P300) to activate γ-globin expression. NF-Y activates γ-globin similarly in erythroid cells harboring mutations at the BCL11A binding motif or *BCL11A* gene, either in genetically modified cellular models[12] or in individuals with HPFH.

In addition to BCL11A-NF-Y competition, the coordinated action of other factors probably contributes to γ-globin repression. LRF/ZBTB7A, a second HbF repressor, acts at −200 bp and independently recruits NuRD[9,13]. We speculate that the concerted action of BCL11A and LRF/ZBTB7A for NuRD recruitment stabilizes the repressed state of the γ-promoters and that loss of either factor increases chromatin accessibility, thereby allowing NF-Y to gain partial competitive advantage. Other repressors suggested to act at the γ-globin promoters (such as TR2/TR4, SOX6, KLF3 and COUP-TFII)[55] lack supporting evidence to substantiate postulated roles.

The effects following acute depletion of BCL11A suggest that TF binding precedes and directs LCR–globin looping. Distinct proteins recruited by TFs and the histone marks they deposit may stabilize or restrict looping depending on their biochemical compatibility, as determined by their specific interactions[56]. This conclusion is consistent with the finding that TF binding is an upstream event of enhancer–promoter looping at the α-globin locus[57].

Our work presents a parsimonious model for initiation of hemoglobin switching, and highlights competitive TF binding within 35 bp of the γ-globin promoters in determination of stage specificity for the >50-kb β-globin cluster. Spatial constraints have previously been reported for NF-Y promoter binding. In vitro studies revealed that two NF-Y molecules exhibit synergistic binding to double CCAAT motifs when the distance between motifs is 27 bp (three helical turns). NF-Y may still occupy double CCAAT motifs without synergy when the distance is reduced by a few base pairs, but cobinding was not detected when the distance was reduced to 17 bp[40,41]. C/EBP and NF-Y binding are mutually exclusive in the albumin gene promoter, where a CCAAT motif is juxtaposed to a C/EBP site. Nevertheless, the factors may cobind if the distance between the two motifs is increased by 10 bp[58]. Studies with synthetic promoters demonstrate that promoter activity is reduced if the distance between *cis* elements is <10–15 bp[59]. More broadly, instances of competitive protein binding at promoters are observed in lower organisms whose TFs usually lack effector domains[60]: for example, the *Escherichia coli* lac repressor sterically restricts RNA polymerase binding. Collectively, TF competition is widely used for modulation of gene activity. In the context of globin gene regulation, the precise architecture of the γ-globin promoters has evolved to provide a platform for dynamic eviction of a ubiquitous activator by a stage-selective repressor as the basis for the fetal-to-adult globin switch.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00798-y.

Received: 3 June 2020; Accepted: 21 January 2021;
Published online: 1 March 2021

References

1. Hay, D. et al. Genetic dissection of the α-globin super-enhancer in vivo. *Nat. Genet.* **48**, 895–903 (2016).
2. Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. & Fraser, P. Long-range chromatin regulatory interactions in vivo. *Nat. Genet.* **32**, 623–626 (2002).
3. Tolhuis, B. et al. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10**, 1453–1465 (2002).
4. Palstra, R. J. et al. The β-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35**, 190–194 (2003).
5. Chada, K., Magram, J. & Costantini, F. An embryonic pattern of expression of a human fetal globin gene in transgenic mice. *Nature* **319**, 685–689 (1986).
6. Magram, J., Chada, K. & Costantini, F. Developmental regulation of a cloned adult β-globin gene in transgenic mice. *Nature* **315**, 338–340 (1985).
7. Starck, J. et al. Developmental regulation of human gamma- and beta-globin genes in the absence of the locus control region. *Blood* **84**, 1656–1665 (1994).
8. Bender, M. A., Bulger, M., Close, J. & Groudine, M. β-Globin gene switching and DNase I sensitivity of the endogenous β-globin locus in mice do not require the locus control region. *Mol. Cell* **5**, 387–393 (2000).
9. Masuda, T. et al. Transcription factors LRF and BCL11A independently repress expression of fetal hemoglobin. *Science* **351**, 285–289 (2016).
10. Sankaran, V. G. et al. Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322**, 1839–1842 (2008).
11. Sankaran, V. G. et al. Developmental and species-divergent globin switching are driven by BCL11A. *Nature* **460**, 1093–1097 (2009).
12. Liu, N. et al. Direct promoter repression by BCL11A controls the fetal to adult hemoglobin switch. *Cell* **173**, 430–442 (2018).
13. Martyn, G. E. et al. Natural regulatory mutations elevate the fetal globin gene via disruption of BCL11A or ZBTB7A binding. *Nat. Genet.* **50**, 498–503 (2018).
14. Traxler, E. A. 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).
15. Huang, P. et al. Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. *Genes Dev.* **31**, 1704–1713 (2017).
16. Ivaldi, M. S. et al. Fetal g-globin genes are regulated by the BGLT3 long noncoding RNA locus. *Blood* **132**, 1963–1973 (2018).
17. Sankaran, V. G. et al. A functional element necessary for fetal hemoglobin silencing. *N. Engl. J. Med.* **365**, 807–814 (2011).
18. Gaensler, K. M. L. et al. Sequences in the Aγ-δ intergenic region are not required for stage-specific regulation of the human β-globin gene locus. *Proc. Natl Acad. Sci. USA* **100**, 3374–3379 (2003).
19. Kurita, R. et al. Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PLoS ONE* **8**, e59890 (2013).
20. Canver, M. C. et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192–197 (2015).
21. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
22. Shariati, S. A. et al. Reversible disruption of specific transcription factor-DNA interactions using CRISPR/Cas9. *Mol. Cell* **74**, 622–633 (2019).
23. Mantovani, R. The molecular biology of the CCAAT-binding factor NF-Y. *Gene* **239**, 15–27 (1999).
24. Dolfini, D., Zambelli, F., Pedrazzoli, M., Mantovani, R. & Pavesi, G. A high definition look at the NF-Y regulome reveals genome-wide associations with selected transcription factors. *Nucleic Acids Res.* **44**, 4684–4702 (2016).
25. Fleming, J. D. et al. NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. *Genome Res.* **23**, 1195–1209 (2013).
26. Nardini, M. et al. Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell* **152**, 132–143 (2013).
27. Coustry, F., Hu, Q., de Crombrugghe, B. & Maity, S. N. CBF/NF-Y functions both in nucleosomal disruption and transcription activation of the chromatin-assembled topoisomerase IIalpha promoter. Transcription activation by CBF/NF-Y in chromatin is dependent on the promoter structure. *J. Biol. Chem.* **276**, 40621–40630 (2001).
28. Oldfield, A. J. et al. Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. *Mol. Cell* **55**, 708–722 (2014).
29. Bellorini, M. et al. CCAAT binding NF-Y-TBP interactions: NF-YB and NF-YC require short domains adjacent to their histone fold motifs for association with TBP basic residues. *Nucleic Acids Res.* **25**, 2174–2181 (1997).
30. Frontini, M. et al. NF-Y recruitment of TFIID, multiple interactions with histone fold TAFIIs. *J. Biol. Chem.* **277**, 5841–5848 (2002).
31. Kabe, Y. et al. NF-Y is essential for the recruitment of RNA polymerase II and inducible transcription of several CCAAT box-containing genes. *Mol. Cell. Biol.* **25**, 512–522 (2005).
32. Oldfield, A. J. et al. NF-Y controls fidelity of transcription initiation at gene promoters through maintenance of the nucleosome-depleted region. *Nat. Commun.* **10**, 3072 (2019).
33. 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).
34. Duan, Z., Stamatoyannopoulos, G. & Li, Q. Role of NF-Y in in vivo regulation of the gamma-globin gene. *Mol. Cell. Biol.* **21**, 3083–3095 (2001).
35. Fang, X., Han, H., Stamatoyannopoulos, G. & Li, Q. Developmentally specific role of the CCAAT box in regulation of human γ-globin gene expression. *J. Biol. Chem.* **279**, 5444–5449 (2004).
36. Martyn, G. E., Quinlan, K. G. R. & Crossley, M. The regulation of human globin promoters by CCAAT box elements and the recruitment of NF-Y. *Biochim. Biophys. Acta Gene Regul. Mech.* **1860**, 525–536 (2017).
37. Xu, J. et al. Transcriptional silencing of γ-globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev.* **24**, 783–789 (2010).
38. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).
39. Zhu, Q., Liu, N., Orkin, S. H. & Yuan, G. C. CUT and RUNTools: a flexible pipeline for CUT and RUN processing and footprint analysis. *Genome Biol.* **20**, 192 (2019).
40. Liberati, C., Ronchi, A., Lievens, P., Ottolenghi, S. & Mantovani, R. NF-Y organizes the γ-globin CCAAT boxes region. *J. Biol. Chem.* **273**, 16880–16889 (1998).

41. Liberati, C., di Silvio, A., Ottolenghi, S. & Mantovani, R. NF-Y binding to twin CCAAT boxes: role of Q-rich domains and histone fold helices. *J. Mol. Biol.* **285**, 1441–1455 (1999).
42. Luo, D. et al. MNase, as a probe to study the sequence-dependent site exposures in the +1 nucleosomes of yeast. *Nucleic Acids Res.* **46**, 7124–7137 (2018).
43. Hu, Q., Lu, J.-F., Luo, R., Sen, S. & Maity, S. N. Inhibition of CBF/NF-Y mediated transcription activation arrests cells at G2/M phase and suppresses expression of genes activated at G2/M phase of the cell cycle. *Nucleic Acids Res.* **34**, 6272–6285 (2006).
44. Kao, C. Y., Tanimoto, A., Arima, N., Sasaguri, Y. & Padmanabhan, R. Transactivation of the human cdc2 promoter by adenovirus E1A induces the expression and assembly of a heteromeric complex consisting of the CCAAT box binding factor, CBF/NF-Y, and a 110-kDa DNA-binding protein. *J. Biol. Chem.* **274**, 23043–23051 (1999).
45. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
46. Nishimasu, H. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
47. Stevens, A. J. et al. A promiscuous split intein with expanded protein engineering applications. *Proc. Natl Acad. Sci. USA* **114**, 8538–8543 (2017).
48. Zafra, M. P. et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* **36**, 888–893 (2018).
49. Wang, L. et al. Reactivation of γ-globin expression through Cas9 or base editor to treat β-hemoglobinopathies. *Cell Res.* **30**, 276–278 (2020).
50. Mahat, D. B. et al. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* **11**, 1455–1476 (2016).
51. Shao, Z., Zhang, Y., Yuan, G. C., Orkin, S. H. & Waxman, D. J. MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol.* **13**, R16 (2012).
52. Xu, J. et al. Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. *Proc. Natl Acad. Sci. USA* **110**, 6518–6523 (2013).
53. Gillespie, M. A. et al. Absolute quantification of transcription factors reveals principles of gene regulation in erythropoiesis. *Mol. Cell* https://doi.org/10.1016/j.molcel.2020.03.031 (2020).
54. Basak, A. et al. Control of human hemoglobin switching by LIN28B-mediated regulation of BCL11A translation. *Nat. Genet.* **52**, 138–145 (2020).
55. Wilber, A., Nienhuis, A. W. & Persons, D. A. Transcriptional regulation of fetal to adult hemoglobin switching: new therapeutic opportunities. *Blood* **117**, 3945–3953 (2011).
56. van Arensbergen, J., van Steensel, B. & Bussemaker, H. J. In search of the determinants of enhancer–promoter interaction specificity. *Trends Cell Biol.* **24**, 695–702 (2014).
57. Oudelaar, A. M. et al. Dynamics of the 4D genome during in vivo lineage specification and differentiation. *Nat. Commun.* **11**, 2722 (2020).
58. Milos, P. M. & Zaret, K. S. A ubiquitous factor is required for C/EBP-related proteins to form stable transcription complexes on an albumin promoter segment in vitro. *Genes Dev.* **6**, 991–1004 (1992).
59. Stracke, R., Thiedig, K. & Kuhlmann, M. What have we learned about synthetic promoter construction? *Plant Synth. Promot.* **1482**, 1–13 (2016).
60. Rojo, F. Repression of transcription initiation in bacteria. *J. Bacteriol.* **181**, 2987–2991 (1999).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021, corrected publication 2021

Methods

Culture of primary cells. Human peripheral blood stem cells (CD34+) (G-CSF mobilized, CD34+ enriched) were purchased from the Center of Excellence in Hematology at the Fred Hutchinson Cancer Research Center. These deidentified samples were exempted from Boston Children's Hospital Institutional Review Board approval. Cells were thawed and recovered to erythroid differentiation medium (EDM) (IMDM, Corning, no. 15-016-CV) supplemented with 330 μg ml^{-1} holo-human transferrin, 10 μg ml^{-1} recombinant human insulin, 2 IU ml^{-1} heparin, 5% inactivated plasma, 3 IU ml^{-1} erythropoietin and 2 mM L-glutamine, with three supplements (10^{-6} M hydrocortisone, 100 ng ml^{-1} stem cell factor (SCF) and 5 ng ml^{-1} IL-3), for 7 days to allow erythroid differentiation, and were then further differentiated in EDM with one supplement (100 ng ml^{-1} SCF).

Culture of immortalized cell lines. Human HEK293T cells (female) were purchased from the American Type Culture Collection and cultured in DMEM, high-glucose (Thermo Fisher Scientific, no. 11965) with 10% FCS and 2 mM L-glutamine and passaged every 3 days.

HUDEP-1 and HUDEP-2 lines (human umbilical cord blood-derived erythroid progenitor, male) were provided by Y. Nakamura. HUDEP-2 cells were maintained in expansion medium—StemSpan SFEM (STEMCELL Technologies, no. 09650) with SCF (50 ng ml^{-1}), erythropoietin (3 IU ml^{-1}), dexamethasone (10^{-6} M) and doxycycline (1 μg ml^{-1})—and passaged every 3 days. Erythroid differentiation was carried out by replacement of the medium with EDM2 (IMDM, Corning, no. 15-016-CV) supplemented with 330 μg ml^{-1} holo-human transferrin, 10 μg ml^{-1} recombinant human insulin, 2 IU ml^{-1} heparin, 5% inactivated plasma, 3 IU ml^{-1} erythropoietin, 2 mM L-glutamine, 100 ng ml^{-1} SCF and 1 μg ml^{-1} doxycycline).

Design and synthesis of lentiviral sgRNA libraries. We identified every 20-mer sequence upstream of the SpCas9 NGG PAM sequence at the *HBB* locus from hg19 chr11:5,220,000–5,326,000 (106 kb comprising the *HBB* gene cluster inclusive of the HS5 to 3'HS1 distal elements). We designed 10,383 spacer sequences (including 100 negative control genome-targeting spacers with 95 safe-targeting[61] and five *AAVS1*-targeting spacers and 990 positive controls at *BCL11A*). The sgRNA oligos (synthesized by CustomArray) were cloned using a Gibson Assembly master mix (New England Biolabs) into lentiGuide-Puro (Addgene plasmid no. 52963, ref. [62]) that had been BsmBI digested, gel purified and dephosphorylated. Gibson Assembly products were transformed to electrocompetent cells (E. cloni 10 G ELITE Electrocompetent Cell, Lucigen, no. 60052). Sufficient colonies were isolated to ensure >1,000 colonies per spacer sequence. Plasmid libraries were deeply sequenced to confirm representation.

To produce lentivirus, HEK293T cells were cultured with DMEM (Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Omega Scientific) and 2% penicillin/streptomycin (Life Technologies) in 15-cm tissue-culture-treated Petri dishes. HEK293T cells were transfected at 80% confluence in 12 ml of medium with 13.3 μg of psPAX2 (Addgene plasmid no. 12260, a gift from D. Trono), 6.7 μg of VSV-G (Addgene plasmid no. 14888, a gift from T. Reya) and 20 μg of the lentiviral construct plasmid of interest using 180 μg of linear polyethylenimine (Polysciences). Medium was changed 16–24 h after transfection. Lentiviral supernatant was collected at 48 and 72 h post transfection and subsequently concentrated by ultracentrifugation (24,000 r.p.m. for 2 h at 4 °C with a Beckman Coulter SW 32 Ti rotor).

Transducution of HUDEP-2 cells with lentiviral library. HUDEP-2 cells were transduced with lentivirus encoding various Cas9 or dCas9 variants (Cas9: LentiCas9-Blast, Addgene plasmid no. 52962 (ref. [62]); dCas9: pGH125_dCas9-Blast, Addgene plasmid no. 85417 (ref. [63]): dCas9-KRAB: pHR-SFFV-dCas9-BFP-KRAB, Addgene plasmid no. 46911 (ref. [64]); dCas9-VP64: pHRdSV40-dCas9-10xGCN4_v4-P2A-BFP, Addgene plasmid no. 60903 and pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS, Addgene plasmid np. 60904 (ref. [65])) to produce stably expressing cells. The cells were then transduced with the sgRNA lentivirus library and selected with 1 μg ml^{-1} puromycin. Cells were transduced at multiplicity of infection = 0.3. Cell numbers were maintained throughout the experiment, with at least 1,000 cells per spacer sequence. After sgRNA transduction, cells were cultured in EDM2 medium for 12 days.

HbF staining and fluorescence-activated cell sorting. After differentiation, intracellular staining was performed by fixing cells with 0.05% glutaraldehyde (grade II; Sigma) for 10 min at room temperature. Cells were centrifuged for 5 min at 600*g* and then resuspended in 0.1% Triton X-100 (Life Technologies) for 5 min at room temperature for permeabilization. Triton X-100 was diluted with PBS plus 0.1% BSA and then centrifuged at 600*g* for 15 min. Cells were stained with antibodies for HbF (clone HbF-1 with fluorescein isothiocyanate or allophycocyanin conjugation; Life Technologies, no. MHFH01) for 20 min in the dark. Cells were washed to remove unbound antibody before fluorescence-activated cell sorting (FACS). HbF antibodies (0.2 μg per 5 million cells) were used, where a population of cells with the top 10% HbF expression was sorted by FACS. A total of 12 samples were collected: three biological replicates each of Cas9-sorted, Cas9-unsorted, dCas9-sorted and dCas9-unsorted.

Amplification of sgRNA and deep sequencing. After collection of unsorted total and HbF-high-sorted cell populations at the end of erythroid maturation culture, library preparation and deep sequencing were performed. Briefly, genomic DNA was extracted using the Qiagen Blood and Tissue kit. sgRNA integrant PCR for amplification of spacer sequences was performed with Herculase II reaction buffer (1×), forward and reverse primers (0.5 μM each), 8% dimethyl sulfoxide (DMSO), deoxynucleotide triphosphates (0.25 μM each) and Herculase II Fusion DNA Polymerase (0.5 reactions) under the following cycling conditions: 95 °C for 2 min; 20 cycles of 95 °C for 15 s, 60 °C for 20 s and 72 °C for 30 s; and 72 °C for 5 min. Multiple reactions of ≤200 ng each were used to amplify from 6.6 μg of gDNA (~1×10^6 cell genomes) per pool. Samples were subjected to additional rounds of PCR to add sequencing adapters and indices before Illumina sequencing.

Data analysis for Cas9 and dCas9 screens. In total, 8,639 sgRNAs (including 977 positive and 100 negative control guides) were retained after filtering by MIT specificity score ≥ 5. For each sgRNA, read counts were calculated from raw fastq files with custom codes. Quality control metadata are shown in Supplementary Table 1. DESeq2 (ref. [21]) (run on R 4.0.1) was then used to compare sgRNA counts between sorted and unsorted samples expressing Cas9 and dCas9 variants; $\log_2$(fold change) values for the *HBB* locus were generated by DESeq2. gRNA positions in the bedGraph files were displayed at default cleavage positions 17–18 for Cas9 and at the center of the guide for dCas9. Analysis of dense perturbation data was also performed with CRISPR–screening uncharacterized region function (CRISPR–SURF)[66]. First, sgRNA counts were used as input for CRISPR–SURF count with the following parameters: -nuclease cas9, -pert indel (for Cas9 samples) or -pert crispri (for dCas9 samples). With the resulting output, CRISPR–SURF deconvolution was applied with the following parameters: -pert cas9, -sim_n 1000, and the perturbation range parameter -range 7 (for Cas9) or -range 20 (for dCas9). CRISPR–SURF deconvolution analysis provides *P* values and statistical power for CRISPR dense perturbation data.

Validation of dCas9 disruption at the γ-globin promoters. Plasmids expressing the eight gRNAs shown in Fig. 5a were individually constructed, followed by lentivirus packaging with the protocols described above except using a small scale. *BCL11A* KO HUDEP-2 cells expressing dCas9 were transduced with each virus and selected using puromycin. After 5 days of erythroid differentiation, cells were collected for gene expression analysis and CUT&RUN. gRNA sequences are listed in Supplementary Table 2.

RT–qPCR. Total RNA was extracted from 1 million freshly collected cells or frozen pellets using a Qiagen RNeasy Mini Kit. Synthesis of complementary DNA was carried out using the iScript cDNA Synthesis Kit (Bio-Rad). The cDNA was diluted five times and 1 μl was used for each qPCR reaction. qPCR was performed with iQ SYBR Green Supermix (Bio-Rad) with the Bio-Rad CFX Real-Time PCR Detection system. HPRT1 was used as endogenous control to normalize between samples. All RT–qPCR primers are listed in Supplementary Table 2. The data were analyzed using the $2^{-\Delta\Delta CT}$ method and visualized with GraphPad. Three technical replicates and at least two biological replicates were performed for each experiment.

shRNA knockdown. The vectors of shRNAs were purchased from Sigma-Aldrich; entrance numbers are listed in Supplementary Table 3. Lentivirus-encoding shRNAs were packaged using HEK293T cells: 100 μl of culture supernatant containing virus particles was added to 1 ml of a suspension of either HUDEP or CD34+ cells. The next day, fresh medium supplemented with puromycin was added to select transduced cells. Knockdown efficiency was validated through RT–qPCR or immunoblot.

CUT&RUN. Briefly, 1 million cells were collected and frozen in 10% DMSO in FBS, as necessary. Frozen cells were thawed, pelleted, washed once with PBS and processed according to the following steps: (1) addition of 0.5 ml of NE buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 0.5 mM spermidine, 0.1% Triton X-100, 20% glycerol and 1× protease inhibitor cocktails (Sigma)) for nuclear isolation. Nuclei were pelleted at 600*g* for 3 min and resuspended with 0.4 ml of NE buffer. (2) 25 μl of BioMagPlus Concanavalin A bead slurry was washed twice, resuspended in 200 μl of binding buffer (20 mM HEPES pH 7.5, 10 mM KCl, 1 mM $CaCl_2$ and 1 mM $MnCl_2$) and added to the nuclear suspension. The mixture was thoroughly mixed on a rocker for 10 min to allow binding of nuclei to the beads. (3) The nuclei were pelleted with a magnet stand and blocked with wash buffer (20 mM HEPES-NaOH pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.1% BSA and 1× protease inhibitor cocktail (Sigma)) supplemented with 2 mM EDTA. After washing once with wash buffer, nuclei were resuspended in 200 μl of wash buffer containing 2 μg of antibody and incubated overnight to allow antibody binding. (4) The next day, nuclei were washed twice on a magnet stand with wash buffer and incubated with 1:1,000 Protein A–micrococcal nuclease fusion protein (pA-MN) in 200 μl of wash buffer. After 1 h the mixture was washed twice to remove unbound pA-MN. (5) Nuclei were resuspended in 150 μl of wash buffer and chilled on a metal block at 0 °C in a water–ice mixture; 3 μl of 100 mM $CaCl_2$ was added to activate pA-MN with incubation at 0 °C for 60 min. (6) The reaction was stopped by the addition of 150 μl of 2× STOP buffer (200 mM NaCl, 20 mM EDTA, 4 mM

EGTA, 50 μg ml^{-1} RNase A and 40 μg ml^{-1} glycogen). The protein–DNA complex was released by centrifugation and then digested by proteinase K at 50 °C overnight. DNA was extracted by ethanol precipitation, followed by Qubit fluorometry and bioanalyzer quality control.

The antibodies used for CUT&RUN were BCL11A (Abcam, no. ab191401), NFYA (Santa Cruz Biotechnology, no. sc-17753) and TBP (Abcam, no. ab220788).

CUT&RUN library preparation. Library preparation of CUT&RUN was performed using a NEB Ultra II library preparation kit, with important modifications. Briefly, <30 ng of DNA was used as input. End preparation was performed at 20 °C for 30 min and then 50 °C for 60 min (rather than 65 °C for 30 min as recommended; the reduced temperature selected prevents melting of short DNA fragments). A 5-pmol adapter was added and ligated to end-preparation products at 20 °C for 15 min. USER enzyme was then utilized to cleave the uracil in the loop, and a 1.7× volume of AMPure beads was used to purify the ligation product; the use of 1.7× beads is essential for recovery of short fragments. To amplify the library, the ligation product was mixed with 2× Ultra II Q5 mix, universal primer and index primers. PCR was carried out as follows: 98 °C for 30 s; 12 cycles of 98 °C for 10 s and 65 °C for 10 s; and a final extension at 65 °C for 5 min. The PCR product was subjected to double-size selection, first with 0.8× and then with 1.2× AMPure beads. The purified library was quantified with Qubit and Tapestation. Libraries were pooled at similar molar amounts and sequenced using the Nextseq500 platform. Paired-end sequencing was performed (read length, 42 bp × 2; index, 6 bp).

The detailed protocol is given at https://doi.org/10.17504/protocols.io.wvgfe3w

CUT&RUN data processing. Raw data were processed using FastQC and CUT&RUNTools[39]. The global settings were: fastq_sequence_length=42, organism_build=hg19, num_bp_from_summit=100, num_peaks=5000, total_peaks=15000, motif_scanning_pval=0.001 and num_motifs=15. Parameters of individual software called by CUT&RUNTools, including Bowtie2 (ref. [67]), MACS2 (ref. [68]), Trimmomatic[69], Picard (http://broadinstitute.github.io/picard/), Samtools[70], MEME[71], Bedtools[72], Bedops[73] and CENTIPEDE[74], were left unchanged.

CUT&RUNTools performs digital footprinting as an integrated step. CENTIPEDE[74] was employed to calculate log-odds, which equals log (P/(1 – P)), where P is the probability of binding based on the strength of footprints reflected by the cut profile. Average footprint profile and log-odds of all motif instances were retrieved in a fimo.result folder. Single-locus cut profiles were retrieved by running a get_cuts_single_locus.sh script, specifying the coordinates of the regions of interest.

The signal-to-noise ratio of CUT&RUN is high. Different noise levels in parallel experiments can have a substantial impact on data normalization. To eliminate such effects, we normalized data based on the fraction of reads in peaks rather than total sequencing depth. First, intersecting peaks between each pair of control and treat expriments were collected using bedtools[72] and the percentage of reads residing in these peaks was caculated. A scaling factor, calculated as percentage_in_peaks_control divided by percentage_in_peak_treat, was used to scale up/down signals in the treatment sample using the bamCoverage tool from deeptools package[75] with the following parameters: scaleFactor $scale – normalizeUsing CPM.

The statistical significance of differential TF binding was quantified with MAnorm[51]; window size used was 500 bp ('-w').

Base editing and single-cell cloning. To base edit the NF-Y motif, the three components of split-intein Target-AID-NG were sequentially expressed in *BCL11A* KO HUDEP-2 cells by lentivirus transduction. Cas9NG-intein-N-expressing cells were purified using FACS, and intein-C-AID cells were purified by blasticidin selection. The expression of each component and the ligated full-length split-intein Target-AID-NG were confirmed by immunoblotting using anti-Flag M2 antibody (no. F1804, Sigma-Aldrich). Each gRNA was introduced to cells through lentivirus transduction.

To base edit the BCL11A motif, we used the codon-optimized base editor FNLS[48]. Base editor and gRNA were introduced to HUDEP-2 cells through lentivirus transduction. After base editing, cells were collected and lysed for genomic DNA extraction. The promoter sequence of γ-globin was amplified and subjected to Sanger sequencing. Sequencing traces were visualized with Snapgene, and the C–T conversion rate quantified with TIDER[76].

To isolate single-cell clones, 30 cells were seeded in a 96-well plate and left unperturbed for 7–10 days. Wells with more than one colony were excluded. Single clones were expanded and genotyped using the method described above. Clones with ≥50% C–T conversion were selected, and those not edited were kept as controls. In addition, clones that had small indels or large deletions in the γ-globin promoters were discarded due to potential indirect effects caused by double-strand DNA breakage and subsequent repair.

The gRNA sequences for base editing and the primers used for amplification of the γ-globin promoters are listed in Supplementary Table 2.

Flow cytometry analysis of single clones. Base-edited clones were stained for intracellular HbF using the same protocol as above, except that this was done in 96-well plates. After staining, cells were analyzed in a BD Accuri C6 flow cytometer and the results were analyzed using FlowJo. For each clone, the percentage of cells with high HbF was quantified.

CRISPR–Cas9-mediated KO in CD34$^+$ or HUDEP-2 cells. Cas9–sgRNA RNP complexes were assembled as follows. For each KO experiment, 500 pmol of FCas9 protein (IDT) and 1,000 pmol of sgRNA (Synthego) were mixed and incubated for 15 min at room temperature. Five million CD34$^+$ cells on day 5 of erythroid differentiation were collected and washed once with PBS. Cells were resuspended with 100 μl of solution P3 for primary cells (Lonza); next, 5 μl of electroporation enhancer (IDT) and preassembled RNP were then added to the cell suspension and nucleofection was performed in a 4D-Nucleofector X unit (Lonza) using program EO-100. Cells were centrifuged and transferred to 10 ml of EDM II for growth under differentiation conditions. For acute depletion of BCL11A, the time of nucleofection was set as day 0; cells were collected at successive times for analyses. To assess genome editing efficiency, cells were lysed to extract genomic DNA. Primers spanning the edited sites were used to amplify the genomic region. The PCR products were subject to Sanger sequencing, and editing efficiency was evaluated by TIDE[77]. The sequences of gRNAs and genotyping primers are listed in Supplementary Table 2. Immunoblotting was used to assess BCL11A levels before and after nucleofection, and band intensities were quantified using ImageJ. Briefly, the pixel densities of target bands and blank regions were measured. The pixel numbers of blank regions represented background and were subtracted from the band. Three independent measurements were performed for each band, and BCL11A signals were normalized using histone H3.

For CRSIPR–Cas9 KO in HUDEP-2 cells, the procedure was the same as that described above except that a smaller volume was used: 50,000 cells were resuspended with 20 μl of solution P3 and combined with the RNP complex assembled with 100 pmol of Cas9 and 150 pmol of sgRNAs. After nucleofection, cells were centrifuged and differentiated for 3 days as described above. Cells were collected after differentiation for either immunoblot or RT–qPCR analysis. The gRNA sequences used are listed in Supplementary Table 2. Antibodies used were C/EBPB (no. NBP1-46179, NovusBio), C/EBPG (no. sc-517003, Santa Cruz Biotechnology), CDP (no. sc-514008, Santa Cruz Biotechnology), NFIA (no. HPA008884, Sigma-Aldrich), NFIC (no. A303-123A-T, Bethyl Laboratories) and histone H3 (no. ab24834, Abcam). All antibodies were used at a concentration of 0.5 μg ml^{-1}.

Statistics. Statistical analysis were performed using GraphPad Prism v.8.0. Details of Student's t-test are indicated in the figure legends. Sample sizes (n) are indicated either in figure legends or by the numbers of individual data points in the figures. For CRISRP–SURF, statistical tests of beta-coefficients were performed empirically through bootstrapping and two-tailed tests. Multiple hypothesis testing was accounted for with the Benjamini–Hochberg procedure.

Additional methods. Methods used for 3C, construction of split Target-AID-NG[46,47,78] and ChIP–seq, qPCR and ChIP–seq data analysis, ATAC-seq and data analysis[79], PRO-seq[80] and PRO-seq data processing can be found in the Supplementary Note.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All raw and processed CRISPR screen, CUT&RUN, ChIP–seq, PRO-seq and ATAC-seq data have been deposited in the NCBI Gene Expression Omnibus under accession number GSE150530. All unprocessed immunoblot gels for Fig. 4a and Extended Data Figs. 2a, 4a,c and 5f can be found in Source data provided with this paper.

Code availability

We made use of publically available software for processing high-throughput sequencing raw data. For single-locus CUT&RUN footprinting, the code can be found at https://bitbucket.org/qzhudfci/cutruntools/src/master/. Code for deconvolution of CRISPR screen data can be found at https://github.com/pinellolab. Custom codes used in this study can be found at https://github.com/yao-qiuming/Nan_NG2020.

References

61. Morgens, D. W. et al. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* **8**, 15178 (2017).
62. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
63. Hess, G. T. et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).
64. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).

65. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
66. Hsu, J. Y. et al. CRISPR-SURF: discovering regulatory elements by deconvolution of CRISPR tiling screen data. *Nat. Methods* **15**, 992–993 (2018).
67. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
68. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
69. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
70. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
71. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
72. Quinlan, A. R. BEDTools: the Swiss-Army tool for genome feature analysis. *Curr. Protoc. Bioinformatics* **2014**, 11.12.1–11.12.34 (2014).
73. Neph, S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
74. Pique-Regi, R. et al. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res.* **21**, 447–455 (2011).
75. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
76. Brinkman, E. K. et al. Easy quantification of template-directed CRISPR/Cas9 editing. *Nucleic Acids Res.* **46**, e58 (2018).
77. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
78. Feng, S. et al. Improved split fluorescent proteins for endogenous protein labeling. *Nat. Commun.* **8**, 370 (2017).
79. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
80. Elrod, N. D. et al. The integrator complex attenuates promoter-proximal transcription at protein-coding genes. *Mol. Cell* **76**, 738–752 (2019).

Acknowledgements

We thank S. Henikoff at the Fred Hutchinson Cancer Research Center for pA-MNase, T. Muir at Princeton University Department of Chemistry for split-intein cDNA and B. Huang at UCSF Department of Pharmaceutical Chemistry for split-mNG2 cDNA. We thank Y. Nakamura at the Cell Engineering Division of RIKEN BioResource Center for HUDEP-1 and HUDEP-2 cell lines. We thank Z. Herbert, M. Berkeley and A. Caruso at the Molecular Biology Core Facilities for high-throughput DNA sequencing, S. Goldman at the Nascent Transcriptomics Core for generation of PRO-seq libraries and performing analyses, and all members at the Hematologic Neoplasia Flow Cytometry core for sorting cells. We also thank M. Cole, M. Canver and C. Smith for critical input and experimental contributions, A. Bowker for technical assistance and members of the Orkin, Bauer and Yuan laboratories for input. D.E.B. was supported by the Burroughs Wellcome Fund and NHLBI (nos. DP2HL137300 and P01HL032262). S.H.O. is an Investigator of the Howard Hughes Medical Institute, supported by both NHLBI (nos. R01HL032259 and P01HL032262) and the Doris Duke Charitable Foundation. G.-C.Y. was supported by NHGRI (no. R01HG009663). L.P. was supported by NHGRI Genomic Innovator Award (no. R35HG010717). N.L. was supported by NIDDK (no. K99DK120925).

Author contributions

N.L. and S.H.O. conceived the study. N.L. designed and performed all experiments, except those detailed below. D.E.B. conceived the Cas9 and dCas9 dense perturbation screens. S.X. performed the screens. Q.Y. analyzed the screen data. J.Y.H. and L.P. performed CRISPR–SURF deconvolution. S.X., Q.Y. and N.L. performed dCas9 disruption validation. Q.Y. and N.L. analyzed PRO-seq data. N.L., Q.Z. and Y.K. analyzed CUT&RUN data under the supervision of G.-C.Y. P.S. assisted with immunoblotting and RT–qPCR. N.L. and S.H.O. wrote the manuscript with input from all authors.

Competing interests

The authors declare no competing interests.

Additional information

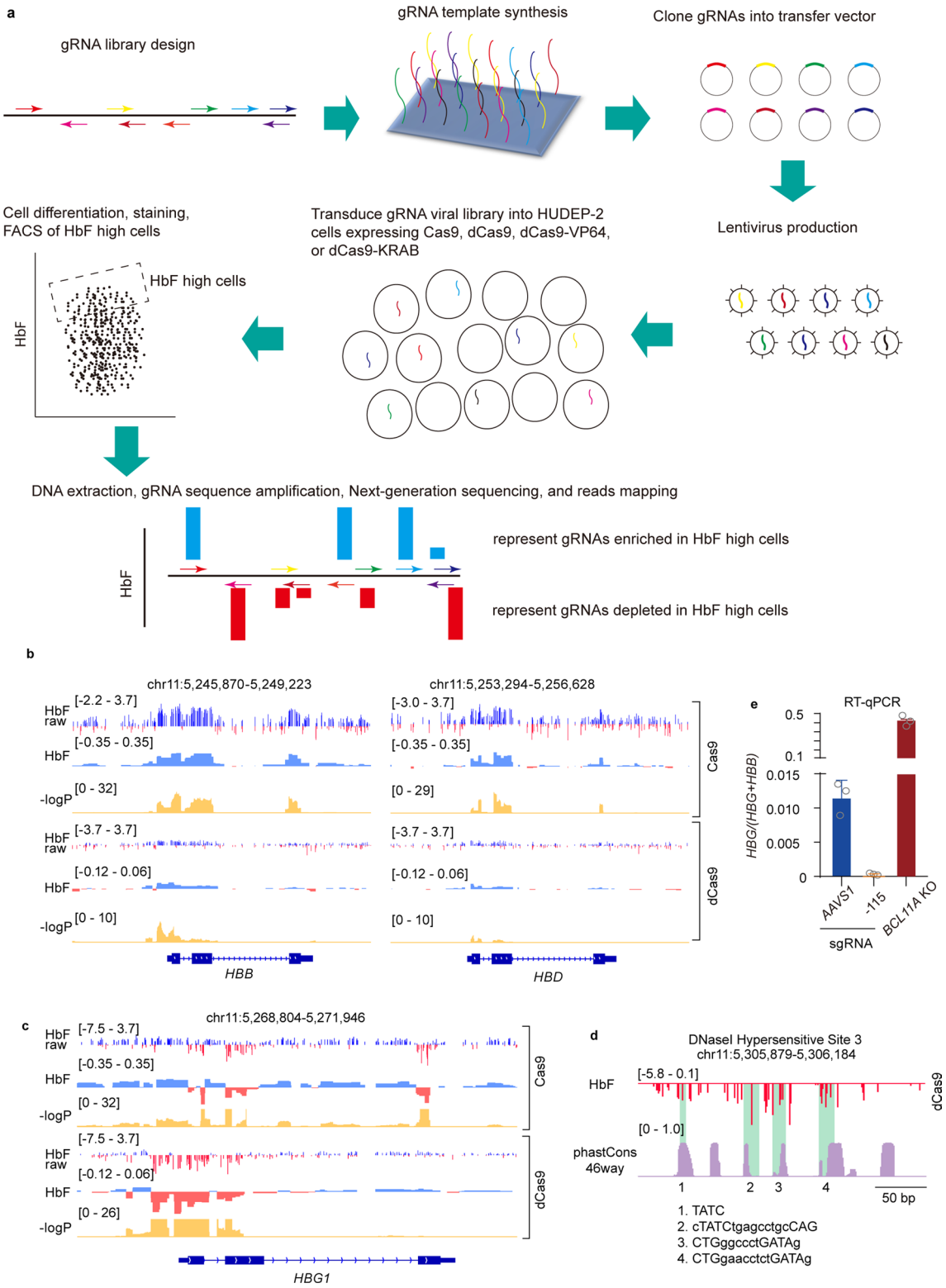
Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00798-y.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00798-y.

Correspondence and requests for materials should be addressed to S.H.O.

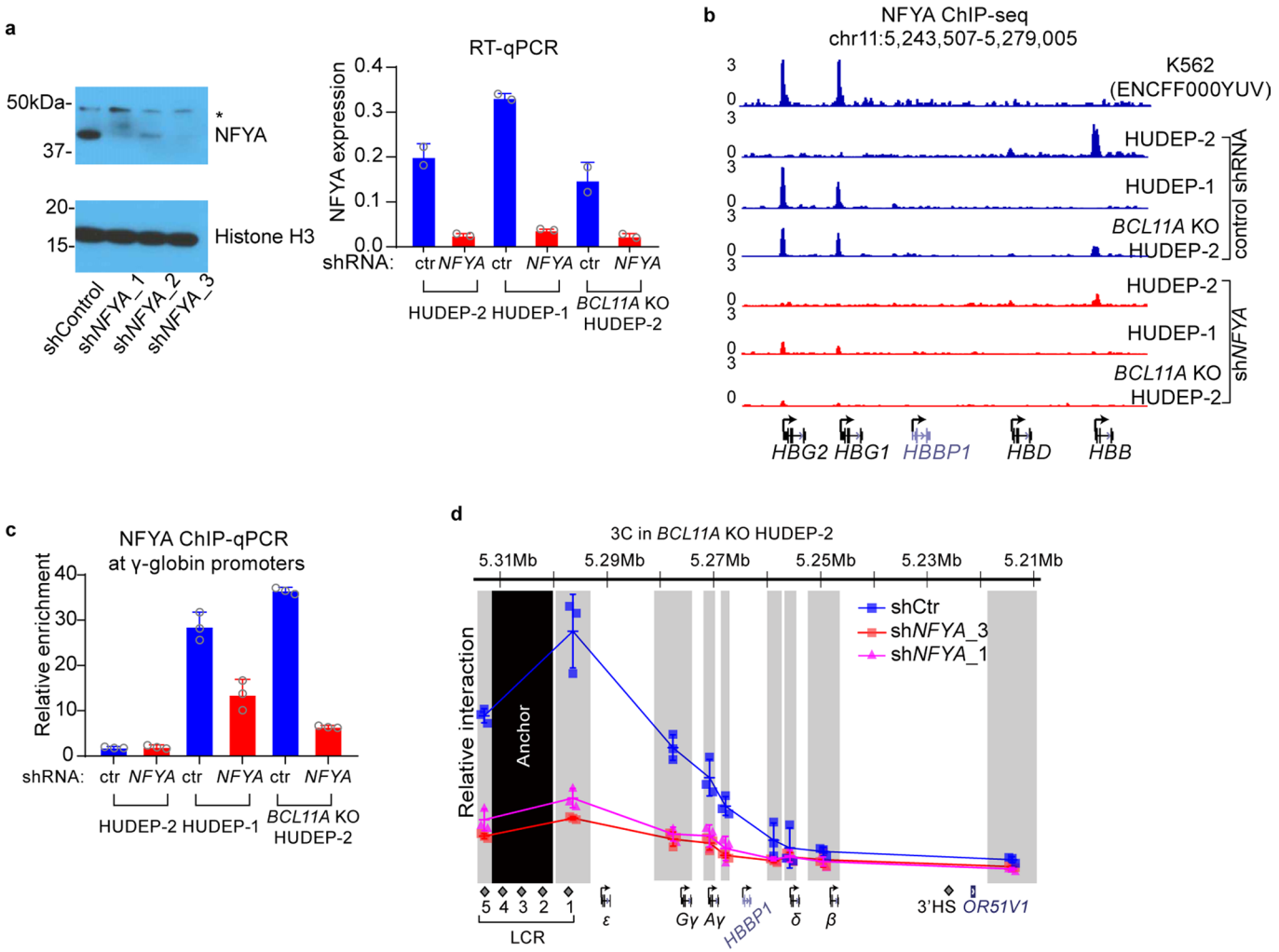
Peer review information *Nature Genetics* thanks Emery Bresnick, Douglas Higgs and Sjaak Philipsen for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

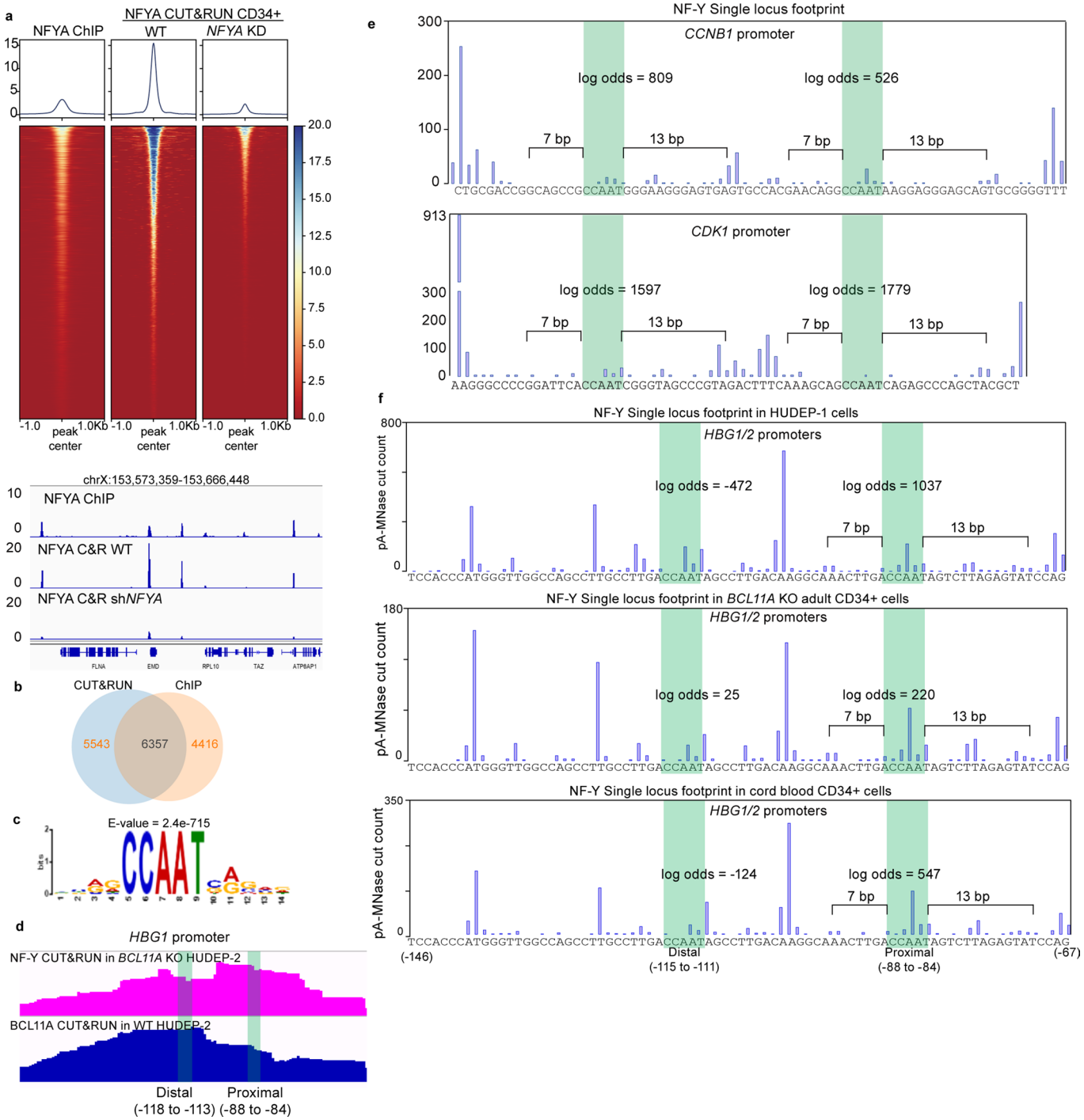


Extended Data Fig. 1 | See next page for caption.

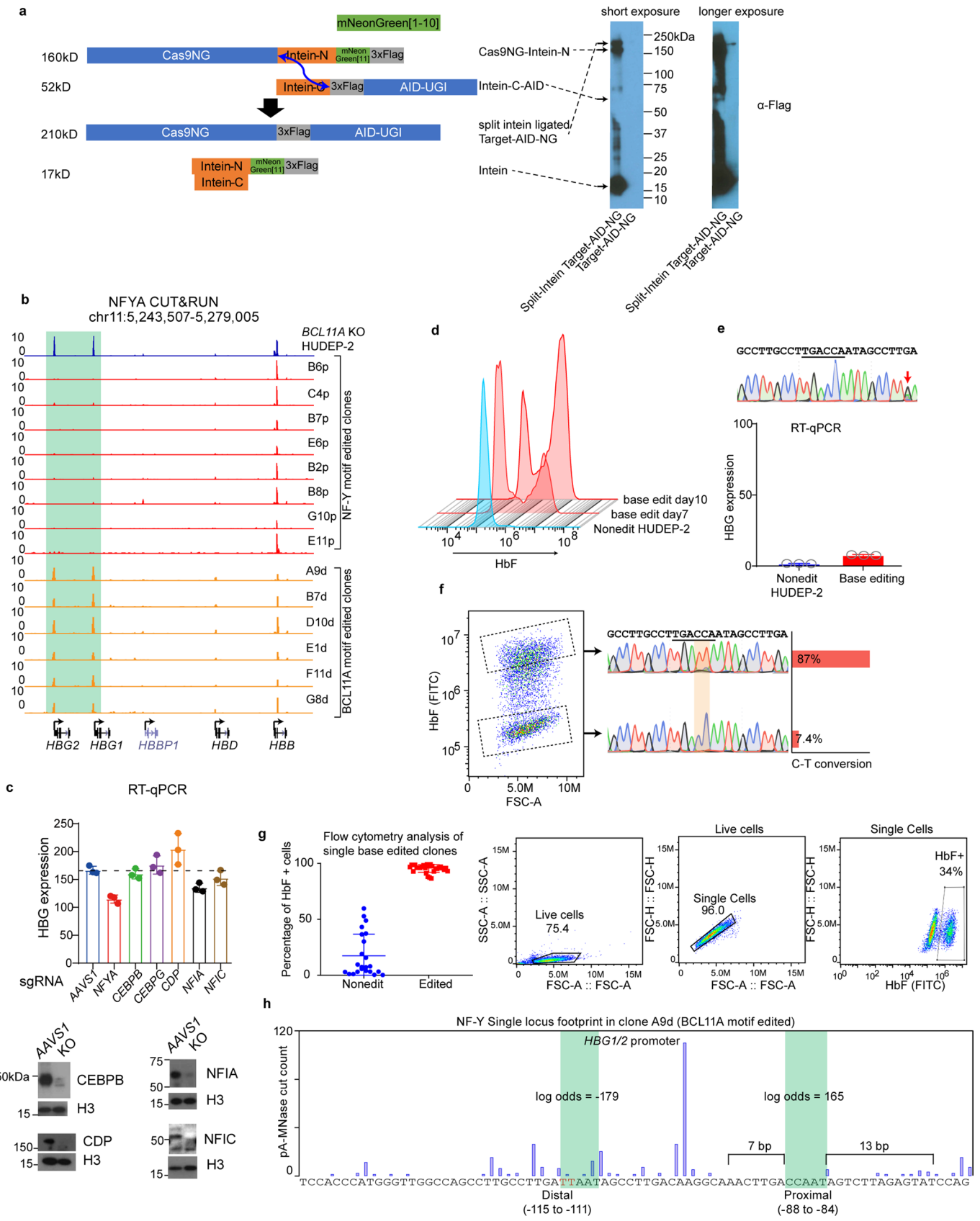
Extended Data Fig. 1 | Dense perturbation of the β-globin locus. a, Flow chart of the dense perturbation experiment design. **b**, Zoomed in view of the dense perturbation results at *HBB* and *HBD* genes. gRNAs that target the exons are enriched in Cas9 experiment. HbF raw score is enrichment of individual gRNAs in HbF-high compared to unsorted population at end of erythroid maturation, plotted as $\log_2$ fold change. HbF score shows deconvoluted underlying genomic regulatory signal with corresponding p-values shown on $-\log_{10}$ scale. **c**, Zoomed in view of the dense perturbation results at *HBG1* gene. Not that gRNAs that target the exons are depleted in Cas9 experiment. **d**, Zoomed in view of the dCas9 dense perturbation result at HS3 of the LCR aligned to PhastCons46way scores. The four regions highlighted in green contain GATA1 or GATA1-TAL1 composite motifs (CTG[N8-9]GATA), with the sequences shown below. **e**, RT-qPCR showing that dCas9/sgRNA binding at -115 of γ-globin promoters reduced γ-globin expression in HUDEP-2 cells. Note that the γ-globin is only expressed at a basal level in cells expressing *AAVS1* control sgRNA. The result is shown as mean (SD) of three technical replicates. Statistical tests of the beta coefficients were performed empirically through bootstrapping and two-tailed tests. Multiple hypothesis testing was accounted for with the Benjamini-Hochberg (BH) procedure.



Extended Data Fig. 2 | NFYA binds to γ-globin promoters and is required for LCR-γ-globin interaction. a, Left, western blot gel showing validation of *NFYA* knockdown efficiency (cropped). All three shRNAs tested showed efficient depletion of NFYA. Right, validation of *NFYA* knockdown efficiency at mRNA level using RT-qPCR. shRNA3 exhibited efficient knockdown of *NFYA* mRNA in all three cells tested and was used thereafter. The result is shown as mean (SD) of two technical replicates. **b**, ChIP-seq tracks of NFYA in HUDEP-2, HUDEP-1, *BCL11A* KO HUDEP-2 cells with or without *NFYA* knockdown. **c**, ChIP-qPCR validation of NFYA binding at the γ-globin promoters in HUDEP-1 and *BCL11A* KO HUDEP-2 cells. No strong binding was detected in HUDEP-2 cells which does not express γ-globin. The result is shown as mean (SD) of three technical replicates. **d**, Chromosome Conformation Capture qPCR in *BCL11A* KO HUDEP-2 cells with or without *NFYA* knockdown. EcoRI fragment encompassing HS2-4 of the LCR was used as anchor point to evaluate LCR-globin interaction. The result is shown as mean (SD) of three technical replicates.

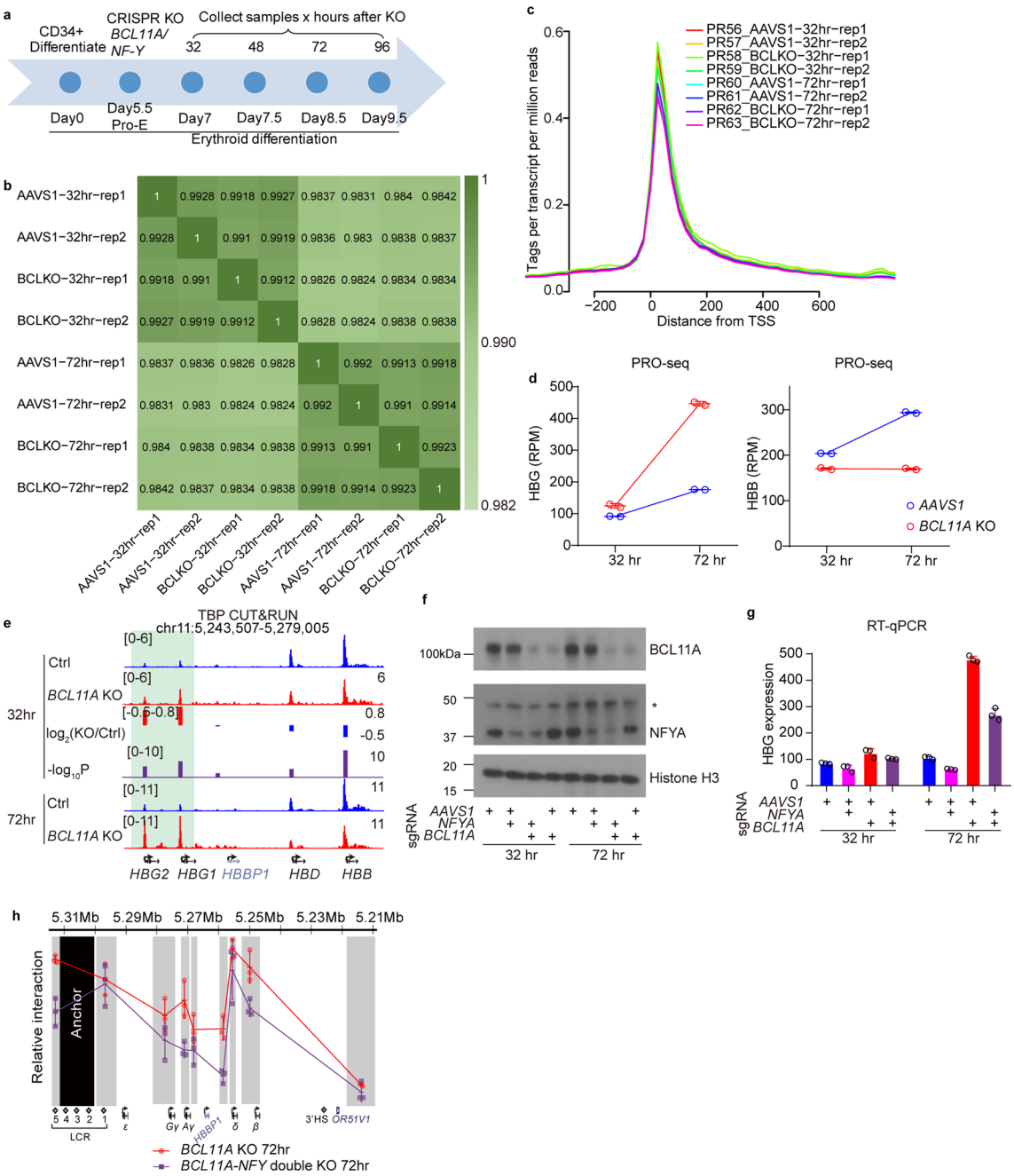


Extended Data Fig. 3 | NF-Y binds to the proximal CCAAT in the γ-globin promoters. a, Upper panel, heatmap comparison of NFYA ChIP-seq in HUDEP-2 cells, NFYA CUT&RUN in primary human CD34⁺ derived erythroid cells with or without *NFYA* knockdown. Lower panel, comparing the signal of the above three experiments at a representative genomic region. **b**, Venn diagram showing the overlap between NFYA CUT&RUN and ChIP-seq peaks. **c**, Motif analysis from 5000 random peaks of NFYA CUT&RUN identifies CCAAT as the highest ranked motif. E-value is reported by MEME. **d**, Zoomed in view of BCL11A CUT&RUN in HUDEP-2 and NFYA CUT&RUN in *BCL11A* KO HUDEP-2 cells at the γ-globin promoters. Distal (−118 to −113) indicates the distal TGACCA motif that BCL11A binds, and proximal (−88 to −84) indicates the proximal CCAAT motif. **e**, Single locus footprint of NF-Y at the *CCNB1* promoter (upper) and *CDK1* promoter (lower). Both CCAAT motifs show strong NF-Y footprints in the two promoters. **f**, Single locus footprint of NF-Y at the γ-globin promoters in HUDEP-1 (upper), *BCL11A* KO adult CD34⁺ derived erythroid cells (middle) and cord blood CD34⁺ derived erythroid cells. Only the proximal motif shows NF-Y footprint.



Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Base editing of the BCL11A and NF-Y motif. a, Left, split-intein mediated ligation of Cas9NG-Intein-N and Intein-C-AID, producing full-length Target-AID-NG. Blue arrow indicates the ligation sites. Right, immunoblot validating the expression of each component and the ligation products. The ligation is incomplete, but the level of ligated product is much higher than the original vector (cropped). **b**, NFYA binding at the γ-promoters diminished in all the NF-Y motif-edited clones (red), and increased in all the BCL11A motif-edited clones (orange), as revealed by NFYA CUT&RUN. NF-Y motif editing was carried out in *BCL11A* KO HUDEP-2 cells while BCL11A motif editing was carried out in wild-type HUDEP-2 cells. **c**, Upper, RT-qPCR analysis of γ-globin expression after acute depletion of C/EBPβ, C/EBPγ, CDP, NFIA and NFIC. Lower, immunoblot validating protein depletion (cropped). *BCL11A* KO HUDEP-2 cells were differentiated for 3 days after nucleofection. The result is shown as mean (SD) of three technical replicates. **d**, Flow cytometry analysis of HbF levels for BCL11A base-edited clones at day 7 and 10. Longer editing resulted in higher base editing rate (Fig. 3e) and higher percentage of HbF positive cells. **e**, A control base editing experiment in which a nucleotide 9 bp away from the BCL11A motif was edited. Sanger sequencing confirmed C-T conversion. **f**, Left, FACS of BCL11A motif base-edited bulk cells into high and low HbF populations. The C-T conversion rate of BCL11A motif in each population was measured by Sanger sequencing and quantified with TIDER. HbF high cells show 87% conversion and HbF low cells show only 7.4% conversion. **g**, Left, flow cytometry analysis of HbF level in individual clones derived from BCL11A motif base editing. Data is showed as mean (SD) of multiple independent clones. Nonedit: n = 23, base edited: n = 30. Right, gating strategy. **h**, Single locus footprint of NF-Y at the γ-promoters in clone A9d, a BCL11A motif-edited clone.



Extended Data Fig. 5 | Acute depletion of BCL11A leads to rapid binding of NF-Y. a, Schematic diagram of primary human CD34⁺ differentiation and acute depletion of BCL11A using CRISPR/Cas9. **b**, Pairwise correlation of PRO-seq experiments. All the experiments in each time point showed high degree of correlation, indicating very minor transcriptional fluctuation upon BCL11A depletion. **c**, Average PRO-seq signal at -200 to +600 bp relative to TSS exhibited promoter pausing of PolII. **d**, Quantification of PRO-seq reads on *HBG1/2* and *HBB* genes after 32 or 72 hrs of BCL11A acute depletion. The y-axis shows Reads Per Million (RPM) for *HBG1*+*HBG2* or *HBB*. The result is shown as mean (SD) of two biologically independent samples (independent cell cultures and CRISPR KO). **e**, CUT&RUN of TBP in CD34⁺ cells undergoing erythroid differentiation after 32 or 72 hrs of BCL11A acute depletion. The result shown is representative of two biological replicates. Quantification of KO/Ctrl and the corresponding p-values are reported by MAnorm. **f**, Western blot for BCL11A and NFYA in adult primary human CD34⁺ derived erythroid cells upon KO of *NFYA*, *BCL11A* or both (cropped). **g**, RT-qPCR analysis of γ-globin expression in adult primary human CD34⁺ derived erythroid cells upon KO of *NFYA*, *BCL11A* or both. Knockout of *NFYA* after 72 hours decreases γ-globin expression. The result is shown as mean (SD) of three technical replicates. **h**, Chromosome Conformation Capture qPCR in adult primary human CD34⁺ derived erythroid cells, comparing *BCL11A* KO and *BCL11A/NF-Y* double KO. EcoRI fragment encompassing HS2-4 of the LCR was used as anchor point to evaluate LCR-globin interaction. The result is shown as mean (SD) of three technical replicates.

nature research

Corresponding author(s): Stuart H. Orkin

Last updated by author(s): Jan 14, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Raw data of high throughput sequencing were collect on Illumina Nextseq or Novaseq platforms with the default factory softwares.
Data analysis	For dCas9 and Cas9 dense perturbation analysis, these softwares were used: DESeq2, R 4.0.1, CRISPR-SURF 1.0. For ChIP-seq, CUT&RUN, ATAC-seq analysis, FastQC 0.11.9, CUT&RUNTools, MAnorm 1.1.4 and deeptools 2.0 were used. CUT&RUNTools integrates the following softwares: Trimmomatic 0.36, Bowtie 2.2.9, Samtools 1.3.1, Picard 2.8.0, MACS 2.1.1, MEME 4.12.0, Bedops 2.4.30, Bedtools 2.26.0. Statistic analysis of qPCR was carried out with GraphPad Prism 8. For PRO-seq analysis, Cutadapt 1.14, Bowtie 1.2.2, Samtools 1.3.1 were used. Custom scripts were used to quantify signals on each gene (https://github.com/yao-qiuming/Nan_NG2020). For quantification of Cas9 editing or base editing efficiency, TIDE 3.2.0 and TIDER 1.0.2 were used. For flow cytometry analysis, FlowJo V10 was used. Western blot was quantified with ImageJ 2.0 Genome tracks were viewed using IGV 2.7.0 Protein Structure was visualized with PyMOL 1.8.6.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All raw and processed CRISPR screen, CUT&RUN, ChIP-seq, PRO-seq and ATAC-seq data have been deposited in the NCBI Gene Expression Omnibus under accession number GSE150530.
All unprocessed western blot gels for figure 4a, Extended Data Figures 2a, 4a, c, 5f can be found in Source Data.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	High throughput experiments including PRO-seq, CUT&RUN were conducted in duplicates as indicated. No additional sample size calculation was performed. Two replicates are commonly used for high throughput sequencing experiments to ensure reproducibility while controlling cost.
Data exclusions	Base edited cells that showed insertion or deletions at the edited sites were considered not qualified and excluded from the downstream analysis in Figures 3c, g. No other data were excluded from analysis.
Replication	Biological replicates were performed and indicated in Figure legends to ensure reproducibility of the key data.
Randomization	Samples were allocated based on genotypes only. There are no covariates and no randomization was required.
Blinding	All the data were generate by machines and do not involve human intervention of values. Therefore no blinding were performed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	BCL11A, ab191401, Abcam; NFYA, sc-17753, Santa Cruz Biotechnology; NFYA, ab139402, Abcam; TBP, ab220788, Abcam; α-Flag M2 antibody, F1804, Sigma-Aldrich;C/EBPB, NBP1-46179, NovusBio; C/EBPG, sc-517003, Santa Cruz Biotechnology; CDP, sc-514008, Santa Cruz Biotechnology; NFIA, HPA008884, Sigma-Aldrich; NFIC, A303-123A-T, Bethyl Laboratories; Histone H3, ab24834, Abcam; HbF, MHFH01, Life Technologies.
Validation	Flag M2 antibody was validated by the manufacturer through immunostaining. HbF antibody was validated by flow cytometry (Extended Data Figure 3g).

All other antibodies were validated by the manufacturers through western blot to confirm specific recognition of the human antigens. All WB images were shown in manufacturers' websites.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HUDEP-2 cells, HUDEP-1 cells were established and shared by Nakamura group at RIKEN, Japan. 293T cells were purchased from ATCC.
Authentication	These cells were not authenticated.
Mycoplasma contamination	Cells were confirmed to be negative for Mycoplasma as determined by PCR.
Commonly misidentified lines (See ICLAC register)	no commonly misidentified lines were used in this study

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150530 token: crmpygswtvipred
Files in database submission	please see below in Methodology-Sequencing depth
Genome browser session (e.g. UCSC)	https://genome.ucsc.edu/s/nanliu/NFY_CUTRUN_for_NG

Methodology

Replicates	two replicates of CUT&RUN were performed for NFYA CUT&RUN or TBP CUT&RUN in CD34+ cells as indicated in the file names by "rep1" and "rep2". One experiment was done for those experiments without labeling "rep".
Sequencing depth	All experiments were sequenced as paired end, with read length of 42 bp each strand. The numbers of total reads and concordantly aligned reads were listed below. (Uniquely mapped reads were not applicable to this project due to the fact that the HBG1 and HBG2 genes are duplicated.) NFYA_ChIP_BCLKOHDP2_ctr,37164569,34771514 NFYA_ChIP_BCLKOHDP2_shNFYA,36723066,34500130 NFYA_ChIP_HDP1_ctr,42005143,39225567 NFYA_ChIP_HDP1_shNFYA,50820724,47620992 NFYA_ChIP_HDP2_ctr,34634556,32575748 NFYA_ChIP_HDP2_shNFYA,49777590,46855899 NFYA_CUTRUN_BCLKOdCas9_102_rep1,33495349,30674512 NFYA_CUTRUN_BCLKOdCas9_102_rep2,15000000,11920433 NFYA_CUTRUN_BCLKOdCas9_115_rep1,36819206,33553876 NFYA_CUTRUN_BCLKOdCas9_115_rep2,15000000,12138125 NFYA_CUTRUN_BCLKOdCas9_124_rep1,38921171,35940594 NFYA_CUTRUN_BCLKOdCas9_124_rep2,15000000,12697117 NFYA_CUTRUN_BCLKOdCas9_139_rep1,33681061,30854506 NFYA_CUTRUN_BCLKOdCas9_139_rep2,15000000,11807286 NFYA_CUTRUN_BCLKOdCas9_158_rep1,39688300,35082998 NFYA_CUTRUN_BCLKOdCas9_158_rep2,15000000,9213025 NFYA_CUTRUN_BCLKOdCas9_197_rep1,41422415,37376825 NFYA_CUTRUN_BCLKOdCas9_197_rep2,15000000,8202314 NFYA_CUTRUN_BCLKOdCas9_208_rep1,46242216,42301794 NFYA_CUTRUN_BCLKOdCas9_208_rep2,15000000,11279120 NFYA_CUTRUN_BCLKOdCas9_62_rep1,32979815,30333424 NFYA_CUTRUN_BCLKOdCas9_62_rep2,15000000,12421347 NFYA_CUTRUN_BCLKOdCas9_AAVS_rep1,30639854,27773935 NFYA_CUTRUN_BCLKOdCas9_AAVS_rep2,15000000,10369371 NFYA_CUTRUN_BCLKOHDP2_rep1,43920791,40823178 NFYA_CUTRUN_BCLKOHDP2_rep2,34270965,29805299 NFYA_CUTRUN_CD34_AAVS32_rep1,19057148,15307722 NFYA_CUTRUN_CD34_AAVS32_rep2,14637450,13197549 NFYA_CUTRUN_CD34_AAVS72_rep2,25988936,22767333 NFYA_CUTRUN_CD34_KO32_rep1,17456600,15834599 NFYA_CUTRUN_CD34_KO32_rep2,20485966,19147012 NFYA_CUTRUN_CD34_KO72_rep2,9283584,8254157 NFYA_CUTRUN_CD34_shBCL,39674277,34823019

NFYA_CUTRUN_CD34_shCtr,33222760,28880730
NFYA_CUTRUN_CD34_shNFYA,47578352,42263166
NFYA_CUTRUN_cloneA9d,23624493,15919597
NFYA_CUTRUN_cloneB2p,18004344,9288219
NFYA_CUTRUN_cloneB6p,14853815,3893985
NFYA_CUTRUN_cloneB7d,16272000,9493221
NFYA_CUTRUN_cloneB7p,29818667,16787151
NFYA_CUTRUN_cloneB8p,24889493,10095503
NFYA_CUTRUN_cloneC4p,15589029,9744071
NFYA_CUTRUN_cloneD10d,19563736,11344188
NFYA_CUTRUN_cloneE11p,10817294,3269153
NFYA_CUTRUN_cloneE1d,27306601,20014302
NFYA_CUTRUN_cloneE6p,18850494,13870224
NFYA_CUTRUN_cloneF11d,19154878,13213573
NFYA_CUTRUN_cloneG10p,10822717,3922287
NFYA_CUTRUN_cloneG8d,15036288,7312922
NFYA_CUTRUN_Cord_CD34,18609228,16381146
NFYA_CUTRUN_HDP1_rep1,28560486,26192514
NFYA_CUTRUN_HDP1_rep2,25445084,19593285
NFYA_CUTRUN_HDP2_rep1,31546932,29406956
NFYA_CUTRUN_HDP2_rep2,32928500,29206382
TBP_CUTRUN_CD34_AAVS32_rep1,20255744,17996530
TBP_CUTRUN_CD34_AAVS32_rep2,16770308,15517167
TBP_CUTRUN_CD34_AAVS72_rep2,28013959,24829072
TBP_CUTRUN_CD34_KO32_rep1,21978816,20072549
TBP_CUTRUN_CD34_KO32_rep2,24492508,22990688
TBP_CUTRUN_CD34_KO72_rep2,20770535,18348659

Antibodies	BCL11A, ab191401, Abcam, for CUT&RUN; NFYA, sc-17753, Santa Cruz Biotechnology, for CUT&RUN; NFYA, ab139402, Abcam, for ChIP-seq; TBP, ab220788, Abcam, for CUT&RUN
Peak calling parameters	Bowtie2 was used to align sequences to hg19, and MACS2 was used to call peaks using -f BAMPE -q 0.01 -B -SPMR. Duplicates were retained for CUT&RUN analysis due to the fact that pA-MN digestion (like Tn5 transposition) frequently results in the same DNA fragments.
Data quality	Raw data was quality checked with FastQC. Low quality reads and unpaired reads were removed by trimmomatic. Additional read through adapters were trimmed with kseq script in CUT&RUNTools.
Software	CUT&RUNTools was used to process ChIP-seq and CUT&RUN data. CUT&RUNTools integrates the following softwares: Trimmomatic 0.36, Bowtie 2.2.9, Samtools 1.3.1, Picard 2.8.0, MACS 2.1.1, MEME, Bedops 2.4.30, Bedtools 2.26.0.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	intracellular staining was performed by fixing cells with 0.05% glutaraldehyde (grade II) (Sigma) for 10 min at room temperature. Cells were centrifuged for 5 min at 600x g and then resuspended in 0.1% Triton X-100 (Life Technologies) for 5 min at room temperature for permeabilization. Triton X-100 was diluted with phosphate buffered saline (PBS) with 0.1% BSA and then centrifuged at 600x g for 15 min. Cells were stained with antibodies for HbF (clone HbF-1 with FITC or APC conjugation; Life Technologies) for 20 min in the dark. Cells were washed to remove unbound antibody before flow cytometry.
Instrument	BD Accuri™ C6 Flow Cytometer
Software	Data were collected using built in software in BD Accuri™ C6 Flow Cytometer and analyzed with FlowJo V10
Cell population abundance	10000-50000 cells were analyzed. Dead cells or doublets were excluded with the gating strategy described below.

Gating strategy

FSC/SSC was used to exclude dead cells. FSC-A/FSC-H were used to select single cells. FL1-A channel recorded the FITC fluorescence. Positive and negative populations were defined according to 1) a non-stained sample as negative control. 2) an HbF expressing cell as positive control. 3) an obvious separation of HbF+ and HbF- populations.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.