Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch

Graphical Abstract

Highlights

- BCL11A recognizes TGACCA motif *in vitro* and *in vivo* through zinc-finger domain

- BCL11A acts at distal TGACCA in human γ-globin promoters to silence expression

- TGACCA motif is mutated in individuals with HPFH syndrome

- Editing the distal TGACCA prevents BCL11A binding and reactivates γ-globin expression

Authors

Nan Liu, Victoria V. Hargreaves,
Qian Zhu, ..., Jian Xu, Martha L. Bulyk,
Stuart H. Orkin

Correspondence

stuart_orkin@dfci.harvard.edu

In Brief

The developmental transition between fetal and adult hemoglobin is controlled by a repressor that acts directly at the γ-globin gene promoter, suggesting a simplified control mechanism that could be manipulated in treatment of γ-hemoglobin disorders.
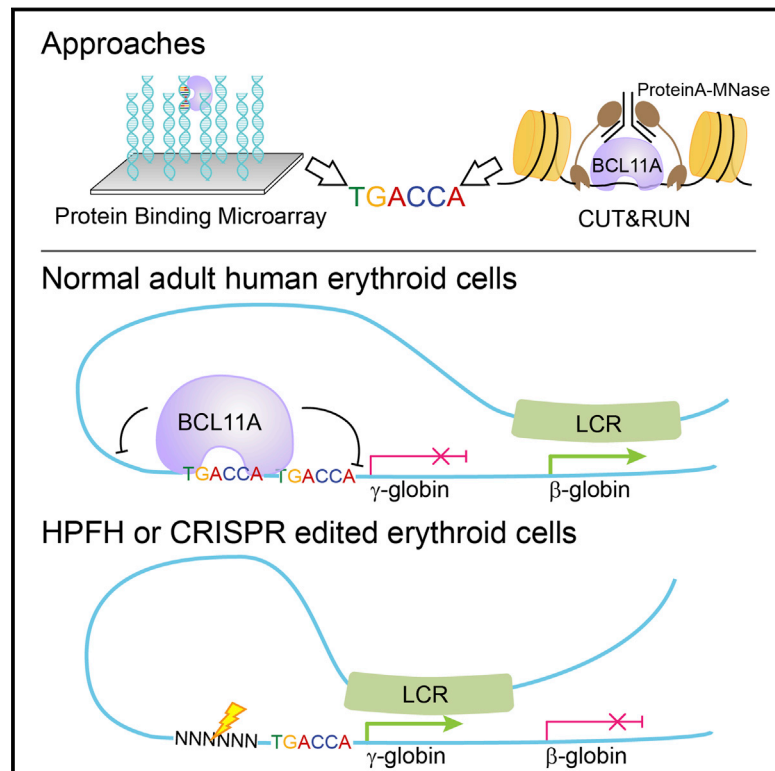
CellPress

Direct Promoter Repression by BCL11A Controls the Fetal to Adult Hemoglobin Switch

Nan Liu,[1,9] Victoria V. Hargreaves,[1,9] Qian Zhu,[2,9] Jesse V. Kurland,[3] Jiyoung Hong,[1] Woojin Kim,[1] Falak Sher,[1] Claudio Macias-Trevino,[1] Julia M. Rogers,[3,4] Ryo Kurita,[5] Yukio Nakamura,[5] Guo-Cheng Yuan,[2] Daniel E. Bauer,[1] Jian Xu,[6] Martha L. Bulyk,[3,4,7] and Stuart H. Orkin[1,8,10,*]

[1]Cancer and Blood Disorders Center, Dana Farber Cancer Institute and Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
[2]Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute and Harvard T.H. Chan School of Public Health, Boston, MA, USA
[3]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[4]Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, MA, USA
[5]Cell Engineering Division, RIKEN Bioresource Center, Tsukuba, Japan
[6]Children's Medical Center Research Institute, Department of Pediatrics, University of Texas at Southwestern Medical Center, Dallas, TX, USA
[7]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[8]Howard Hughes Medical Institute, Boston, MA, USA
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: stuart_orkin@dfci.harvard.edu
https://doi.org/10.1016/j.cell.2018.03.016

SUMMARY

Fetal hemoglobin (HbF, $\alpha_2\gamma_2$) level is genetically controlled and modifies severity of adult hemoglobin (HbA, $\alpha_2\beta_2$) disorders, sickle cell disease, and β-thalassemia. Common genetic variation affects expression of BCL11A, a regulator of HbF silencing. To uncover how BCL11A supports the developmental switch from γ- to β- globin, we use a functional assay and protein binding microarray to establish a requirement for a zinc-finger cluster in BCL11A in repression and identify a preferred DNA recognition sequence. This motif appears in embryonic and fetal-expressed globin promoters and is duplicated in γ-globin promoters. The more distal of the duplicated motifs is mutated in individuals with hereditary persistence of HbF. Using the CUT&RUN approach to map protein binding sites in erythroid cells, we demonstrate BCL11A occupancy preferentially at the distal motif, which can be disrupted by editing the promoter. Our findings reveal that direct γ-globin gene promoter repression by BCL11A underlies hemoglobin switching.

INTRODUCTION

During human development, the major hemoglobin expressed in red blood cells changes from fetal hemoglobin (HbF, $\alpha_2\gamma_2$) to adult hemoglobin (HbA, $\alpha_2\beta_2$). The transcriptional shift of the γ- to the β-globin gene is commonly referred to as the "fetal-to-adult hemoglobin switch." The level of HbF, which is ∼1% of total hemoglobin in adults, is genetically controlled and a crit-

ical modifier of the clinical severity of the major β-hemoglobin disorders, β-thalassemia and sickle cell disease (SCD). In β-thalassemia, where β-globin is deficient, increased γ-globin expression reduces the imbalance of the α- and β-globin chains that underlies the pathophysiology of anemia in this condition. In SCD, increased HbF levels interfere with the polymerization (sickling) of HbS ($\alpha_2\beta^S_2$), thereby reducing damage to red blood cells and extending their lifespan. Common genetic variation modestly affects HbF level, whereas rare alleles with deletions within the β-globin gene cluster or single-base substitutions (or microdeletion) in the γ-globin gene promoters lead to substantial elevations (10%–30% of total hemoglobin) in the benign hereditary persistence of fetal hemoglobin (HPFH) syndrome.

Nuclear regulatory factors controlling HbF level were largely elusive until genome-wide association study (GWAS) provided potential candidates (Menzel et al., 2007; Uda et al., 2008), including BCL11A, which has been validated as a major HbF silencer through surrogate genetics in cells (Sankaran et al., 2008; Xu et al., 2010), mouse knockout (Sankaran et al., 2009), and study of rare haploinsufficient individuals (Basak et al., 2015; Dias et al., 2016). Despite clear genetic evidence for its critical role, how and where BCL11A acts to repress γ-globin gene expression has remained unclear. Literature conflicts with regard to whether BCL11A, or its paralog BCL11B (Wakabayashi et al., 2003), binds DNA directly and, if so, the sequence(s) it may recognize (Avram et al., 2002; Ippolito et al., 2014; Liu et al., 2006; Longabaugh et al., 2017; Marban et al., 2005; Senawong et al., 2005; Tang et al., 2011; Wiles et al., 2013). Prior work suggested that BCL11A acts at a distance from the γ-globin gene promoters but was unable to map precisely sites of occupancy within the β-globin locus (Jawaid et al., 2010; Sankaran et al., 2008, 2011). To understand how BCL11A acts, we began by characterizing its essential domains and DNA binding specificity both *in vitro* and in chromatin.
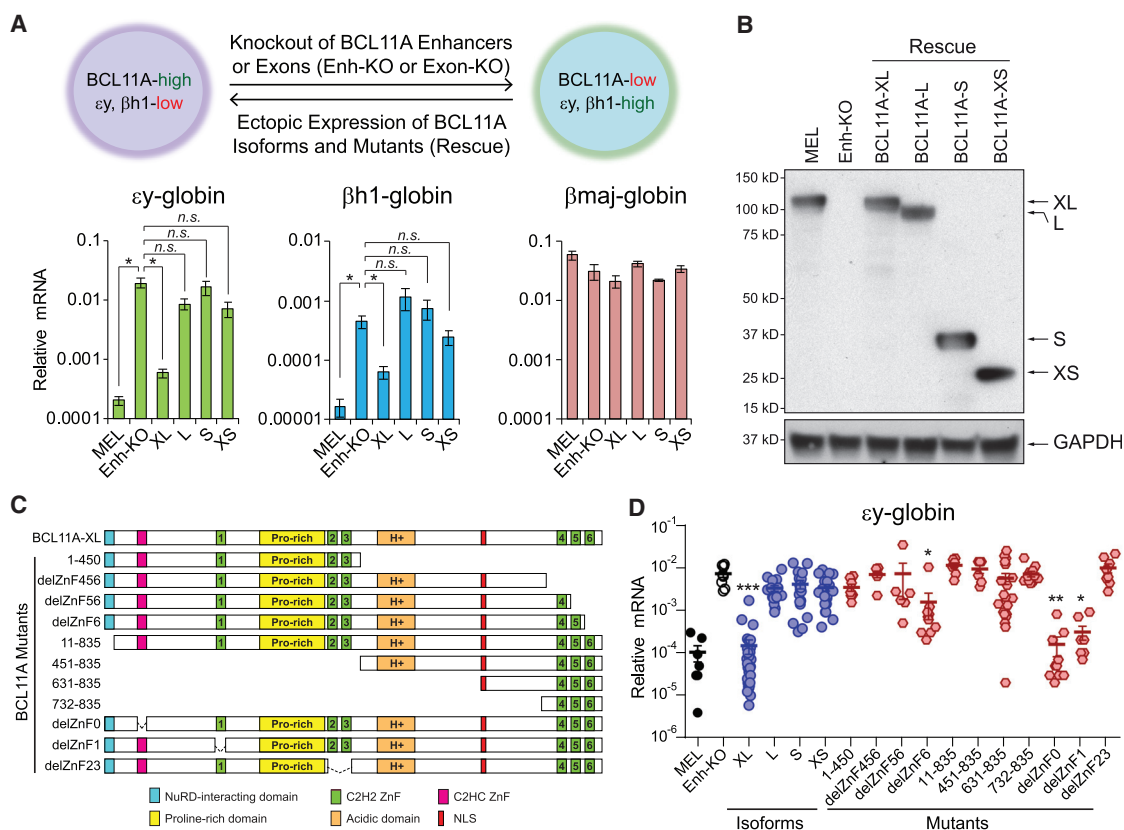
Figure 1. Functional Rescue Analysis Identifies BCL11A Isoform XL and Domains Required for Transcriptional Repression

(A) Schematic of functional rescue analysis in *BCL11A* enhancer or exon knockout mouse erythroleukemia (MEL) cells by ectopic expression of various BCL11A isoforms or domain mutants. Expression of BCL11A-XL, but no other known isoforms, restored the transcriptional repression of the embryonic εγ- and βh1-globin genes, whereas the adult βmajor-globin remained unaffected. Results are mean ± SEM of three experiments and analyzed by two-sided t test. *p < 0.05. *n.s.*, not significant.

(B) Western blot analysis validated the ectopic expression of BCL11A isoforms in Enh-KO cells to levels that are comparable to the endogenous BCL11A expression in MEL cells.

(C) Schematic of the structure-function domains of BCL11A and various mutants. The numbers denote the amino acids retained in each BCL11A mutant relative to the full-length BCL11A-XL isoform.

(D) Repression of εγ-globin expression was restored by full-length BCL11A-XL, but not by domain mutants lacking the NuRD-interacting domain, ZnF23 or ZnF456. Each circle denotes an independent single-cell-derived stable cell clone. Results are mean ± SEM of multiple independent clones and analyzed by one-way ANOVA with repeated-measures. *p < 0.05, **p < 0.01, ***p < 0.001.

See also Figure S1.

RESULTS

Domains of BCL11A Required for Globin Repression

Several isoforms of BCL11A protein are reported (Figure S1A). In adult erythroid cells, two species are expressed: L and a larger, more abundant isoform XL, which differs by inclusion of a terminal exon via RNA splicing (Liu et al., 2006; Sankaran et al., 2008). To determine domains of BCL11A required for globin repression, we developed a functional rescue assay in which cDNA constructs are expressed in murine erythroleukemia (MEL) cells deficient in *BCL11A* either through deletion of the essential erythroid enhancer (Bauer et al., 2013) or TALEN-edited removal of exon 1 or exons 1–5 (Figures 1A–1D, S1B, and S1C). Enhancer- and exon-deleted cells express <5% and no detectable BCL11A RNA (or protein), respectively. *BCL11A*-deficient cells exhibit ≥100-fold derepression of β-like embryonic εγ-globin

gene (Xu et al., 2010) and modest derepression of the weakly expressed βh1-globin gene (Figures 1A, 1D, and S1D–S1H). Forced expression of BCL11A-XL, but not BCL11A-L, restored repression of embryonic globins (Figures 1A, 1D, S1G, and S1H). As BCL11A-XL differs from isoform L only by the presence of a C-terminal zinc-finger cluster (ZnF456), these results implicate this domain as essential for globin repression. Consistent with this conclusion, deletion of ZnF6, ZnF56, or ZnF456 from BCL11A-XL impaired repression (Figures 1C, 1D, and S1).

Expression of additional BCL11A constructs in the rescue assay revealed requirement for the N terminus (BCL11A 11–835) and an internal ZnF cluster, ZnF23 (delZnF23) (Figures 1C and 1D). As the N terminus contains a canonical NuRD-binding peptide (Lejon et al., 2011), NuRD is present in protein complexes with BCL11A (Xu et al., 2013), and downregulation of NuRD subunits induces HbF expression (Gnanapragasam
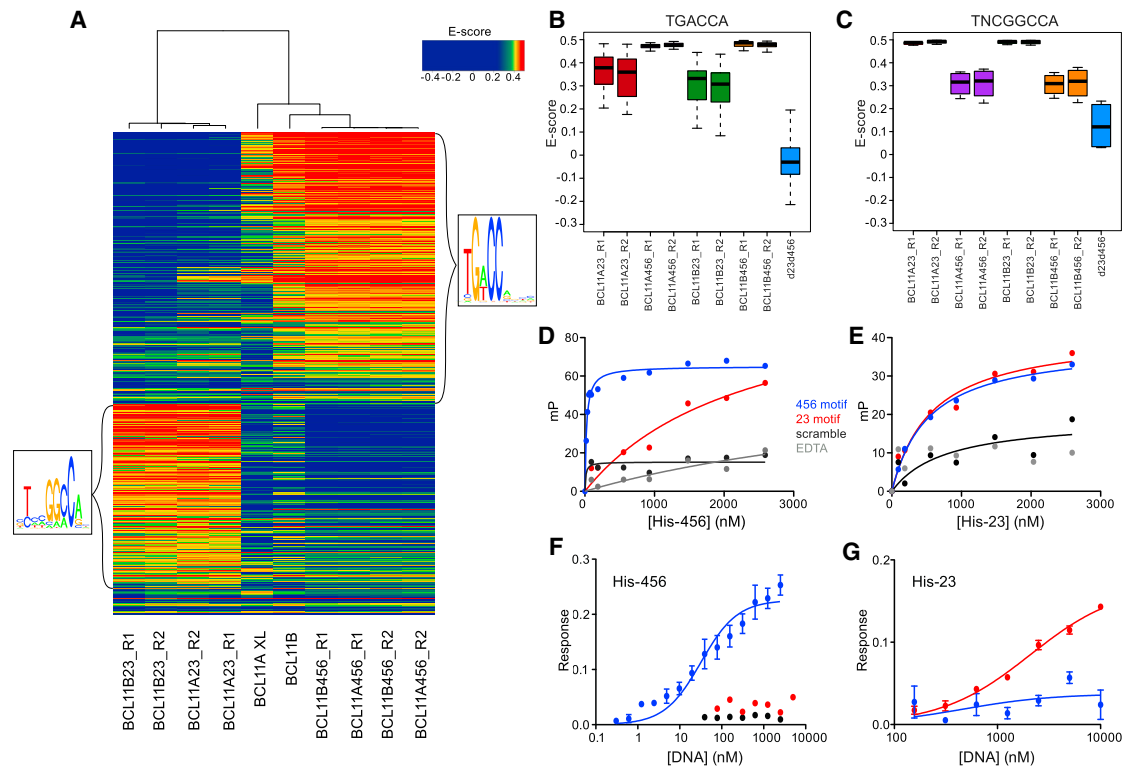
Figure 2. BCL11A Binds DNA in a Sequence-Specific Manner through its C-Terminal ZnF Cluster

(A) Heatmaps depict results of clustering PBM enrichment (denoted as E) scores (rows) and BCL11 proteins (either BCL11A or BCL11B) (columns) for all 8-mers bound (PBM E > 0.40) by at least one protein. Replicates are denoted by R1 and R2. Logos generated from clusters of 8-mers depict binding specificities by the various proteins and match the motifs (information content on y axis, nucleotide position on x axis) derived by the Seed-and-Wobble algorithm.

(B and C) Boxplots depicting PBM E scores of all 8-mers containing TGACCA (B) or TNCGGCCA (C). Each plot illustrates the respective motif preference for the different BCL11 proteins. The middle line of each box represents the median E score value, the upper and lower whiskers portray the min(max(x), Q_3 + 1.5* interquartile range [IQR]) and the max(min(x), Q_1-1.5*IQR), respectively, and points beyond the whiskers represent outliers.

(D and E) Fluorescence polarization curves for ZnF456 (D) and ZnF23 (E) upon binding a 6-FAM-labeled 10-bp double-stranded oligonucleotide containing the 456 motif (blue), the 23 motif (red), or a scrambled control sequence (black) or with the addition of 1 mM EDTA (gray). The curves were fit to a single-site binding model (hyperbola).

(F and G) Octet curves for ZnF456 (F) and ZnF23 (G) upon binding the 456 motif (blue), the 23 motif (red), or a scrambled control sequence (black). The calculated affinity of ZnF456 binding its preferred motif was 31.9 ± 6.8 nM. The calculated affinity of ZnF23 binding to the 23 motif was 2,079 ± 245.8 nM and indeterminate for binding the 456 motif.

The results are shown as mean ± SEM of three experiments.

et al., 2011; Sankaran et al., 2008; Xu et al., 2013), we infer that physical association with NuRD is necessary for globin repression. In contrast, other domains, including ZnF0 and ZnF1, appear dispensable for BCL11A-mediated globin repression (Figures 1C, 1D, and S1).

DNA Recognition by BCL11A

Various DNA sequences have been reported as potential consensus binding motifs for BCL11A (or its paralog BCL11B) (Ai et al., 2017; Avram et al., 2002; Consortium, 2012; Ippolito et al., 2014; Liu et al., 2006; Longabaugh et al., 2017; Marban et al., 2005; Senawong et al., 2005; Tang et al., 2011; Wakabayashi et al., 2003; Wiles et al., 2013). Published studies are internally inconsistent and compromised by the failure to recognize existence of an XL isoform and/or adequately validate chromatin immunoprecipitation sequencing (ChIP-seq) findings. Where no dominant motif has been recovered, some have reported

E26-transformation-specific (ETS)- and Runx-family motifs and deduced their transcription factors (TFs) as possible cofactors (Ippolito et al., 2014; Longabaugh et al., 2017). To interrogate in an unbiased manner sequences recognized by BCL11A, we used the universal protein-binding microarray (PBM) platform that permits high-throughput determination of potential DNA binding sequences (Berger et al., 2006). All possible 8-mer sequences were represented on PBMs with 32-fold coverage. Because ZnF23 and ZnF456 are individually required for globin repression (Figure 1D) and DNA-binding ZnF proteins typically bind DNA through ZnF clusters (Choo and Klug, 1997; Wolfe et al., 2000), we first searched for DNA sequences recognized by BCL11A-XL and the two isolated finger domains, ZnF23 and ZnF456. We performed experiments with the corresponding ZnF clusters of both BCL11A and BCL11B. For ZnF23 and ZnF456, we deduced consensus sequences of (T/C)NNGG(C/A)C(A/G/C) and TG(A/T)CC(A/C/T), respectively (Figure 2A). Sequence selectivity
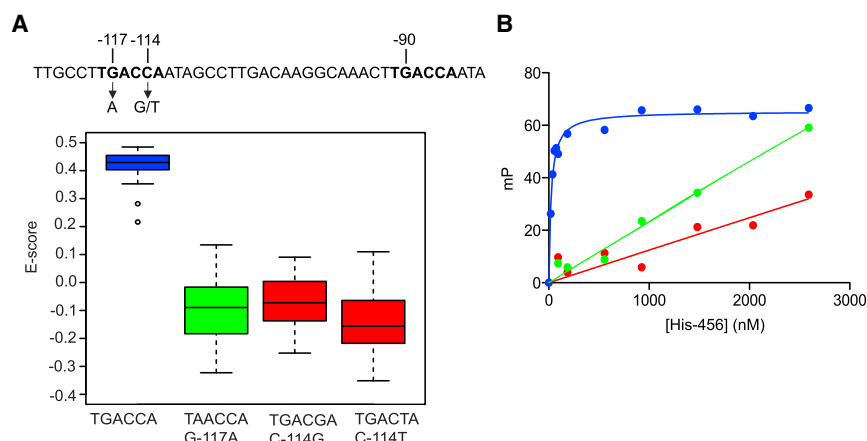
A

B

Figure 3. HPFH Mutations Abrogate BCL11A-DNA Binding

(A) Boxplot representing PBM E scores of all 8-mers containing the depicted 6-mers and highlighting how the HPFH mutations impair BCL11A-XL binding. Boxplot formats are as in Figures 2B and 2C.

(B) Fluorescence polarization curves for ZnF456 upon binding a 6-FAM-labeled 10-bp double-stranded oligonucleotide containing the 456 motif (blue), the 456 motif with the G-117A mutation (green), or the C-114G/T mutation (red). The curves were fit to a single-site binding model (hyperbola). See also Table S1.

was quantified with a PBM enrichment (E) score between −0.5 and +0.5. E scores of 0.45 or greater represent highly selective binding (Badis et al., 2009; Berger et al., 2008; Jiang et al., 2013). The specific sequence that yielded the highest average E score for ZnF456 for both BCL11A and BCL11B was TGACCA (A456, E = 0.472; B456, E = 0.477) (Figure 2A). Of all possible 8-mers, TNCGGCCA was the highest-scoring sequence for ZnF23 from BCL11A or BCL11B (A23, E = 0.489; B23, E = 0.490) (Figure 2A). ZnF23 and ZnF456 each bound their respective sequences selectively (Figures 2B and 2C). A double mutant lacking both ZnF domains (d23d456) failed to bind either motif specifically (Figures 2B and 2C) or to other sequences, thereby excluding DNA binding contribution from elsewhere in the protein. As ZnF456 and BCL11A-XL recognize the same sequence (Figure 2A), whereas ZnF23 recognizes a different sequence, we infer that ZnF456 largely dictates DNA binding of the intact protein. The apparent lack of contribution by ZnF23 to DNA binding of BCL11A-XL may reflect weak affinity for its cognate sequence or inaccessibility in the context of the entire protein. Taken together, these results identify a preferred DNA motif, TGACCA (bound directly by BCL11A-XL *in vitro*), which is mediated through ZnF456.

To measure the affinity, as opposed to selectivity, of binding by ZnF23 and ZnF456 to their respective DNA sequences, we performed two biophysical assays, fluorescence polarization (Do et al., 2008) and biolayer interferometry (BLI)-based octet analysis (Concepcion et al., 2009), with highly purified recombinant proteins. In fluorescence polarization, plane-polarized light excited a fluorescent tag, 6-FAM, on one 5′ end of a 10-bp double-stranded oligonucleotide containing the optimal PBM-derived motif of ZnF23 or ZnF456 at the center. Binding of protein to DNA causes the fluorophore to tumble more slowly, leading to an increase in polarization. Titration of a 6-FAM-labeled ZnF456 motif with a purified ZnF456 protein resulted in increased polarization (Figure 2D), consistent with a calculated K_D of 23.1 ± 2.7 nM, which is within the range typical of ZnF DNA-binding proteins (Pedone et al., 1996). The purified ZnF23 protein bound its motif (Figure 2E) with a much lower affinity, 571.3 ± 149.9 nM, which may account for the apparent lack of contribution of ZnF23 to overall BCL11A-XL binding specificity.

To confirm that the DNA motifs were specific to each ZnF domain, we performed the fluorescence polarization assay with each fluorescently labeled motif and the alternative ZnF protein. The affinity of ZnF456 for the ZnF23 motif was 2349 ± 1293 nM, which is approximately 100-fold weaker than ZnF456 bound to its own motif (Figures 2D and 2E). Surprisingly, ZnF23 bound its cognate motif and the ZnF456 motif with similar affinity (K_D = 526.3 ± 57.05 nM), which is 22-fold weaker than the binding of ZnF456 to its motif. Despite the weak affinity of ZnF23 for its preferred DNA motif, its requirement for globin repression suggests an alternative role, perhaps in mediating protein or RNA interactions. Similar analyses with the alternative method, octet, confirmed the affinity measurements obtained by fluorescence polarization (Figures 2F and 2G).

Mutation of Preferred Embryonic/Fetal Recognition Site in HPFH

Of note, the preferred binding motif for BCL11A, TGACCA, is present in the promoters of all embryonic and fetal-expressed globin genes of humans and mice and none of the adult expressed genes (Table S1). Moreover, the motif is duplicated in each of the human γ-globin genes (−118 to −113 and −91 to −86). The distal motif is altered by single base substitutions at −117 and −114 or by a 13-bp deletion in rare individuals with HPFH (Collins et al., 1985; Gelinas et al., 1985; Gilman et al., 1988). In heterozygous individuals, the HbF level is 10%–30%. PBM analysis revealed that the −117 and −114 HPFH substitutions abrogate binding by BCL11A (Figure 3A). Affinity determinations by fluorescence polarization (Figure 3B) and octet (data not shown) confirmed PBM findings. Thus, these HPFH alleles represent rare sequence variants impaired for BCL11A binding.

Mapping BCL11A Chromatin Occupancy by CUT&RUN

Independent laboratories failed to detect occupancy of BCL11A at the γ-globin promoters by ChIP-PCR or ChIP-chip and instead suggested binding within the locus control region (LCR) and the intergenic Aγ-δ region (Jawaid et al., 2010; Sankaran et al., 2008, 2011). These findings are most consistent with a model in which BCL11A silences HbF primarily through long-distance interactions, perhaps in the context of chromosomal loops (Deng et al., 2014; Xu et al., 2010). Given the presence of a preferred

recognition sequence for BCL11A within the γ-globin promoters and mutation of this motif in HPFH, we revisited *in vivo* BCL11A chromatin occupancy with an alternative approach. We adapted CUT&RUN (Skene and Henikoff, 2017), a recently described method that maps protein binding sites and their long-range genomic contact regions. CUT&RUN is a nuclease-based procedure that eliminates use of protein crosslinking and affords higher resolution than ChIP-seq due to recruitment of micrococcal nuclease to DNA-bound protein with subsequent release of small DNA cleavage fragments.

We performed pilot experiments in immortalized adult-stage human erythroid HUDEP-2 cells (Kurita et al., 2013) and compared findings for localization of CTCF and GATA1 to ChIP-seq datasets (Canver et al., 2017). CUT&RUN identified CTCF and GATA1 binding sites with high sensitivity and resolution at the genome-wide level (Figure S2A). The average width of peaks identified from CUT&RUN CTCF was 159 bp compared to 273 bp from CTCF ChIP-seq. Over 46% and 51% of CTCF and GATA1 ChIP-seq peaks were recovered by the respective CUT&RUN experiments. Upon analysis, we observed that DNA fragments <40 bp were underrepresented. To improve recovery of short fragments, we modified several steps of the procedure and library preparation and optimized the data processing pipeline. The modified protocol increased the recovery of short fragments, enhanced the peak signal over noise (Figure S2B and STAR Methods), and increased the percentage of peak overlap with ChIP-seq by 9%. These optimizations were important for BCL11A CUT&RUN due to its short recognition motif and generation of short fragments from its binding sites.

From six antibodies, we selected one (Figure S2C) for BCL11A CUT&RUN experiments based on enrichment of its preferred binding sequence, overall peak number, and peak overlap between antibodies. BCL11A CUT&RUN was performed in expansion-phase HUDEP-2 cells (Kurita et al., 2013) and primary human stem/progenitor CD34+ cells (Xu et al., 2012) undergoing erythroid differentiation for 3, 5, 7, and 9 days (Figures S2D–S2G). Three or four nuclease cleavage time courses were performed for each cell type and stage. *BCL11A* knockout HUDEP-2 cells (Canver et al., 2015) or immunoglobulin G (IgG) were used as negative controls for HUDEP-2 or CD34+ cell experiments, respectively (Figure S2F). 23000–74000 peaks were identified in these experiments (Figure S2H). The distributions of peaks were highly similar and were most abundant within promoter (35%–39%), intron (30%–34%), and distal intergenic (23%–25%) regions (Figures S2I and S2J).

Unbiased motif discovery of BCL11A CUT&RUN yielded the motif TG(A/G)CC(A/T/G), which was centrally enriched in peaks and contained the BCL11A recognition sequences determined by PBM (Figures 4A, 4B, and S3A). Motifs for GATA1, CTCF, EKLF family proteins, and other TFs were also identified. Among the several sequences conforming to TG(A/G)CC(A/T/G), TGACCA was the most favored sequence (Figures S3B and S3C), which is consistent with PBM E scores for all sequences (data not shown).

As a nuclease-based method, CUT&RUN allows base-resolution digital footprints, which may reflect precise protein binding sites (Neph et al., 2012; Skene and Henikoff, 2017). We enumer-

ated the fragment ends for each base of the genome in BCL11A CUT&RUN and adapted an algorithm that detects and scores footprints *de novo* (Neph et al., 2012). We performed motif discovery within these footprints. TG(A/T)CC(A/C/T) was identified as the top enriched motif (Figure 4C). This motif matches well with the peak-based *de novo* discovery approach above and matches more closely the PBM-derived motif. We next plotted the frequency of nuclease cutting on TGACCA and surrounding sequences and observed that TGACCA was well protected in BCL11A CUT&RUN (Figures 4D, 4E [left], and S4A). In contrast, none of the random hexamers, GATA1 motif, or the mutant HPFH sequences was protected (Figures S4B and S4C). No footprint was evident on analysis of TGACCA in experiments performed with *BCL11A* knockout cells or with IgG (Figures 4D and 4E [right]). We trained an unsupervised Bayes mixture model on the raw counts around −100 to +100 bp of all occurrences of TGACCA to compute the binding probability of each occurrence (STAR Methods). As we ranked TGACCA sites by their computed binding log odds, the rank percentile matched well with the strength of the footprint (Figures 4F and 4G). From this analysis, we found ∼3000 TGACCA sites (with log odds >4.5 or >0.99 probability) that are potential binding sites of the protein. Taken together, these observations support TGACCA as the preferred *in vivo* binding sequence for BCL11A.

Chromatin Localization of BCL11A in Globin Loci

Within the β-globin locus, strong peaks were observed starting from day 5 of CD34+ cell differentiation and in HUDEP-2 cells within the LCR, in the *HBBP1* pseudogene, in the β-globin gene (*HBB*), in 3′HS, and of special interest, in the γ-globin gene promoters (Figures 5A and 5B). The general pattern was similar for different nuclease incubation times starting at 30 min (data not shown). Hypersensitivity sites and *HBB* gained greater accessibility as differentiation progressed in CD34+ cells, reflecting progressive opening of the region upon erythroid differentiation. The TGACCA motif is present at 35 sites in the entire locus. Among 12 that are encompassed by CUT&RUN peaks, the motifs within the γ-globin promoter lie in the peak center and have highest binding probability (Figure 5A [last track] and Table S2). Peaks encompassing HS4, HS3, HS1, *HBD*, and *HBB*, which do not contain a TGACCA motif, may reflect frequent chromatin contacts within the locus rather than direct sites of BCL11A occupancy on DNA (see Discussion). Strong peaks with a centered motif within the γ-globin promoters are highly likely to reflect authentic, direct binding sites of BCL11A.

Strong peaks were also found within the α-globin locus, including the promoter of *HBZ* (Figures S5A and S5B), which encodes the embryonic-expressed α-like ζ-globin—a finding consistent with derepression of ζ-globin on the loss of BCL11A (Sankaran et al., 2008).

High-Resolution Localization of BCL11A at the γ-Globin Promoters

The enhanced resolution of CUT&RUN permits discrimination of BCL11A binding to the duplicated TGACCA motifs in the γ-globin promoters despite their close proximity (21 bp separate the motifs; Figure 3A). Two independent analyses reveal a
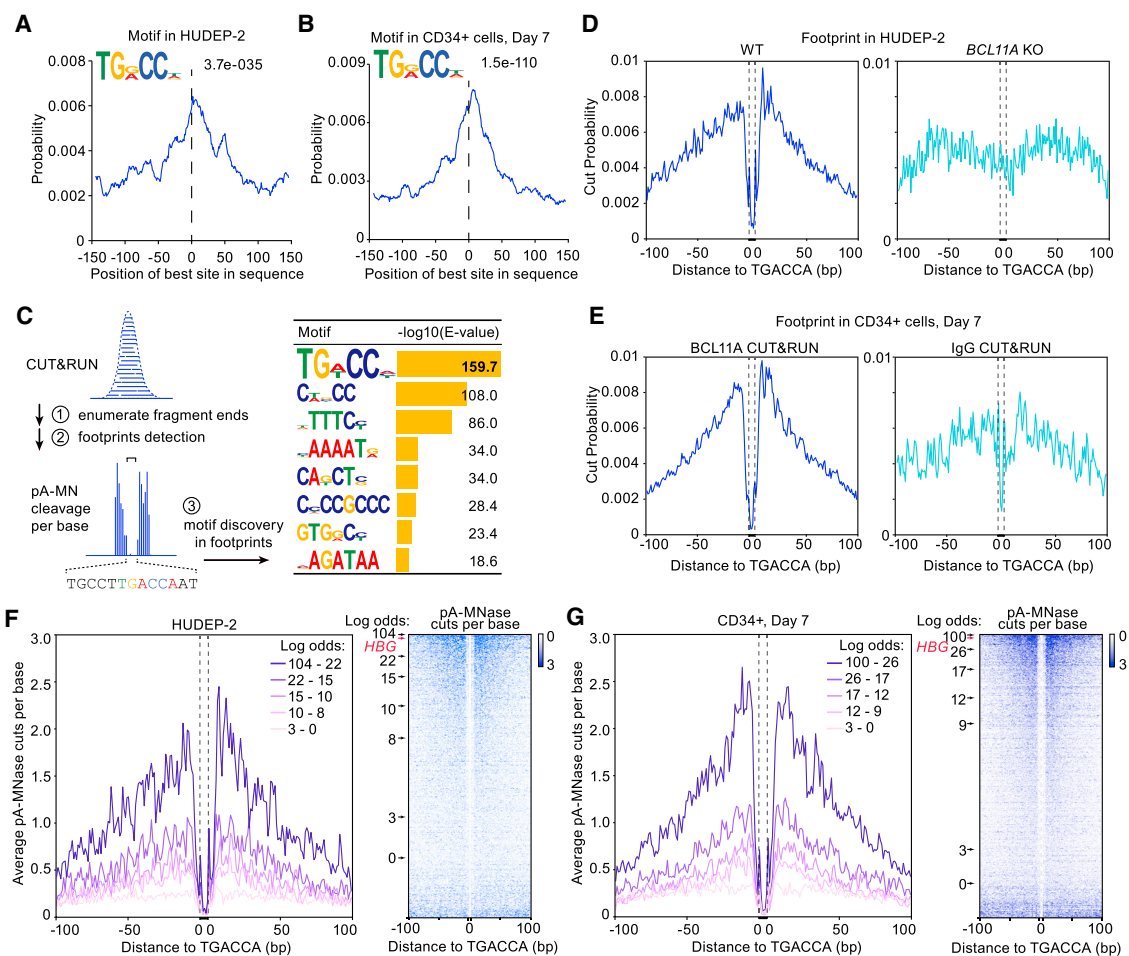
Figure 4. BCL11A CUT&RUN Experiments Reveal an Enriched Binding Motif Consistent with PBM

(A and B) Motif discovery analysis in BCL11A CUT&RUN peaks in HUDEP-2 cells (A) and in CD34+ cells at day 7 of erythroid differentiation (B). The plots showed the position of the motif relative to peak centers. Other motifs discovered are in Figure S3A. E values shown in upper right were reported by MEME.

(C) (Left) Schematic showing the principal of footprint detection. (Right) Motifs discovered within footprints.

(D and E) Targeted motif footprint analysis of BCL11A CUT&RUN. Cut probability of each base surrounding and within TGACCA motifs was plotted. (Left) Footprint analysis on TGACCA in BCL11A CUT&RUN in wild-type HUDEP-2 cells (D) and in CD34+ cells at day 7 of differentiation (E). (Right) control analysis performed in *BCL11A* KO HUDEP-2 cells (D) or IgG CUT&RUN in CD34+ cells (E). Data for other controls and CD34+ cells at days 3, 5, and 9 of differentiation are shown in Figure S4.

(F and G) Footprint analysis for five tiers of BCL11A peaks ranked by binding log odds. Log odds are defined as $\log(P_A/(1-P_A))$, where P_A is posterior binding probability reported by CENTIPEDE. (Left) The average cut count of each base surrounding and within TGACCA was plotted for five tiers of BCL11A peaks. (Right) Heatmap depicts the number of pA-micrococcal nuclease (MNase) cuts per base (column) for each peak (row). The peaks are ordered by decreasing binding log odds. The γ-globin promoter distal motif's binding log odds is denoted by a red arrow.

See also Figures S2, S3, and S4.

binding preference of BCL11A for the distal versus proximal motif. First, we observed that the distal motif (−118 to −113) was more often covered with fragments than the proximal motif (p = 0.0004), and the distal motif was closer to the peak summit position than the proximal motif (p = 0.0001) (Figures 5B–5D). Second, a discriminative TF binding model built for BCL11A revealed that the log odds of binding at the distal motif was 82 in CD34+ cells and 61 in HUDEP-2 cells (Figures 4F and 4G), which ranked in the top 1%–2% of all TGACCA sites— higher than the proximal motif (log odds of 65 in CD34+ cells and 43 in HUDEP-2 cells). In accord with this estimate, we

enumerated the number of fragment ends for each base of the duplicated γ-globin promoters and found that consistently across independent experiments, there were fewer cuts within the distal motif than within the proximal motif (p = 0.0129) (Figures 5E and 5F). Taken together, these findings indicate that BCL11A preferentially binds the distal TGACCA motif in the γ-globin promoters in adult erythroid chromatin. Similarly, the TGACCA motif within the ζ-globin promoter was also protected within a footprint (Figures S5B and S5C), suggesting that the *HBZ* gene promoter harbors an authentic binding site for BCL11A in adult erythroid cells.
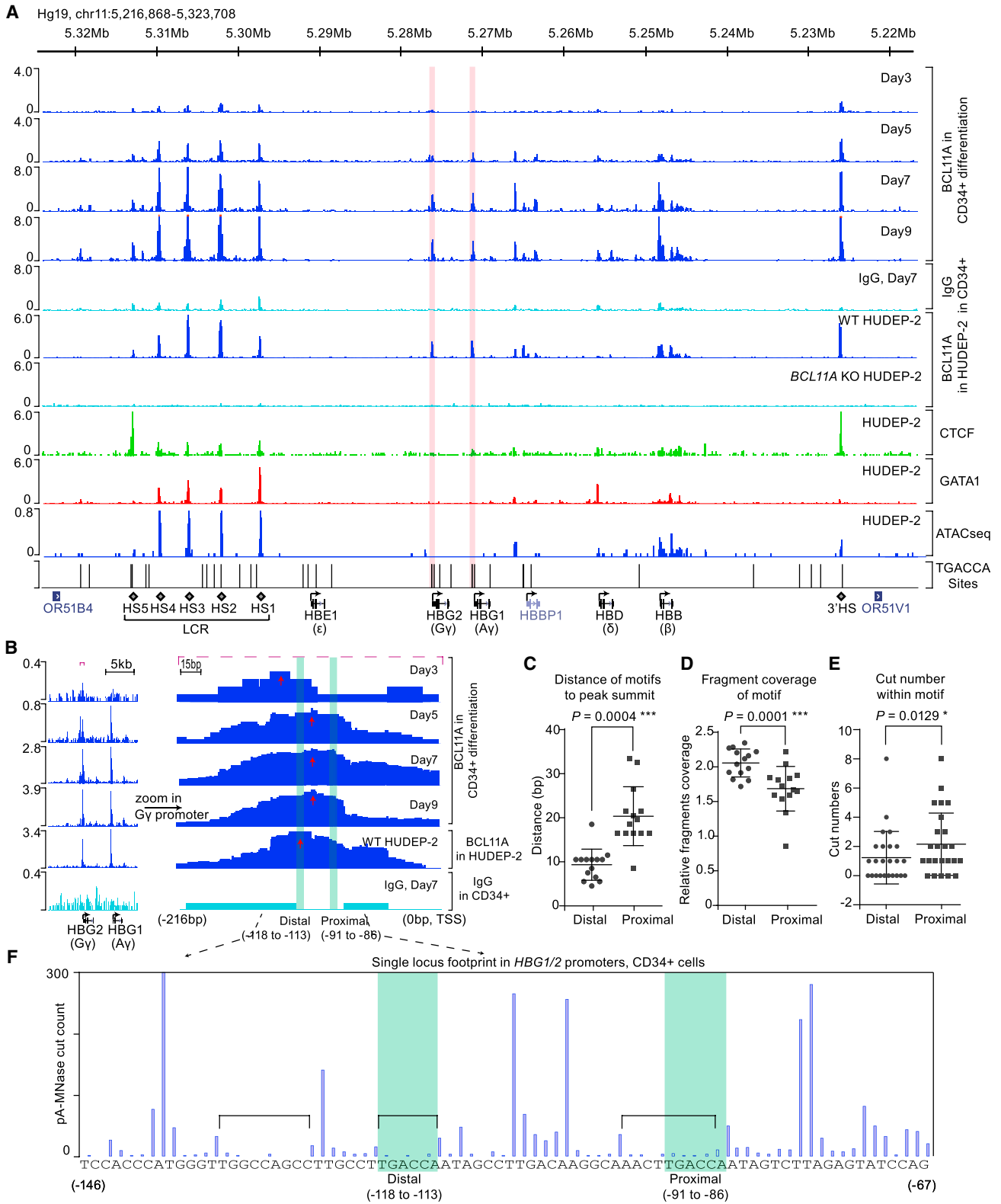
Figure 5. CUT&RUN Reveals BCL11A Binding in γ-Globin Promoters

(A) CUT&RUN profiles in β-globin cluster. Antibodies and cell types for each track are shown on the right. The promoters of duplicated γ-globin genes (*HBG2* and *HBG1*) are highlighted in pink. The bottom track shows all 35 TGACCA sites in non-repetitive regions.

(legend continued on next page)

Gene Editing of Distal BCL11A-Binding Motif in γ-Globin Promoter

To provide definitive evidence that the distal TGACCA motif is bound by BCL11A and also required for repression, we generated a HUDEP-2 cell line (clone D3) in which the distal motifs of all four γ-globin promoters were mutated by CRISPR/Cas9 genome editing (Figure 6A). Fluorescence-activated cell sorting (FACS) and real-time quantitative PCR (qPCR) analysis showed that HbF was expressed in 97% of cells (Figure 6B), and γ-globin comprised 77% of total β-like globin RNA (Figure 6C). Chromosome conformation capture (3C) revealed a switch in enhancer-gene interaction from LCR-β-globin to LCR-γ-globin in clone D3 (Figure S6B) similar to prior observations with a human β-globin locus examined in the context of *BCL11A* knockout transgenic mice (Xu et al., 2010). We performed BCL11A CUT&RUN in clone D3. To facilitate mapping of mutated sequences, we generated a mutant reference genome in which *HBG1* carried a ΔC allele and *HBG2* carried a Δ13-bp allele (Figure S6A). The ΔA mutation can be efficiently mapped to the ΔC allele, since there is only one mismatch.

The CUT&RUN results were initially somewhat surprising, as peaks appeared stronger at the γ-globin promoters in D3 than in wild-type cells (Figure S6D). This seemingly paradoxical observation can be explained, we believe, by consideration of two contributors. First, CUT&RUN is principally a nuclease-based method. Without immunoprecipitation, it maps sites that are directly occupied by the protein, as well as sites that are in proximity to the protein (Skene and Henikoff, 2017). Therefore, major changes in overall chromatin accessibility and long-distance chromosomal interactions may lead to differences in peak sizes and distributions. Second, within the globin locus, transcription of γ- and β-globin genes correlates with DNase sensitivity of their respective promoters. Indeed, as evidenced by assay for transposase-accessible chromatin sequencing (ATAC-seq), the γ-globin promoter in wild-type is virtually inaccessible but is accessible in *BCL11A* knockout and promoter-edited D3 cells (Figure S6E).

Moreover, close inspection of the BCL11A CUT&RUN in D3 cells revealed a valley at the distal motif and a much broader overall pattern (Figure 6D, S6C, and S6D). Footprint analysis showed that the mutated distal motif became hypersensitive to nuclease digestion (Figure 6E). Within position −146 to −67 of the γ-globin promoters, 89 out of 840 cuts occurred in the mutant distal motif (TGCCAA and TGACAA); in contrast, only 2 out of 412 cuts within the distal motif in wild-type cells (Figure 6E). Consistent with this, log odds of binding to distal motif

decreased to an average between alleles of −38 in edited cells as compared to +61 in wild-type cells. Importantly, the gene-edited, mutant motifs were no longer bound by BCL11A, despite overall increased accessibility of the γ-globin promoter. To confirm that these findings are not limited to this specific edited clone of cells, we performed the analysis in a population of cells subjected to CRISPR/Cas9 editing as a pool. Similar results, including derepression of γ-globin and loss of protection at the mutant motifs, were obtained (Figures S6F and S6G). Taken together, these findings indicate that repression of γ-globin gene expression requires direct binding of BCL11A to the distal promoter motif.

DISCUSSION

In addition to its prominent role in HbF regulation, BCL11A and its paralog BCL11B are critical for differentiation in diverse contexts, including B- and T-lymphoid cells (Albu et al., 2007; Liu et al., 2003, 2006) and the developing brain (Dias et al., 2016; Simon et al., 2012; Tang et al., 2011). As numerous difficulties have been encountered in mapping chromatin occupancy of these proteins (Avram et al., 2002; Consortium, 2012; Ippolito et al., 2014; Liu et al., 2006; Longabaugh et al., 2017; Marban et al., 2005; Senawong et al., 2005; Tang et al., 2011; Wiles et al., 2013), prior conclusions regarding their target genes and transcriptional networks merit careful reassessment.

We found that the nuclease-based CUT&RUN procedure (Skene and Henikoff, 2017) provides a powerful, high-resolution method for chromatin localization of a previously intractable protein. Unbiased motif search in BCL11A CUT&RUN experiments yielded the PBM-derived ZnF456-preferred sequence at the center of peaks, indicative of direct protein binding. The utility of CUT&RUN in the context of BCL11A is underscored by our inability to detect the binding motif in ChIP-seq experiments performed with the same antibody (data not shown).

In addition to establishing *in vivo* occupancy of the γ-globin promoters by BCL11A, the high resolution afforded by CUT&RUN revealed preferential binding to the distal motif of the duplicated pair in the promoters. Furthermore, the nature of this method, a targeted nuclease assay, allowed *in vivo* footprinting of BCL11A at base resolution, which provided further evidence of BCL11A occupancy at the distal motif. Although the rare HPFH alleles discovered to date only represent a few of the many possible mutations in the promoter region, it is nonetheless striking that four HPFH mutations target the distal motif and none target the proximal motif. This correlation, supported

---

(B) (Left) BCL11A binding at the *HBG1/2* promoter across multiple CUT&RUN experiments. (Right) Zoomed-in view of 216 bp of Gγ promoter region. The distal and proximal TGACCA motifs are highlighted in green. Arrows indicate positions of peak summits.

(C) Comparison of the distances of distal motif or proximal motif to peak summit.

(D) Comparison of the fragment coverage of distal and proximal motifs. Each fragment originates from one DNA molecule released by protein A-MNase digestion. Relative fragment coverage was defined as average fragment number per base on motifs divided by average fragment number per base within the peak.

(E) Comparison of cut numbers occurred in distal motif or proximal motif.

(F) Single-locus footprint analysis shows the cut frequency at each nucleotide of the γ-globin promoter region (from −146 to −67 in relation to transcription start site [TSS]). Three footprints (regions protected from protein A-MNase cut) are indicated with brackets, and two TGACCA motifs are highlighted in green. 14 CUT&RUN experiments in CD34+ cells are included in (C), (D), and (F); 24 CUT&RUN experiments in HUDEP-2 and CD34+ cells are included in (E); and the p value was calculated by paired t test.
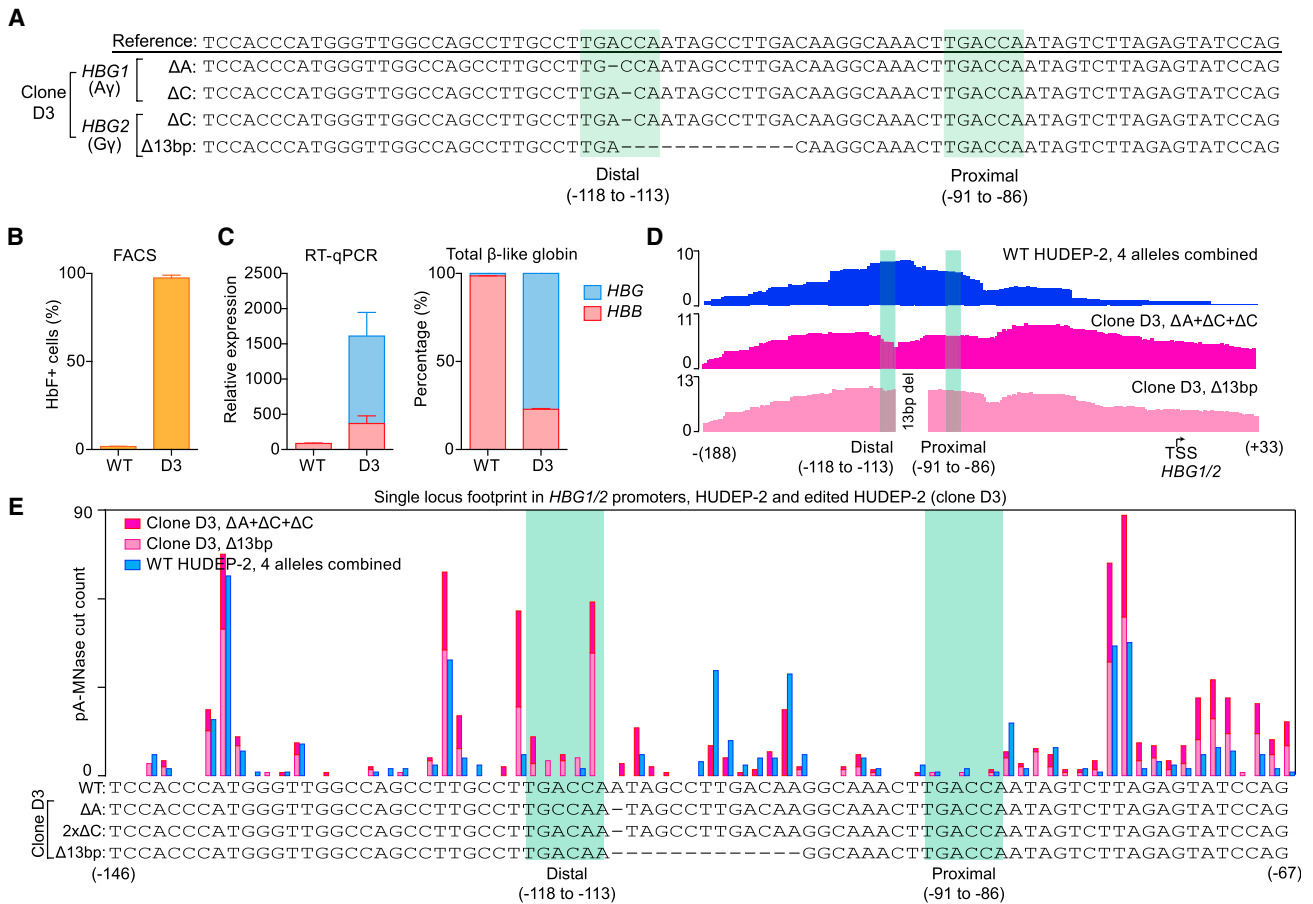
See also Figure S5 and Table S2.

Figure 6. BCL11A Fails to Bind Distal Motif in γ-Globin Promoter Edited HUDEP-2 Cells

(A) Sequence alignment of four alleles of the γ-globin promoter region of HUDEP-2 promoter-edited clone D3 cells.

(B) FACS analysis showing the percentages of HbF-positive cells in wild-type HUDEP-2 and clone D3. Results are shown as mean ± SEM of three experiments.

(C) (Left) Real-time qPCR analysis of mRNA levels for γ-globin and β-globin in indicated cell lines. (Right) Percentages of β- or γ-globin mRNA. Results are shown as mean ± SEM of three experiments.

(D) CUT&RUN tracks showing the γ-globin promoter region. The top track shows combination of 4 γ-globin alleles in wild-type HUDEP-2 cells, the middle track shows combination of 3 alleles (ΔA+ΔC+ΔC) in clone D3, and the bottom track shows allele Δ13 bp. The positions of proximal and distal motifs are highlighted in green. Note that the distal motifs in clone D3 are mutated. Three replicates were merged. Figures S6C and S6D show the replicates separately.

(E) Single-locus footprint showing the γ-globin promoter region in HUDEP-2 and clone D3. Edited cells in clone D3 show a high number of cuts interrupting the distal motif (highlighted green), whereas the wild-type has few to no cuts.

See also Figure S6.

by demonstration of preferential occupancy of the distal motif and loss of binding in promoter-edited cells *in vivo*, provides persuasive evidence that BCL11A acts through this element to direct γ-globin gene repression and hemoglobin switching.

TGACCA motifs overlap with CCAAT motifs in most embryonic and fetal stage globin gene promoters (Table S1). CCAAT motifs of globin genes are primarily bound by transcriptional activators, such as NF-Y, and are required for their transcription (Kim and Sheffery, 1990; Martyn et al., 2017). Within γ-globin promoters, duplication of TGACCAAT sequences offers the potential of multiple modes of regulation. Prior electrophoretic mobility shift assay (EMSA) data suggested that nuclear transcription factor Y (NF-Y) prefers the proximal motif (Superti-Furga et al., 1988), whereas our data demonstrate preferential *in vivo* occupancy

of BCL11A at the distal motif. In fetal cells, in which BCL11A abundance is low, NF-Y may direct active transcription of γ-globin through the proximal motif. With increased BCL11A abundance at the adult stage, BCL11A binding to the distal motif in the γ-globin promoter may transform the local chromatin into a condensed state through NuRD recruitment, thereby preventing NF-Y binding and silencing γ-globin expression.

Besides peaks at the γ-globin promoters, four other regions that contain a TGACCA motif within a footprint are also potential direct BCL11A binding sites (Table S2). These sites include the promoter of the pseudogene lying between the Aγ- and δ-globin genes (*HBBP1*), a subregion of the β-globin LCR (HS2), and both upstream and downstream insulator elements of the β-globin locus (HS5 and 3'HS). Log odds calculations are supportive for

binding to HS5, *HBBP1*, and 3′HS. The presence of potential binding sites outside the γ-globin promoters suggests a broader role for BCL11A in the β-globin locus. Testing the functional impact of these potential sites will necessitate precise editing of these regions and subsequent assessment of effects, if any, on HbF silencing. Of interest, on close inspection, we have observed that one of the peaks in the *HBBP1* region, which lies within the broader Aγ-δ region implicated historically in HbF silencing, was selectively diminished in the γ-promoter-edited cells (Figure S6C). Although the larger significance of this finding remains to be elucidated, it is notable that numerous chromosomal interactions within the β-globin locus map to the *HBBP1* region, as determined by capture-C (Huang et al., 2017) and CAPTURE-3C-seq (Liu et al., 2017). *HBBP1*, however, lies outside the 3.5-kb domain previously proposed as a *cis* element for HbF silencing (Sankaran et al., 2011). These sequences, just upstream of the δ-globin gene, do not contain a detectable BCL11A binding site, which had been suggested previously by ChIP-chip (Sankaran et al., 2011), yet they constitute another region with a high frequency of chromosomal interactions (Liu et al., 2017).

Peaks in CUT&RUN also reflect chromosomal interactions at a distance (Skene and Henikoff, 2017). The presence of BCL11A protein, either bound directly or recruited through interaction with other DNA-bound factors at numerous positions within the β-globin locus, may signify a role in establishing or maintaining proper architecture of the locus. CUT&RUN peaks that are not associated with direct binding of BCL11A to a TGACCA motif encompassed within the peak reflect chromatin-bound BCL11A at other sites brought into proximity through chromosomal interactions. The increased frequency of contacts of the highly accessible γ-globin promoter in promoter-edited cells with other regions of the β-globin locus, we propose, accounts for the seemingly paradoxical increase in CUT&RUN peaks in the promoter, despite lack of direct binding of BCL11A to the mutated distal motif.

The findings reported herein provide compelling evidence that HbF silencing in adult cells is achieved in large part through direct promoter-mediated repression. Thus, prior models proposing that BCL11A acts primarily at a distance from the γ-globin promoters to repress HbF (Sankaran et al., 2011; Xu et al., 2010) are untenable. Hemoglobin switching appears somewhat less complex than previously suspected. As chromosomal looping between the LCR and downstream globin genes is believed necessary for high-level expression (Deng et al., 2014), and loss of BCL11A shifts LCR interactions to the γ-globin gene (Xu et al., 2010), we suspect the presence of BCL11A at the γ-globin promoter may be incompatible with stable loop formation. However, additional roles for BCL11A elsewhere in the β-globin cluster, e.g., in the vicinity of *HBBP1* or the LCR, are by no means excluded.

BCL11A brings together two complementary themes in human genetics. Common genetic variation, as reflected in GWAS, led to recognition of BCL11A as a candidate factor responsible for HbF silencing (Menzel et al., 2007; Uda et al., 2008). Rare genetic variants, reflected by classical HPFH alleles (Collins et al., 1985; Gelinas et al., 1985), prevent BCL11A occupancy at the γ-globin promoters. Whereas common variation affects BCL11A expres-

sion through an erythroid-specific enhancer (Bauer et al., 2013; Canver et al., 2015), rare variation disrupts a pivotal *cis*-regulatory motif in the γ-globin promoter bound by BCL11A. Both genetic elements represent attractive, discrete targets for therapeutic genome editing to reactivate HbF in patients with sickle cell disease or β-thalassemia.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Human Blood and Tissue Samples
  - ○ Immortalized Cell Lines
- METHOD DETAILS
  - ○ Functional Rescue Experiments
  - ○ Western Blot
  - ○ Protein Binding Microarray (PBM) Experiments
  - ○ Protein Purification and Biophysical Assessment of Binding
  - ○ Generation of *HBG* Promoter-edited Cells
  - ○ Screening of −115 Edited HUDEP-2 Clones
  - ○ Validation of HbF Enrichment
  - ○ Chromatin Immunoprecipitation
  - ○ CUT&RUN
  - ○ Library Preparation and Sequencing for CUT&RUN
  - ○ CUT&RUN Data Processing
  - ○ Peak Calling and Overlap with ChIP-seq
  - ○ De Novo Motif Discovery from Peaks
  - ○ Footprint Detection and Motif Discovery within Footprints
  - ○ Single-locus Footprinting, and Global Footprinting Analysis Targeted toward TGACCA Motif
  - ○ Chromosome Conformation Capture Analysis
  - ○ ATAC-seq
  - ○ ATAC-seq Data Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

AUTHOR CONTRIBUTIONS

S.H.O. conceived this study. J.X. designed the functional rescue assay, and V.V.H. and J.X. executed experiments. M.L.B. conceived and supervised the PBM experiments, and V.V.H., J.V.K., and J.M.R. executed experiments and analyzed data. V.V.H, J.H., and W.K. expressed and purified recombinant BCL11A protein and performed fluorescence polarization and octet experiments. R.K. and Y.N. provided the HUDEP-2 cell line. N.L. designed and executed CUT&RUN experiments and conducted ChIP-seq, ATAC-seq, and 3C experiments. Q.Z. and N.L. designed and performed computational analyses and interpreted the results of CUT&RUN. G.-C.Y. and S.H.O provided feedback. F.S., C.M.-T., and D.E.B. generated promoter edited HUDEP-2 cells. S.H.O., V.V.H., N.L., and Q.Z. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

The authors have no competing interests related to this work.

REFERENCES

Ai, S., Peng, Y., Li, C., Gu, F., Yu, X., Yue, Y., Ma, Q., Chen, J., Lin, Z., Zhou, P., et al. (2017). EED orchestration of heart maturation through interaction with HDACs is H3K27me3-independent. eLife 6.

Albu, D.I., Feng, D., Bhattacharya, D., Jenkins, N.A., Copeland, N.G., Liu, P., and Avram, D. (2007). BCL11B is required for positive selection and survival of double-positive thymocytes. J. Exp. Med. 204, 3003–3015.

Avram, D., Fields, A., Senawong, T., Topark-Ngarm, A., and Leid, M. (2002). COUP-TF (chicken ovalbumin upstream promoter transcription factor)-interacting protein 1 (CTIP1) is a sequence-specific DNA binding protein. Biochem. J. 368, 555–563.

Badis, G., Berger, M.F., Philippakis, A.A., Talukder, S., Gehrke, A.R., Jaeger, S.A., Chan, E.T., Metzler, G., Vedenko, A., Chen, X., et al. (2009). Diversity and complexity in DNA recognition by transcription factors. Science 324, 1720–1723.

Bailey, T.L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics 27, 1653–1659.

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. 2, 28–36.

Basak, A., Hancarova, M., Ulirsch, J.C., Balci, T.B., Trkova, M., Pelisek, M., Vlckova, M., Muzikova, K., Cermak, J., Trka, J., et al. (2015). BCL11A deletions result in fetal hemoglobin persistence and neurodevelopmental alterations. J. Clin. Invest. 125, 2363–2368.

Bauer, D.E., Kamran, S.C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M.C., Smith, E.C., Pinello, L., et al. (2013). An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. Science 342, 253–257.

Bembom, O. (2017). Sequence logos for DNA sequence alignments (Division of Biostatistics, University of California, Berkeley).

Berger, M.F., and Bulyk, M.L. (2009). Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nat. Protoc. 4, 393–411.

Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep, P.W., 3rd, and Bulyk, M.L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat. Biotechnol. 24, 1429–1435.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell 133, 1266–1276.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. 109, 21.29 1–9.

Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., et al. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature 527, 192–197.

Canver, M.C., Lessard, S., Pinello, L., Wu, Y., Ilboudo, Y., Stern, E.N., Needleman, A.J., Galactéros, F., Brugnara, C., Kutlar, A., et al. (2017). Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. Nat. Genet. 49, 625–634.

Choo, Y., and Klug, A. (1997). Physical basis of a protein-DNA recognition code. Curr. Opin. Struct. Biol. 7, 117–125.

Collins, F.S., Metherall, J.E., Yamakawa, M., Pan, J., Weissman, S.M., and Forget, B.G. (1985). A point mutation in the A gamma-globin gene promoter in Greek hereditary persistence of fetal haemoglobin. Nature 313, 325–326.

Concepcion, J., Witte, K., Wartchow, C., Choo, S., Yao, D., Persson, H., Wei, J., Li, P., Heidecker, B., Ma, W., et al. (2009). Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization. Comb. Chem. High Throughput Screen. 12, 791–800.

Consortium, E.P.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74.

Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A., and Blobel, G.A. (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell 158, 849–860.

Dias, C., Estruch, S.B., Graham, S.A., McRae, J., Sawiak, S.J., Hurst, J.A., Joss, S.K., Holder, S.E., Morton, J.E., Turner, C., et al.; DDD Study (2016). BCL11A Haploinsufficiency Causes an Intellectual Disability Syndrome and Dysregulates Transcription. Am. J. Hum. Genet. 99, 253–274.

Do, T., Ho, F., Heidecker, B., Witte, K., Chang, L., and Lerner, L. (2008). A rapid method for determining dynamic binding capacity of resins for the purification of proteins. Protein Expr. Purif. 60, 147–150.

Gelinas, R., Endlich, B., Pfeiffer, C., Yagi, M., and Stamatoyannopoulos, G. (1985). G to A substitution in the distal CCAAT box of the A gamma-globin gene in Greek hereditary persistence of fetal haemoglobin. Nature 313, 323–325.

Gilman, J.G., Mishima, N., Wen, X.J., Stoming, T.A., Lobel, J., and Huisman, T.H. (1988). Distal CCAAT box deletion in the A gamma globin gene of two black adolescents with elevated fetal A gamma globin. Nucleic Acids Res. 16, 10635–10642.

Gnanapragasam, M.N., Scarsdale, J.N., Amaya, M.L., Webb, H.D., Desai, M.A., Walavalkar, N.M., Wang, S.Z., Zu Zhu, S., Ginder, G.D., and Williams, D.C., Jr. (2011). p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. Proc. Natl. Acad. Sci. USA 108, 7487–7492.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017–1018.

Huang, J., Liu, X., Li, D., Shao, Z., Cao, H., Zhang, Y., Trompouki, E., Bowman, T.V., Zon, L.I., Yuan, G.C., et al. (2016). Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. Dev. Cell 36, 9–23.

Huang, P., Keller, C.A., Giardine, B., Grevet, J.D., Davies, J.O.J., Hughes, J.R., Kurita, R., Nakamura, Y., Hardison, R.C., and Blobel, G.A. (2017). Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. Genes Dev. 31, 1704–1713.

Ippolito, G.C., Dekker, J.D., Wang, Y.H., Lee, B.K., Shaffer, A.L., 3rd, Lin, J., Wall, J.K., Lee, B.S., Staudt, L.M., Liu, Y.J., et al. (2014). Dendritic cell fate is determined by BCL11A. Proc. Natl. Acad. Sci. USA 111, E998–E1006.

Jameson, D.M., and Seifried, S.E. (1999). Quantification of protein-protein interactions using fluorescence polarization. Methods 19, 222–233.

Jawaid, K., Wahlberg, K., Thein, S.L., and Best, S. (2010). Binding patterns of BCL11A in the globin and GATA1 loci and characterization of the BCL11A fetal hemoglobin locus. Blood Cells Mol. Dis. 45, 140–146.

Jiang, B., Liu, J.S., and Bulyk, M.L. (2013). Bayesian hierarchical model of protein-binding microarray k-mer data reduces noise and identifies transcription factor subclasses and preferred k-mers. Bioinformatics 29, 1390–1398.

Kim, C.G., and Sheffery, M. (1990). Physical characterization of the purified CCAAT transcription factor, alpha-CP1. J. Biol. Chem. 265, 13362–13369.

Kurita, R., Suda, N., Sudo, K., Miharada, K., Hiroyama, T., Miyoshi, H., Tani, K., and Nakamura, Y. (2013). Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. PLoS ONE 8, e59890.

Lejon, S., Thong, S.Y., Murthy, A., AlQarni, S., Murzina, N.V., Blobel, G.A., Laue, E.D., and Mackay, J.P. (2011). Insights into association of the NuRD complex with FOG-1 from the crystal structure of an RbAp48·FOG-1 complex. J. Biol. Chem. 286, 1196–1203.

Liu, P., Keller, J.R., Ortiz, M., Tessarollo, L., Rachel, R.A., Nakamura, T., Jenkins, N.A., and Copeland, N.G. (2003). Bcl11a is essential for normal lymphoid development. Nat. Immunol. 4, 525–532.

Liu, H., Ippolito, G.C., Wall, J.K., Niu, T., Probst, L., Lee, B.S., Pulford, K., Banham, A.H., Stockwin, L., Shaffer, A.L., et al. (2006). Functional studies of BCL11A: characterization of the conserved BCL11A-XL splice variant and its interaction with BCL6 in nuclear paraspeckles of germinal center B cells. Mol. Cancer 5, 18.

Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., Cao, H., Ni, M., Liu, Y., Gu, Z., et al. (2017). In situ capture of chromatin interactions by biotinylated dCas9. Cell 170, 1028–1043.e19.

Longabaugh, W.J.R., Zeng, W., Zhang, J.A., Hosokawa, H., Jansen, C.S., Li, L., Romero-Wolf, M., Liu, P., Kueh, H.Y., Mortazavi, A., and Rothenberg, E.V. (2017). Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. Proc. Natl. Acad. Sci. USA 114, 5800–5807.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics 27, 1696–1697.

Marban, C., Redel, L., Suzanne, S., Van Lint, C., Lecestre, D., Chasserot-Golaz, S., Leid, M., Aunis, D., Schaeffer, E., and Rohr, O. (2005). COUP-TF interacting protein 2 represses the initial phase of HIV-1 gene transcription in human microglial cells. Nucleic Acids Res. 33, 2318–2331.

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

Menzel, S., Garner, C., Gut, I., Matsuda, F., Yamaguchi, M., Heath, S., Foglio, M., Zelenika, D., Boland, A., Rooks, H., et al. (2007). A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15. Nat. Genet. 39, 1197–1199.

Nakagawa, S., Gisselbrecht, S.S., Rogers, J.M., Hartl, D.L., and Bulyk, M.L. (2013). DNA-binding specificity changes in the evolution of forkhead transcription factors. Proc. Natl. Acad. Sci. USA 110, 12349–12354.

Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 489, 83–90.

Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2017). Biostrings: Efficient manipulation of biological strings. Version 2.46.0.

Pedone, P.V., Ghirlando, R., Clore, G.M., Gronenborn, A.M., Felsenfeld, G., and Omichinski, J.G. (1996). The single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is sufficient for high-affinity specific DNA binding. Proc. Natl. Acad. Sci. USA 93, 2822–2826.

Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. 21, 447–455.

Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr. Protoc. Bioinformatics 47, 11.12.1–34.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44 (W1), W160–W165.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods 11, 783–784.

Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K., Hirschhorn, J.N., Cantor, A.B., and Orkin, S.H. (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science 322, 1839–1842.

Sankaran, V.G., Xu, J., Ragoczy, T., Ippolito, G.C., Walkley, C.R., Maika, S.D., Fujiwara, Y., Ito, M., Groudine, M., Bender, M.A., et al. (2009). Developmental and species-divergent globin switching are driven by BCL11A. Nature 460, 1093–1097.

Sankaran, V.G., Xu, J., Byron, R., Greisman, H.A., Fisher, C., Weatherall, D.J., Sabath, D.E., Groudine, M., Orkin, S.H., Premawardhena, A., and Bender, M.A. (2011). A functional element necessary for fetal hemoglobin silencing. N. Engl. J. Med. 365, 807–814.

Schatz, M.C., and Langmead, B. (2013). The DNA Data Deluge: Fast, efficient genome sequencing machines are spewing out more data than geneticists can analyze. IEEE Spectr. 50, 26–33.

Senawong, T., Peterson, V.J., and Leid, M. (2005). BCL11A-dependent recruitment of SIRT1 to a promoter template in mammalian cells results in histone deacetylation and transcriptional repression. Arch. Biochem. Biophys. 434, 316–325.

Simon, R., Brylka, H., Schwegler, H., Venkataramanappa, S., Andratschke, J., Wiegreffe, C., Liu, P., Fuchs, E., Jenkins, N.A., Copeland, N.G., et al. (2012). A dual function of Bcl11b/Ctip2 in hippocampal neurogenesis. EMBO J. 31, 2922–2936.

Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. eLife 6.

Superti-Furga, G., Barberis, A., Schaffner, G., and Busslinger, M. (1988). The -117 mutation in Greek HPFH affects the binding of three nuclear factors to the CCAAT region of the gamma-globin gene. EMBO J. 7, 3099–3107.

Tang, B., Di Lena, P., Schaffer, L., Head, S.R., Baldi, P., and Thomas, E.A. (2011). Genome-wide identification of Bcl11b gene targets reveals role in brain-derived neurotrophic factor signaling. PLoS ONE 6, e23691.

Uda, M., Galanello, R., Sanna, S., Lettre, G., Sankaran, V.G., Chen, W., Usala, G., Busonero, F., Maschio, A., Albai, G., et al. (2008). Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia. Proc. Natl. Acad. Sci. USA 105, 1620–1625.

Wakabayashi, Y., Watanabe, H., Inoue, J., Takeda, N., Sakata, J., Mishima, Y., Hitomi, J., Yamamoto, T., Utsuyama, M., Niwa, O., et al. (2003). Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. Nat. Immunol. 4, 533–539.

Warnes, G.R., Bolker, B., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., et al. (2016). gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1.

Wiles, E.T., Lui-Sargent, B., Bell, R., and Lessnick, S.L. (2013). BCL11B is up-regulated by EWS/FLI and contributes to the transformed phenotype in Ewing sarcoma. PLoS ONE 8, e59369.

Wolfe, S.A., Nekludova, L., and Pabo, C.O. (2000). DNA recognition by Cys2His2 zinc finger proteins. Annu. Rev. Biophys. Biomol. Struct. *29*, 183–212.

Xu, J., Sankaran, V.G., Ni, M., Menne, T.F., Puram, R.V., Kim, W., and Orkin, S.H. (2010). Transcriptional silencing of gamma-globin by BCL11A involves long-range interactions and cooperation with SOX6. Genes Dev. *24*, 783–798.

Xu, J., Shao, Z., Glass, K., Bauer, D.E., Pinello, L., Van Handel, B., Hou, S., Sta-matoyannopoulos, J.A., Mikkola, H.K., Yuan, G.C., and Orkin, S.H. (2012).

Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. Dev. Cell *23*, 796–811.

Xu, J., Bauer, D.E., Kerenyi, M.A., Vo, T.D., Hou, S., Hsu, Y.J., Yao, H., Trow-bridge, J.J., Mandel, G., and Orkin, S.H. (2013). Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A. Proc. Natl. Acad. Sci. USA *110*, 6518–6523.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-BCL11A [EPR14943-44]	Abcam	Cat#ab191401;
Mouse monoclonal anti-BCL11A [15E3AC11]	Abcam	Cat#ab18688; RRID:AB_444633
Mouse monoclonal anti-BCL11A [14B5]	Abcam	Cat#ab19487; RRID:AB_444947
Mouse monoclonal anti-Histone H3	Abcam	Cat#ab24834; RRID:AB_470335
Rabbit polyclonal anti-GATA1	Abcam	Cat#ab11852; RRID:AB_298635
Normal rabbit IgG	Millipore	Cat#12-370; RRID:AB_145841
Rabbit polyclonal anti-CTCF	Millipore	Cat#07-729; RRID:AB_441965
FITC-conjugated anti-Human Fetal Hemoglobin antibody	Life Technologies	Cat#MHFH01; RRID:AB_1500107
Mouse monoclonal ANTI-FLAG M2	Sigma-Aldrich	Cat#F1804; RRID:AB_262044
Rabbit polyclonal anti-GAPDH	Santa Cruz	Cat#sc-25778; RRID:AB_10167668
Rabbit polyclonal HA-probe antibody (HRP)	Santa Cruz	Cat#sc-805 HRP
Mouse monoclonal anti-HA (Alexa Fluor 488)	Life Technologies	Cat#A-21287; RRID:AB_2535829
Rabbit polyclonal anti -GST (Alexa Fluor 488)	Life Technologies	Cat#A-11131; RRID:AB_2534137
Bacterial and Virus Strains		
Rosetta(DE3)pLysS	Millipore	Cat#70956-4
Biological Samples		
Human peripheral blood stem cell, CD34 enriched	Fred Hutchinson Cancer Research Center	Lot:R002895
Chemicals, Peptides, and Recombinant Proteins		
T4 DNA ligase	Thermo Fisher Scientific	Cat#EL0011
Protein A-Micrococcal nuclease	Skene and Henikoff, 2017	N/A
EcoRI-HF	NEB	Cat#R3101
Insulin solution human	Sigma-Aldrich	Cat#I9278-5ML
Heparin	Sigma-Aldrich	Cat#H3149-500KU
holo-Transferrin human	Sigma-Aldrich	Cat#T0665-1G
Hydrocortisone	Sigma-Aldrich	Cat#H0888-1G
Erythropoietin	Amgen	NDC 55513-144-10
Solvent Detergent Pooled Plasma AB	Rhode Island Blood Center	Cat#X0004
L-Glutamine	Thermo Fisher Scientific	Cat#25030-164
Fetal Bovine Serum Source: Australia Heat Inactivated	Omega Scientific	Cat#FB-22
Human IL-3	R&D	Cat#203-IL
Human SCF	R&D	Cat#255-SC
Doxycycline	Sigma-Aldrich	Cat#D9891-5G
Dexamethasone	Sigma-Aldrich	Cat#D4902-100MG
Puromycin	Sigma-Aldrich	Cat#p8833-25mg
Blasticidin	Invivogen	Cat#ant-bl-1
Dynabeads protein G	Thermo Fisher Scientific	Cat#10004D
T4 DNA Polymerase	Millipore	Cat#70099
cOmplete ULTRA Tablets, EDTA-free, glass vials Protease Inhibitor Cocktail	Sigma-Aldrich	Cat# 5892953001
BioMag Plus Concanavalin A	Polysciences	Cat# 86057-3
Critical Commercial Assays		
NEBNext Ultra II DNA Library Prep Kit	NEB	Cat#E7645S
NextSeq 500/550 High Output Kit v2 (75 cycles)	Illumina	Cat#FC-404-2005

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Nextera DNA Library Preparation Kit	Illumina	Cat#FC-121-1030
iQ SYBR Green Supermix	Bio-rad	Cat#1708880
RNeasy Plus Mini Kit	QIAGEN	Cat#74136
Gateway Technology	Invitrogen	Cat#12535-019
PURExpress *In Vitro* Protein Synthesis Kit	New England Biolabs	Cat#E6800S
1-Step Human Coupled IVT kit	Thermo Fisher Scientific	Cat#88881
Deposited Data		
Raw and processed sequencing data	This paper	GEO: GSE104676
PBM data	This paper	UniPROBE (http://the_brain.bwh.harvard.edu/uniprobe/): LIU18A
Experimental Models: Cell Lines		
Human: HUDEP-2 cells	Kurita et al., 2013	N/A
Mouse: MEL cells	Coriell Institute for Medical Research	Cat# GM00086; RRID:CVCL_3988
Human: HEK293T cells	ATCC	Cat#CRL-3216; RRID:CVCL_0063
Oligonucleotides		
Primers for RT-qPCR, see Table S3	This paper	N/A
Primers for genotyping CRISPR edited cells, see Table S3	This paper	N/A
Primers for sgRNAs, see Table S3	This paper	N/A
Primers for 3C-qPCR, see Table S3	Xu et al., 2010	N/A
dsDNA for FP, see Table S3	This paper	N/A
dsDNA for Octet, see Table S3	This paper	N/A
Recombinant DNA		
pLentiGuide	Sanjana et al., 2014	Addgene #52963
lentiCas9-Blast	Sanjana et al., 2014	Addgene #52962
pLVX-EF1a-IRES-zsGreen1	Clontech	Cat#631982
pDEST15	Invitrogen	Cat#11802014
pT7CFE	Thermo Fisher Scientific	Cat#88861
pMCSG7	Midwest Center for Structural Genomics	N/A
Software and Algorithms		
MACS2	Zhang et al., 2008	https://github.com/taoliu/MACS
Bowtie2	Schatz and Langmead, 2013	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Trimmomatic	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic; RRID:SCR_011848
Picard	Broad Institute	http://broadinstitute.github.io/picard
SAMtools	Badis et al., 2009	http://samtools.sourceforge.net/; RRID:SCR_006525
CENTIPEDE	Pique-Regi et al., 2011	http://centipede.uchicago.edu/
MEME	Machanick and Bailey, 2011	http://meme-suite.org/; RRID:SCR_001783
plotFingerprint	Ramírez et al., 2016	https://deeptools.readthedocs.io/en/latest/#
make_cut_matrix.py	Parker lab	https://github.com/Parkerlab/atactk
BEDtools	Quinlan, 2014	http://bedtools.readthedocs.io/en/latest/; RRID:SCR_006646
fp2012	Neph et al., 2012	https://github.com/sjneph/footprinting2012
Universal PBM Analysis Suite	Berger and Bulyk, 2009	http://the_brain.bwh.harvard.edu/PBMAnalysisSuite/index.html

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
R: gplots	Warnes et al., 2016	https://cran.r-project.org/web/packages/gplots/index.html
seqLogo	Bembom, 2017	https://bioconductor.org/packages/release/bioc/html/seqLogo.html
makePWM	Pagès et al., 2017	http://bioconductor.org/packages/release/bioc/html/Biostrings.html
pbm.motif	Jiang et al., 2013	http://thebrain.bwh.harvard.edu/hierarchicalANOVA
GraphPad Prism 5	Graph Pad Software	https://www.graphpad.com/; RRID:SCR_002798
Other		
Illumina NextSeq500 instrument	Illumina	Cat#SY-415-1001
Qubit fluorometer	Thermo Fisher Scientific	Cat#Q33216
Pippin prep	Sage Science	Cat#PIP0001
Gel Cassettes (3% agarose) for Pippin Prep	Sage Science	Cat#CSD3010
Custom-designed universal "all 10mer" oligonucleotide arrays	Aglient Technologies	Cat#AMADID 030236
GenePix 4400A microarray scanner	Molecular Devices	Cat#GENEPIX 4400
EnVision 2102 multimode plate reader	Perkin Elmer	Cat#2102-0020
Octet RED384	ForteBio	Cat#30-5101
Ni-NTA biosensors	ForteBio	Cat#18-5102

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed and will be fulfilled by the Lead Contact, Stuart H. Orkin (stuart_orkin@dfci.harvard.edu)

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Blood and Tissue Samples

Human peripheral blood stem cells (CD34$^+$ cells) from healthy donors (G-CSF mobilized, CD34 enriched) were purchased from Fred Hutchinson Cancer Research Center. The lot R002895 for experiments was collected from a health male and confirmed to express low level of *HBG1/2*. Samples were de-identified prior to purchasing, thus the research consent requirements were exempted from BCH IRB approval. The cells were thawed and recovered to EDM (IMDM (Corning, 15-016-CV) supplemented with 330 μg/mL Holo-Human Transferrin, 10 μg/mL Recombinant Human Insulin, 2 IU/mL Heparin, 5% Inactivated Plasma, 3 IU/mL Erythropoietin, 2 mM L-Glutamine) with three supplements (10^{-6} M hydrocortisone, 100 ng/mL SCF, 5 ng/mL IL-3) for 7 days to allow erythroid differentiation, then further differentiated in EDM with one supplement (100 ng/mL SCF) for another 2 days. Samples were collected on day 3, 5, 7, and 9 for CUT&RUN, RNA extraction and protein analysis.

Immortalized Cell Lines

Human HEK293T cells (female) were purchased from ATCC. The cells were cultured in DMEM, high glucose (Thermo Fisher Scientific, 11965) with 10% FCS and 2 mM L-Glutamine, and passaged every three days.

MEL (Murine Erythroleukemia, male) cells were purchased from Coriell Institute for Medical Research. The cells were cultured in DMEM, low glucose (Thermo Fisher Scientific, 11885) with 10% FCS, 2 mM L-Glutamine, and passaged every three days.

HUDEP-2 cell line (human umbilical cord blood-derived erythroid progenitor, male) was generated and kindly shared by Dr. Yukio Nakamura. HUDEP-2 cells were maintained in expansion medium: StemSpan SFEM (STEMCELL Technologies, 09650) with SCF (50 ng/ml), EPO (3 IU/ml), Dexamethasone (10^{-6} M), Doxycycline (1 μg/ml), and passaged every three days. The cell density was maintained within 20 thousand to 500 thousand cells/mL. Erythroid differentiation was carried out by replacing the medium to EDM (Erythroid Differentiation Media: IMDM (Corning, 15-016-CV) supplemented with 330 μg/mL Holo-Human Transferrin, 10 μg/mL Recombinant Human Insulin, 2 IU/mL Heparin, 5% Inactivated Plasma, 3 IU/mL Erythropoietin, 2 mM L-Glutamine) with two supplements (100 ng/mL SCF, 1 μg/mL doxycycline). After 5 days differentiation, cells were collected for chromosome conformation capture assay.

METHOD DETAILS

Functional Rescue Experiments

Bcl11a enhancer knockout (Enh-KO) MEL cells were generated by TALEN-mediated deletion of the 12-kb intronic enhancer(Bauer et al., 2013). *Bcl11a* exon knockout MEL cells were generated by TALEN-mediated deletion of exon1 (E1D) or exons 1 to 5 (E15D). Specifically, the TALENs were designed to generate double strand breaks flanking the *Bcl11a* exon1 or exon1 to 5 regions. The TALENs recognize the following sequences: TCTTGACTGGGCTGAAGC (5′L1), TGAAGTGGGGGGCTGGGGG (5′R1), TGACTGGG CTGAAGCGTC (5′L2), TATTGAAGTGGGGGCTGG (5′R2), TGTTTACAAGCACCGCGT (Intron1L1), TCCTCTGTCTGTTTGTTG (Intron1R1), TTTACAAGCACCGCGTGT (Intron1L2), TCCGTCCTCTGTCTGTTT (Intron1R2), TGTGCTAGCTCTTTGTTG (3′L1), TCCCGGTGGAAGAGGAAC (3′R1), TGCTAGCTCTTTGTTGAT (3′L2), and TCTCCCGGTGGAAGAGGA (3′R2). TALENs were synthesized and cloned into pcDNA3.1 vector with the FokI nuclease domain. 2.5 μg of each of the four TALEN plasmids together with 0.5 μg of pMax-GFP (Lonza) were electroporated to 2 × 10^6 MEL cells following manufacturer's procedure (Lonza, VCA-1005). Single cell-derived biallelic deletion clones were isolated by flow cytometry sorting of GFP positive cells followed by limiting dilution and genotyping analyses. The complete open reading frame (ORF) of human BCL11A isoforms (XL, L, S and XS) and various domain mutants were subcloned into the lentiviral vector pLVX-EF1a-IRES-zsGreen1 (Clontech #631982). Lentiviruses were packaged in HEK293T cells as described(Huang et al., 2016). Briefly, 2 μg of pΔ8.9, 1 μg of VSV-G and 3 μg sgRNA vectors were co-transfected into HEK293T cells seeded in a 10 cm Petri dish. Lentiviruses were harvested from the supernatant 48-72 hr post-transfection. *BCL11A* enhancer or exon knockout MEL cells were transduced with lentiviruses in 6-well plates and GFP-positive cells were FACS sorted 48 hr post-transfection. After sorting, cells were seeded in T25 flasks as pooled populations or in 96-well plates as single cell clones. 6-32 single cell-derived stable cell clones with ectopic expression of BCL11A isoforms or domain mutants were processed for gene expression analysis. Total RNA was isolated using RNeasy Plus Mini Kit (QIAGEN) following manufacturer's protocol. Quantitative RT-PCR (RT-qPCR) was performed using the iQ SYBR Green Supermix (Bio-Rad). Oligonucleotide sequences are listed in Table S3.

Western Blot

Western blot was performed as described(Xu et al., 2012). Briefly, samples were boiled in 1x SDS loading buffer to denature all proteins and separated with 13% or gradient SDS-PAGE gels. Proteins were then transferred to PVDF membrane with a standard wet transfer system at 2.5 mA/cm^2 for 2 hr. Membranes were blocked with 5% nonfat milk for 1 hr and then incubated with primary antibodies for 1 hr at room temperature or overnight at cold room with shaking. Excess antibodies were washed with TBS-T (50 mM Tris pH 8.0, 150 mM NaCl, 0.1% Tween 20) for 3 times and HRP-conjugated secondary antibodies were incubated for 30 min at room temperature. After 3 washes with TBS-T, the membranes were developed with Immobilon Western Chemiluminescent HRP Substrate (Millipore, WBKLS0500). The following antibodies were used: M2-Flag (F1804, Sigma-Aldrich), BCL11A (ab19487 and ab191401, Abcam), and GAPDH (sc-25778, Santa Cruz Biotechnology). All antibodies were used at 1:1,000 dilutions in TBS-T.

Protein Binding Microarray (PBM) Experiments

For PBM experiments, ZnF23 and 456 from BCL11A and BCL11B were cloned into pDEST15 (Invitrogen) with an N-terminal glutathione S-transferase (GST) tag using the Gateway system (Invitrogen). The ZnF domains were expressed by *in vitro* transcription-translation (IVT) using PURExpress (New England Biolabs) supplemented with RNase Inhibitor (New England Biolabs) and 50 μM zinc acetate. Protein concentrations were determined by immunoblotting with a GST dilution series. BCL11A XL and BCL11B were cloned into vector pT7CFE (Thermo Fisher) with an N-terminal HA tag. The full-length proteins were expressed by IVT using the 1-Step Human Coupled IVT kit (Thermo Fisher) supplemented with 50 μM zinc acetate. Concentrations of full-length proteins from IVT reactions were determined by comparison to pure BCL11A-XL of known concentration on western blot with antibodies against BCL11A (Abcam, ab18688) and HA (Santa Cruz Biotechnology, sc-805 HRP).

PBM experiments were performed as described (Berger and Bulyk, 2009; Berger et al., 2006), except as noted here. Custom-designed "all 10mer" universal oligonucleotide arrays in 8 × 60K GSE array format (Agilent Technologies; AMADID, 030236) (Nakagawa et al., 2013) were double-stranded. Each BCL11 wild-type and mutant protein was assayed at a final concentration of 600 nM in a standard PBM binding reaction containing 50 μM zinc acetate on a fresh slide. Protein binding was detected using an anti-HA antibody (Life Technologies, A-21287) or an anti-GST antibody (Life Technologies; catalog, A-11131) at a final concentration of 25 μg/mL and fluorescence measurements were obtained using a GenePix 4400A (Molecular Devices) microarray scanner. Subsequent data quantification, normalization, and DNA binding specificity analysis were performed as described previously using the Universal PBM Analysis Suite and the Seed-and-Wobble motif-derivation algorithm (Berger and Bulyk, 2009; Berger et al., 2006). Average E-scores for TGACCA and TNCGGCCA sequences were calculated by taking the mean of the E-scores of all 8-mers containing matches to these sequences. Average E-scores for 8-mers scoring E > 0.40 by at least one BCL11 construct were plotted with the heatmap.2 function in the gplots R package using the Pearson correlation distance metric and single-linkage clustering. Unweighted clusters of 8-mers were then aligned to generate a PWM as previously described(Jiang et al., 2013) and associated sequence logos were made using the makePWM and seqLogo functions from the Biostrings and seqLogo R packages, respectively. Plots were made using the standard Boxplot function in R.

Protein Purification and Biophysical Assessment of Binding

For fluorescence polarization experiments, 6xHis-tagged ZnF23 and 456 were cloned into vector pMCSG7 (Midwest Center for Structural Genomics) in a ligation-independent manner using T4 DNA Polymerase (Millipore, 70099). The construct was transformed into Rosetta(DE3) pLysS competent cells (Millipore, 70956-4) for protein expression. Protein expression was induced with 1 mM IPTG (Roche) and supplemented with 100 μM $ZnSO_4$ (Sigma-Aldrich) at 20°C for 18 hr. Cells were resuspended in lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole, 1mM DTT, 8 M urea, protease inhibitor cocktail), sonicated, spun, and the lysates were incubated with Ni-NTA agarose resin (QIAGEN). After washing, proteins were eluted with lysis buffer containing 300 mM imidazole, and dialyzed overnight at 4°C against minimal buffer (20 mM Tris pH 8.0, 50 mM NaCl, 1mM DTT). The following day, the proteins were purified on a UNO S1 column (Bio-Rad) and dialyzed overnight at 4°C against storage buffer (20 mM Tris pH 8.0, 50 mM NaCl, 1 mM DTT, 10% glycerol v/v, 20 μM $ZnSO_4$). The proteins were then concentrated using Amicon ultra-15 centrifugal filter units with a 3K MWCO (Millipore), aliquoted, flash frozen, and stored at −80°C.

Fluorescence polarization experiments were performed as described(Jameson and Seifried, 1999) in 384-well plates (Corning, 3820) using the EnVision 2102 multimode plate reader (Perkin Elmer) running version 1.13 of the EnVision Manager software. Briefly, increasing concentrations of His-ZnF23 or His-ZnF456 (0-3 μM) were titrated into a total of 20 μL assay buffer (10 mM Tris pH 8.0, 150 mM NaCl) containing 10 nM 6-FAM-labeled double-stranded oligonucleotide (Integrated DNA Technologies), whose sequences are listed in Table S3. Samples were measured at an excitation wavelength of 480 nm and an emission wavelength of 535 nm. Data were analyzed using GraphPad Prism 5 (Graph Pad Software), plotting mP against protein concentration and fitting the curve to a single-site binding model (hyperbola) to extract a K_D value.

Proteins for Octet analysis were expressed and purified similarly to those for fluorescence polarization with the following exceptions. Rosetta(DE3) pLysS expressing 6xHis-tagged ZnF23 or ZnF456 were resuspended in binding buffer (20 mM sodium phosphate buffer pH 8.0, 500 mM NaCl, 0.1% NP40, 10 mM imidazole) and sonicated (Qsonica sonicators). Cell lysates were spun down and supernatants were run on a Ni-NTA column (MCLAB, NiNTA-300). After washing, the proteins were eluted with 300 mM imidazole, purified on a size exclusion column (GE Healthcare HiLoad 16/60 75 pg), and dialyzed overnight at 4°C against storage buffer. The proteins were then concentrated, aliquoted, flash frozen, and stored at −80°C.

Octet Red384 (ForteBio, Pall Science) was used for binding studies(Do et al., 2008). Ni-NTA biosensors (ForteBio) were used to immobilize 6xHis-tagged proteins. Ni-NTA biosensors were pre-wet for 60 s in kinetic buffer (PBS, 0.02% Tween20, 0.05% sodium azide), then immersed in ligand solution (10 μg/ml His-ZnF456 or 10 μg/ml His-ZnF23) for 180 s, and finally immersed in kinetic buffer for 60 s. The kinetics of DNA association was monitored by moving sensors into wells containing 0.3 nM-10 μM dsDNA (sequences in Table S3) for 180 s. This was followed by dissociation in kinetic buffer for 180 s. During the entire kinetic assay, the sample plate was kept shaking at 1000 rpm. A column of biosensors without ligand was titrated with analyte and used as a parallel reference control. A ligand-loaded biosensor without analyte was also used as baseline. Data were analyzed using GraphPad Prism 5 (Graph Pad Software), plotting Response against DNA concentration and fitting the curve to a single-site binding model to extract a K_D value.

Generation of *HBG* Promoter-edited Cells

Cas9 expression vector (Addgene plasmid ID 52962) and sgRNA targeting common sequences at the *HBG1* and *HBG2* genes 115 bp upstream of the transcription start sites (TSS) were sequentially introduced into HUDEP clone 2 (HUDEP-2) cells via lentiviral transduction (Canver et al., 2015). Briefly, lentivirus packaging was carried out as described in section "Functional Rescue Experiments." HUDEP-2 cells were infected with Cas9 expression virus and selected with 10 μg/mL blasticidin. The sgRNA sequence was cloned into lentiGuide-Puro (Addgene plasmid ID 52963) (Sanjana et al., 2014), and transfected into HEK293T cells to generate virus. 10^5 Cas9-expressing cells were transduced with the virus containing the sgRNA-encoding plasmid and incubated for 7-10 days with 10 μg/mL blasticidin and 1 μg/mL puromycin selection to allow for editing. These bulk cultures were plated clonally at limiting dilution and grown for approximately 14 days.

Screening of −115 Edited HUDEP-2 Clones

Genomic DNA extracted from HUDEP-2 clones was amplified using primer pairs specific to *HBG1* or *HBG2* (Table S3). Unique identification of *HBG1* versus *HBG2* was achieved by designing primer sequences overlapping nucleotide variants between *HBG1* and *HBG2*. Additional variants were present within the amplicon to ensure single gene-specific amplification. PCR was performed using the QIAGEN HotStarTaq 2 × master mix and the following cycling conditions: 95°C for 15 minutes; 35 cycles of 95°C for 15 s, 60°C for 1 minute, 72°C for 1 minute; 72°C for 10 minutes. Amplicons were purified using the QIAquick® PCR Purification Kit (Cat. No. 28106) and Sanger sequenced or sent to amplicon sequencing.

For Sanger sequencing, sequences were aligned to a reference sequence obtained from sequencing HUDEP-2 cells. Edited alleles were identified by comparing the Sanger sequence chromatogram to the HUDEP-2 reference sequence. One clone, D3, was found to contain edited versions of all four alleles for *HBG1* and *HBG2* at the −115 promoter site.

For amplicon sequencing of bulk edited cells, reads were aligned to amplicon sequence of *HBG1* or *HBG2* promoters (∼1700 bp, see Table S3 for genotyping primers) using Bowtie2. Reads which did not align to any amplicon were discarded. Remaining reads were analyzed to determine the mutation type (i.e., deletion) and deleted position by comparing to the wild-type sequence. Reads were tallied to determine the proportion of different deletion variants.

Validation of HbF Enrichment

$5x10^4$ of the D3 single-cell HUDEP-2 clones were stained for HbF expression by intracellular flow-cytometry (Canver et al., 2015). Briefly, cells were fixed with 0.05% glutaraldehyde in PBS, incubated at room temperature for 10 min and centrifuged at 600 x g for 5 min. Cells were then treated with 0.1% Triton in PBS with 0.1% BSA, incubated at room temperature for 5 min, and centrifuged at 600x g for 15 min. Cells were resuspended in PBS/BSA and incubated with 2 μL of FITC-conjugated anti-Human Fetal Hemoglobin antibody in the dark at room temperature for 15 min. Cells were analyzed using an LSRII flow cytometer, recording 10,000 events per condition. In addition, RNA was obtained from single-cell HUDEP-2 clone D3 and control HUDEP-2 cells and prepared for RT-qPCR (Canver et al., 2015) using *CAT* and *HBA1/2* as reference genes. Expression data represent the mean of at least three replicates.

Chromatin Immunoprecipitation

10^7 expansion phase HUDEP-2 cells were collected and fixed with 1% formaldehyde for 5 minutes at room temperature. Fixation was quenched with 125 mM glycine. Cells were washed with ice cold PBS twice and resuspended in 0.13 mL lysis buffer (50 mM Tris-HCl 8.0, 10 mM EDTA, 0.5% SDS) and sonicated in a microtube (Covaris, 520045) with Covaris E220 ultrasonicator (Covaris). 0.12 mL sonicated chromatin was mixed with 1 mL ChIP dilution buffer (20 mM Tris-HCl 8.0, 2 mM EDTA, 1% Triton X-100, 300 mM NaCl, protease inhibitor), 20 μL Dynabeads protein G (Thermo Fisher Scientific) and 3 μg antibody (CTCF, 07-729, Millipore). After overnight rotating, the beads were washed with the following buffers: twice with RIPA150 (20 mM Tris-HCl 8.0, 1 mM EDTA, 1% Triton X-100, 150 mM NaCl, 0.1% Sodium Deoxycholate, 0.1% SDS), twice with RIPA500 (20 mM Tris-HCl 8.0, 1 mM EDTA, 1% Triton X-100, 500 mM NaCl, 0.1% Sodium Deoxycholate, 0.1% SDS), twice with LiCl buffer (10 mM Tris-HCl 8.0, 1 mM EDTA, 0.5% NP-40, 250 mM LiCl, 0.5% Sodium Deoxycholate), twice with TE buffer (10 mM Tris-HCl 8.0, 1 mM EDTA). To elute and decrosslink, 300 μL Elution buffer (50 mM Tris-HCl 8.0, 10 mM EDTA, 1% SDS, 150 mM NaCl, 0.1 mg/mL Proteinase K) was added to the beads and incubated at 65°C overnight. The eluted material was extracted with phenol-chloroform, and DNA was precipitated by adding 750 μL absolute ethanol. DNA was pelleted by centrifuging at 14000 rpm for 15 minutes, washed once with 75% ethanol, then dried and dissolved with 50 μL TE buffer. The quality of ChIP was verified by real-time PCR. To construct ChIP-seq library, we used NEBNext Ultra II DNA Library Prep Kit (NEB) according to manufacturer's protocol. Quality check of library was carried out with Qubit and bioanalyzer. The library was sequenced in the NextSeq 500 platform with NextSeq 500/550 High Output Kit v2 (75 cycles). Paired-end sequencing was performed (2x42 bp, 6 bp index).

CUT&RUN

CUT&RUN experiments were carried out as described (Skene and Henikoff, 2017) with modifications. Briefly, nuclei from $2x10^6$ cells were isolated with NE buffer (20 mM HEPES-KOH, pH 7.9, 10 mM KCl, 0.5 mM Spermidine, 0.1% Triton X-100, 20% Glycerol and 1x protease inhibitor cocktails from Sigma), captured with BioMagPlus Concanavalin A (Polysciences) and incubated with primary antibody for 2 hours. After washing away unbound antibody with wash buffer (20 mM HEPES-NaOH pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 0.1% BSA and 1x protease inhibitor cocktails from Sigma), protein A-MNase was added at a 1:1000 ratio and incubated for 1 hour. The nuclei were washed again and placed in a 0°C metal block. To activate protein A-MNase, $CaCl_2$ was added to a final concentration of 2 mM. The reaction was carried out for different time courses and stopped by addition of equal volume of 2XSTOP buffer (200 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 μg/mL RNase A and 40 μg/mL 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 fluorometer and bioanalyzer quality control. Protein A-MNase (batch 5) was kindly provided by Dr. Steve Henikoff. The antibodies used were: GATA1, ab11852, abcam; CTCF, 07-729, Millipore; BCL11A, ab191401, abcam; normal rabbit IgG, 12-370, Millipore.

Library Preparation and Sequencing for CUT&RUN

To construct the CUT&RUN DNA library for sequencing on an Illumina platform, we modified the protocol of NEBNext Ultra II DNA Library Prep Kit, NEB (Ipswich, USA), aiming to preserve short DNA fragments (30-80 bp). Briefly, dA-tailing temperature was decreased to 50°C to avoid DNA melting, and the reaction time was increased to 1 hour to compensate for lower enzymatic activity. After adaptor ligation, 2.1x volume of AMPure XP beads was added to the reaction to ensure high recovery efficiency of short fragments. After 12 cycles of PCR amplification, the reaction was cleaned up with 1.2x volume of AMPure XP beads. 16-24 barcoded libraries were quantified and mixed at equal molar ratio. PCR dimers were removed with Pippin prep size selection according to manufacturer's manual, and the mixed library was denatured according to the standard protocol from Illumina. 1.3 mL of 1.8 pM diluted library was loaded to a NextSeq 500/550 High Output Kit v2 (75 cycles), and sequenced in the NextSeq 500 platform. To enable determination of fragment length, paired-end sequencing was performed (2x42 bp, 6 bp index).

CUT&RUN Data Processing

Paired-end sequencing reads were trimmed with Trimmomatic (Bolger et al., 2014) to remove adaptor sequences from the 3′ end of each read using parameters ILLUMINACLIP: 2:15:4:4:true SLIDINGWINDOW:4:15 MINLEN:25. Reads which were 25 bp or longer after trimming, and were paired between the mates were kept. Then reads were aligned to the reference human hg19 assembly (Schatz and Langmead, 2013) with settings --end-to-end,--dove-tail,--phred33. The --dove-tail considers mates that overlap with each other, usually when fragment length is less than read length, as a concordant alignment. The resulting alignments, recorded

in BAM file, were sorted, indexed, and marked for duplicates with Piccard (http://broadinstitute.github.io/picard) MarkDuplicate function. Afterward, the BAM file was filtered with SAMtools (Badis et al., 2009) to discard reads, mates that were unmapped, or PCR/optical duplicates (-f 3 -F 4 -F 8 -F 1024). Fragments of size less than 120 bp were kept (Skene and Henikoff, 2017).

Peak Calling and Overlap with ChIP-seq

MACS (Zhang et al., 2008) version 2.1 was used to call peaks from the BAM file with narrowPeak setting, p value cutoff of 1e-5, and with --SPMR which normalizes the signal intensity to facilitate comparison between tracks. To compare CUT&RUN with ChIP-seq, we sampled the number of paired reads per experiment to 18 million, and called peaks with default narrowPeak setting in MACS2. We then computed the Jaccard overlap coefficient among the top 15,000 peaks called per experiment. To evaluate the signal-to-noise ratio of GATA1 CUT&RUN and GATA1 ChIP-seq, we computed the fraction of total reads located in top genomic bins (of 500 bp) with highest read coverage. This was achieved with the plotFingerprint tool using default settings (Ramírez et al., 2016).

De Novo Motif Discovery from Peaks

To discover possible enriched sequences among the BCL11A CUT&RUN peaks, we performed *de novo* motif discovery with MEME motif suite (Bailey and Elkan, 1994; Machanick and Bailey, 2011). Specifically, sequences from −100 bp to +100 bp from the summit position of all peaks called by MACS were analyzed for enrichment. Top enriched motifs (up to 6 motifs) were collected from running DREME (Bailey, 2011) and MEME on each batch of sequences.

Footprint Detection and Motif Discovery within Footprints

A footprint is a signature cut pattern left by the protein A-MNase enzyme as it cuts around regions bound by a TF. A footprint has a characteristic low number of cuts within core region containing consensus motif that is well protected from enzyme cuts, and a high number of cuts in the two flanking regions of the core. To detect footprints, we first enumerate all fragment ends to determine the precise cut location of the enzyme. We used BEDtools (Quinlan, 2014) bamtobed, with parameter -bedpe, to determine the fragment ends. We then ran a simple calculation termed Footprint Occupancy Score (FOS)(Neph et al., 2012) to detect footprints. Significant footprints were next analyzed for enriched motif sequence with the MEME discovery tool. Overall, this analysis allowed ascertainment of binding sites of CUT&RUN which met two important criteria: 1) presence of a consensus motif, and 2) location within a footprint region (i.e., protection from nuclease digestion).

Single-locus Footprinting, and Global Footprinting Analysis Targeted toward TGACCA Motif

We produced a single-base resolution footprinting profile for the beta-globin region. To do so, we first enumerated all fragment ends as above, and then combined cuts from all CD34$^+$ CUT&RUN BCL11A experiments to produce an aggregate footprinting profile.

To fit a binding probability model for BCL11A, we relied on the motif obtained by PBM and by the PWM determined from motif analysis within the footprints. The consensus motif sequence pointed to TGACCA. With this consensus, the binding probability model was estimated as follows. 1) We first used FIMO (Grant et al., 2011) to find all occurrences of TGACCA within peak regions of the genome identified by MACS. 2) For all occurrences of TGACCA, we constructed a cut frequency matrix with make_cut_matrix.py (https://github.com/Parkerlab/atactk) where the rows of this matrix were occurrences of TGACCA, and the columns were the cuts at each base for the 206-base motif-centered region (100-base each flank, 6-base motif). 3) We used CENTIPEDE (Pique-Regi et al., 2011) to learn a binding model from the observed cut frequency matrix with default parameters. This method assumes that the sites unbound by a TF differ in many ways from sites bound by a TF, such as in their cut frequencies. It uses an unsupervised Bayes mixture model to classify each site as either bound or unbound by TF and outputs a probability value. Using this approach, we quantified the binding probability for the *HBG1/2* promoter at the distal and proximal motif locations independently.

It should be noted that since *HBG1* and *HBG2* promoters are duplicates of each other, we summed the cut frequency from both regions and computed one binding log odds for the *HBG1/2* promoters. We also report log-odds scores for the combined CD34$^+$ experiments, sub-sampled to 40 million reads, and proceeded to build a model from the combined experiment.

Chromosome Conformation Capture Analysis

Chromosome conformation capture (3C) was performed as described previously (Xu et al., 2010). 10^7 HUDEP-2 and derived cell lines at day 5 of erythroid differentiation were crosslinked with 2% formaldehyde for 10 minutes at room temperature. The cells were lysed for 15 minutes with ice cold lysis buffer (10 mM Tris-HCl at pH 8.0, 10 mM NaCl, 0.2% NP-40), and resuspended in 0.5 mL 1.2x NEB CutSmart buffer. 10 μL 10% SDS was added to the nuclei, then shaken for 20 minutes at 65°C. Triton X-100 was added to 2% and nuclei were shaken for 1 hour at 37°C. Digestion was done by adding 300 U of EcoRI-HF (NEB, R3101) to the nuclei and incubating overnight at 37°C. 200 U EcoRI-HF was added again and incubated for another 3 hours. 88 μL 10% SDS was added to the nuclei, followed by 65°C incubation for 20 minutes. To ligate the digested chromatin, the following components were added to the nuclei and brought up to 7 mL by distilled water: 0.7 mL 10x ligation buffer (NEB), 350 μL 20% Triton X-100, and 70 U T4 DNA ligase (Thermo Fisher, EL0011). The reaction was carried out at 16°C for 4 hours and room temperature for 30 minutes. The DNA was purified by phenol extraction and isopropanol precipitation. Quantitative RT-PCR was performed with iQ SYBR Green mix (Biorad). Samples were normalized to an independent *ERCC3* locus. Primers are listed in Table S3.

ATAC-seq

ATAC-seq was performed with Illumina Nextera DNA Preparation kit (Illumina, FC-121-1030) as previously described (Buenrostro et al., 2015). 50,000 cells of expansion phase HUDEP-2 or derived cell lines were collected, and permeabilized with 50 μL ice cold lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl$_2$, 0.1% IGEPAL CA-360) for 10 minutes. The transposition reaction was carried out at 37°C for 60 minutes in 50 μL volume containing 25 μL 2x TD buffer and 2.5 μL transpose enzyme (contained in Illumina Nextera DNA Preparation kit). DNA was purified with QIAGEN MinElute PCR purification kit. Library amplification was done with NEBNext Ultra II Q5 Master Mix (contained in NEBNext Ultra II DNA Library Prep Kit). PCR amplification was carried out as follow: 72°C 5 min, 98°C 30 s, 98°C 10 s, 63°C 30 s, 72°C 1 min, repeat steps 3-5 for another 7 cycles, hold at 4°C. The resulting libraries were purified using QIAGEN MinElute PCR purification kit, and quantified with Qubit fluorometer and bioanalyzer, and then sequenced in NextSeq 500 platform. Sequencing parameter was paired end 2x41 bp reads, 8 bp index.

ATAC-seq Data Analysis

In processing ATAC-seq data, paired-end reads were trimmed and aligned to hg19 using Trimmomatic and Bowtie2 respectively. For promoter-edited samples, these were aligned to a custom hg19 genome where there is a 1 bp deletion in *HBG1* promoter and a 13 bp deletion in the *HBG2* promoter (exact locations shown in Figure S6A). Reads were piled-up using MACS2 --SPMR -B parameters and visualized with IGV.

QUANTIFICATION AND STATISTICAL ANALYSIS

Sample size, mean, and significance *P* values are indicated in the text, figure legends, or Method Details. Error bars in the experiments represent standard error of the mean (SEM) from either independent experiments or independent samples. Statistical analyses were performed using GraphPad Prism, or reported by the relevant computational tools. Detailed information about statistical methods is specified in figure legends or Method Details.

DATA AND SOFTWARE AVAILABILITY

All raw and processed CUT&RUN, ChIP-seq and ATAC-seq data have been deposited in the NCBI Gene Expression Omnibus under accession number GEO: GSE104676.

PBM data have been deposited in UniPROBE (http://the_brain.bwh.harvard.edu/uniprobe/) under accession number UniPROBE: LIU18A.
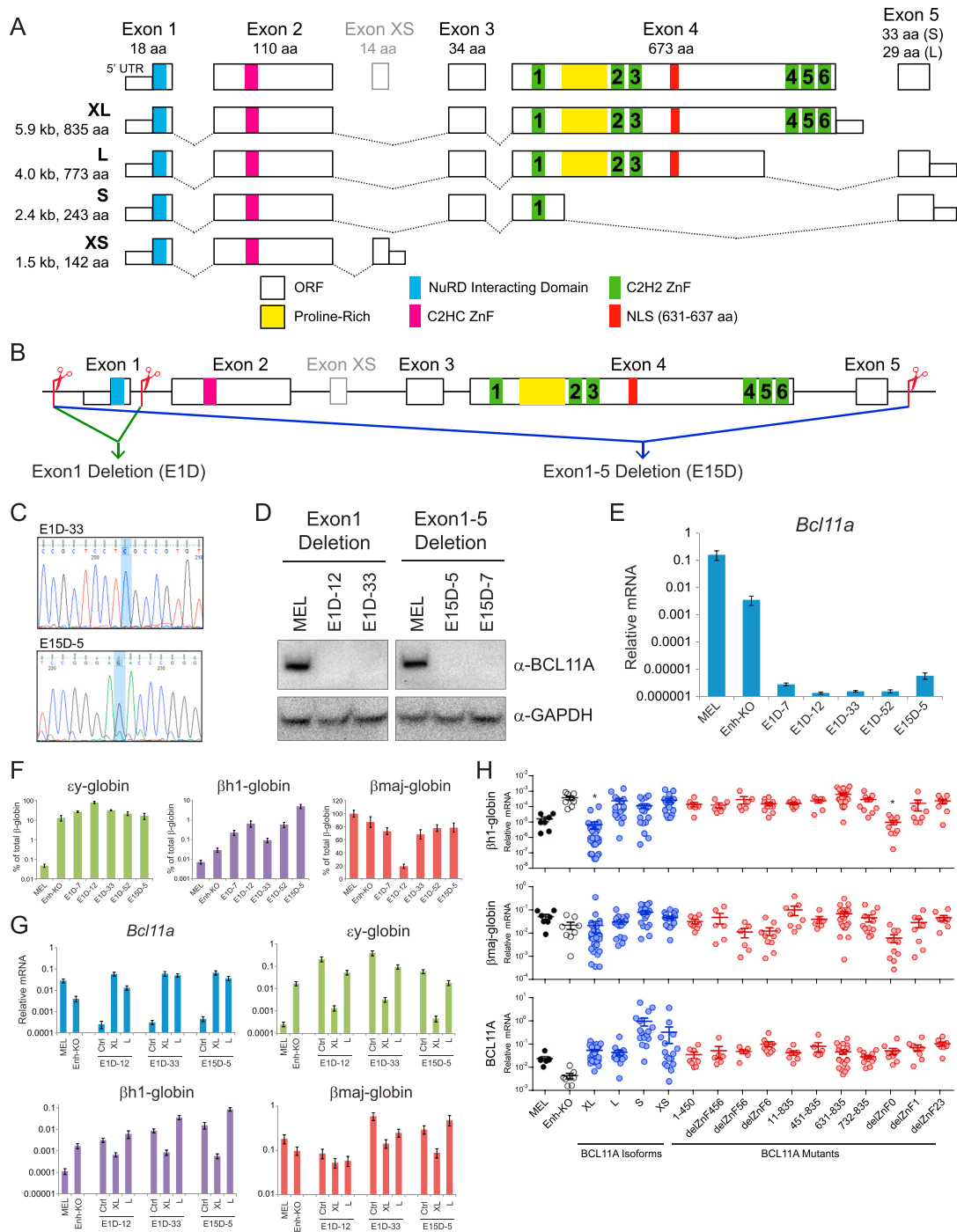
Figure S1. Functional Rescue Analysis in *BCL11A* Exon Knockout Cells, Related to Figure 1

(A) Schematic of human BCL11A isoforms.

(B) Schematic of TALEN-mediated deletion of *BCL11A* exon 1 (E1D) or exon1 to 5 (E15D) in murine MEL cells.

(C) Representative Sanger sequencing cytographs of *BCL11A* exon knockout clones.

(D) western blot analysis validated the complete absence of BCL11A protein expression in exon knockout cells.

(E) RT-qPCR analysis validated the loss of BCL11A mRNA expression relative to the unmodified MEL or Enh-KO cells.

(F) Complete knockout of BCL11A expression led to significant derepression of εy- and βh1-globin genes. The mRNA expression of εy, βh1 and βmajor is shown as the % of total β-like globin genes.

(G) Ectopic expression of BCL11A-XL but not the L isoform restored the stable repression of εy- and βh1-globin genes in multiple independent *BCL11A* exon knockout MEL cell lines.

(H) Expression of full-length BCL11A-XL, but not domain mutants lacking the NuRD-interacting domain, ZnF23 or ZnF456, restored repression of βh1-globin. Expression of βmajor-globin remained largely unaffected. Each circle denotes an independent single-cell-derived stable cell clone.

Results are mean ± SEM of multiple independent clones and analyzed by one-way ANOVA with repeated-measures. *p < 0.05. Results are mean ± SEM of three experiments for (E–G).
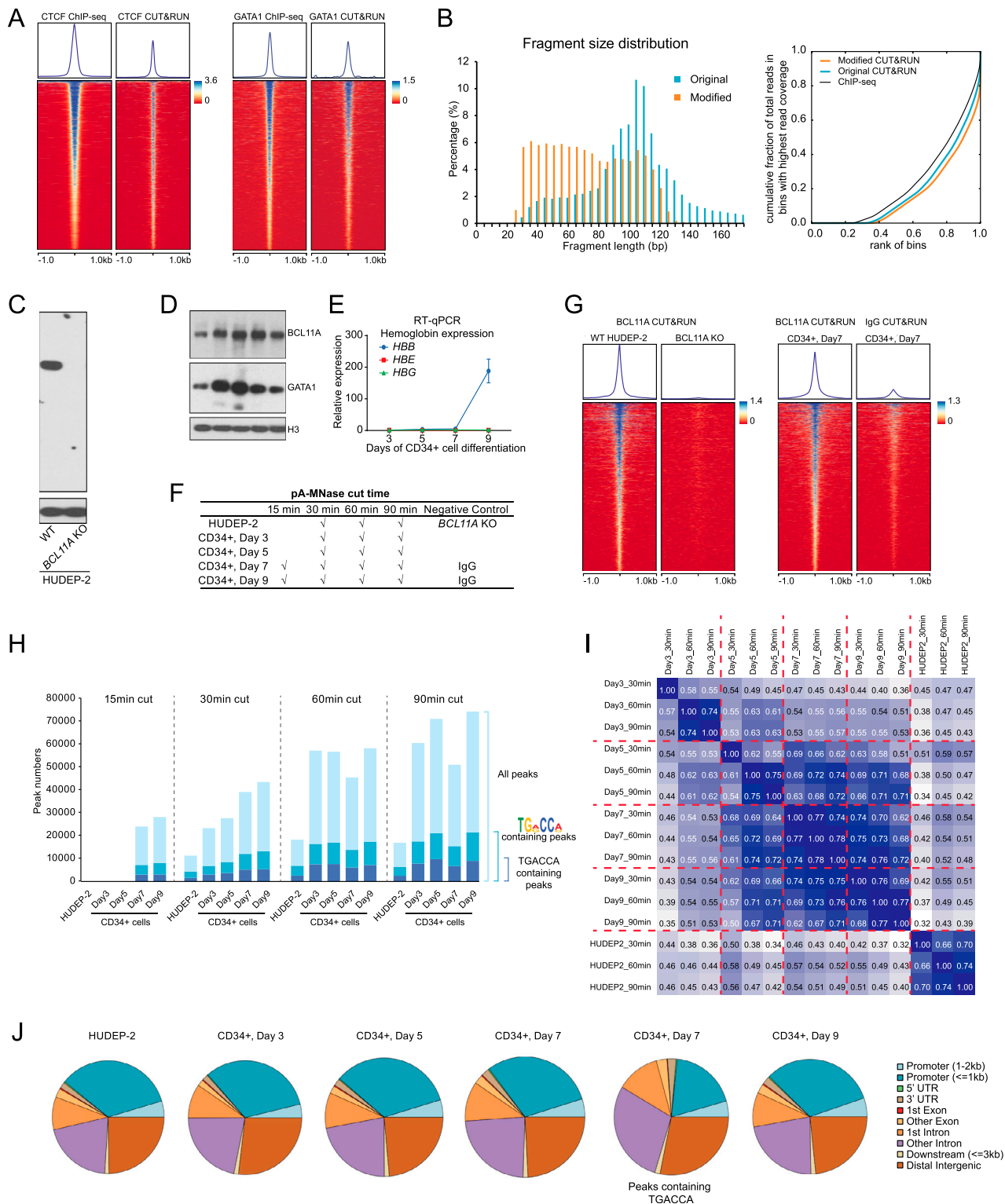
Figure S2. Optimization and Application of CUT&RUN, Related to Figure 4

(A) Heatmap comparison of overlapping peaks between ChIP-seq and CUT&RUN (see Methods).

(B) Left, Fragment length distributions of original and modified CUT&RUN protocols. Fragment ends were enumerated and used to calculate the fragment length. Right, Signal-to-noise measured by the total number of reads in top ranked randomly sampled bins (each 500 bp) with plotFingerprint. A steep rise to the right of the plot indicates a better signal enrichment (see Methods).

(legend continued on next page)

(C) western blot showing the specificity of BCL11A antibody.

(D) Protein levels of BCL11A and GATA1 in expansion phase HUDEP-2 and differentiating CD34$^+$ cells analyzed by western blot of whole cell lysates. Histone H3 was used as loading control. Left to right: HUDEP-2, CD34$^+$ cells on differentiation day 3, 5, 7, and 9.

(E) mRNA level of β-like globin genes in differentiating CD34$^+$ cells analyzed by RT-qPCR. Results are shown as mean ± SEM of three experiments.

(F) Experimental design of BCL11A CUT&RUN in HUDEP-2 cells and differentiating CD34$^+$ cells.

(G) Heatmaps showing the BCL11A CUT&RUN peaks in HUDEP-2, *BCL11A* KO HUDEP-2 cells, and in CD34$^+$ cells.

(H) Peak numbers of all BCL11A CUT&RUN. The fraction of peaks that contain the motif is shown in dark blue.

(I) Pairwise overlap of peaks for all BCL11A CUT&RUN experiments. Peaks were called by MACS2 with narrowPeak setting. Overlap refers to 1-bp overlap between peaks.

(J) Peak distribution of BCL11A CUT&RUN. Data for 60 min or 30 min protein A-MNase digestion are shown for HUDEP-2 and CD34$^+$ cells, respectively.
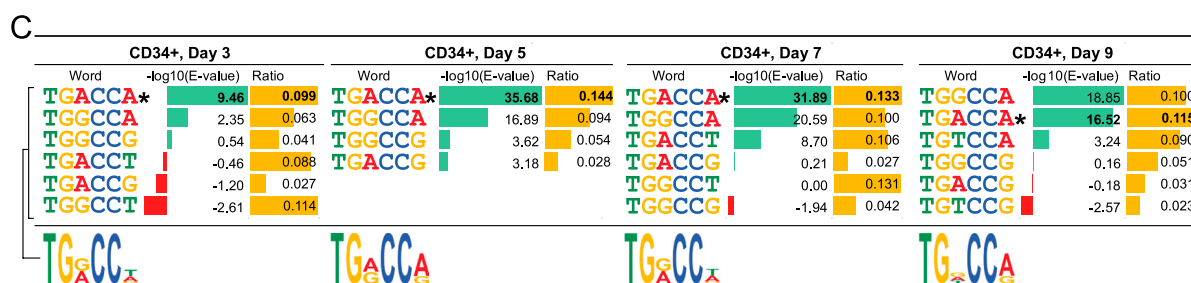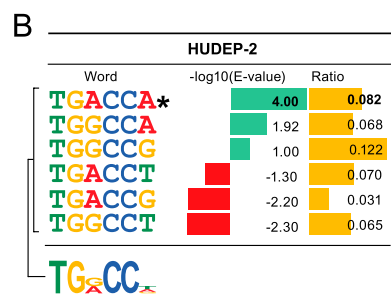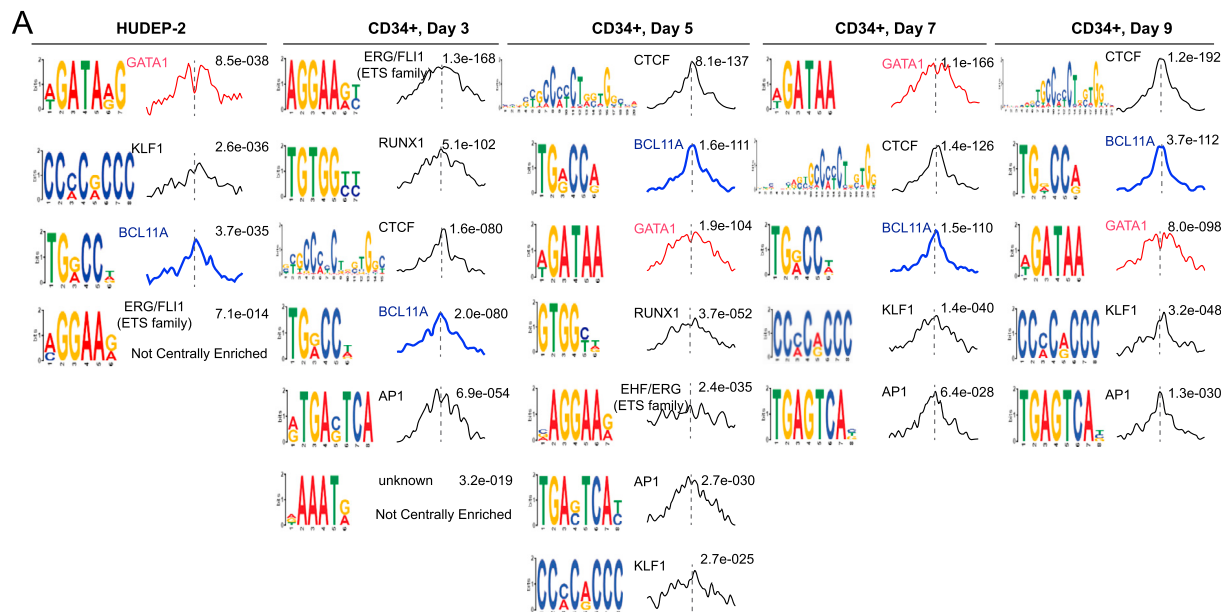
Figure S3. Motif Discovery for BCL11A CUT&RUN, Related to Figure 4

(A) Motifs discovered in HUDEP-2 and each stage of CD34[+] cells. Each motif and its position and likely binding factor are shown. Data for 60 min or 30 min protein A-MNase digestion are shown for HUDEP-2 and CD34[+] cells, respectively. Results of other cut times were similar and not shown. E-values shown in upper right were reported by MEME.

(B and C) Comparison of all combinations of TG(A/G)CC(A/C/T) in BCL11A CUT&RUN in HUDEP-2 (B) or CD34[+] cells (C). The sequences are ranked by –log10(E-value), where E-value is the probability of occurrence reported by MEME. The column "Ratio" shows the proportion of peaks that contain the corresponding sequence. The datasets used are 60 min cut in HUDEP-2 and 30 min cut in CD34[+] cells.
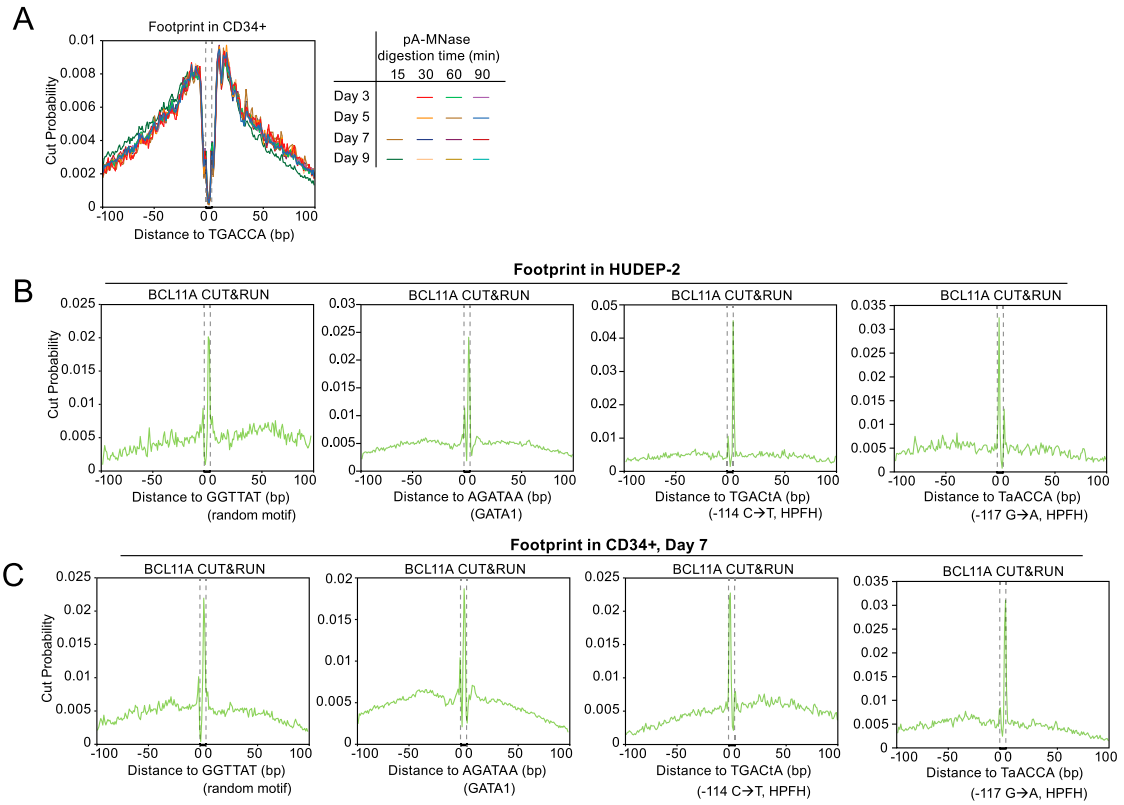
Figure S4. Footprint Analysis for BCL11A CUT&RUN, Related to Figure 4

(A) Targeted motif footprint analysis for BCL11A CUT&RUN in CD34[+] cell experiments. Cut probability for each base on TGACCA and surrounding sequences was plotted.

(B and C) Targeted motif footprints in BCL11A CUT&RUN for control sequences in HUDEP-2 (B) or CD34[+] cells (C). Cut probabilities for each base surrounding indicated motifs were plotted.
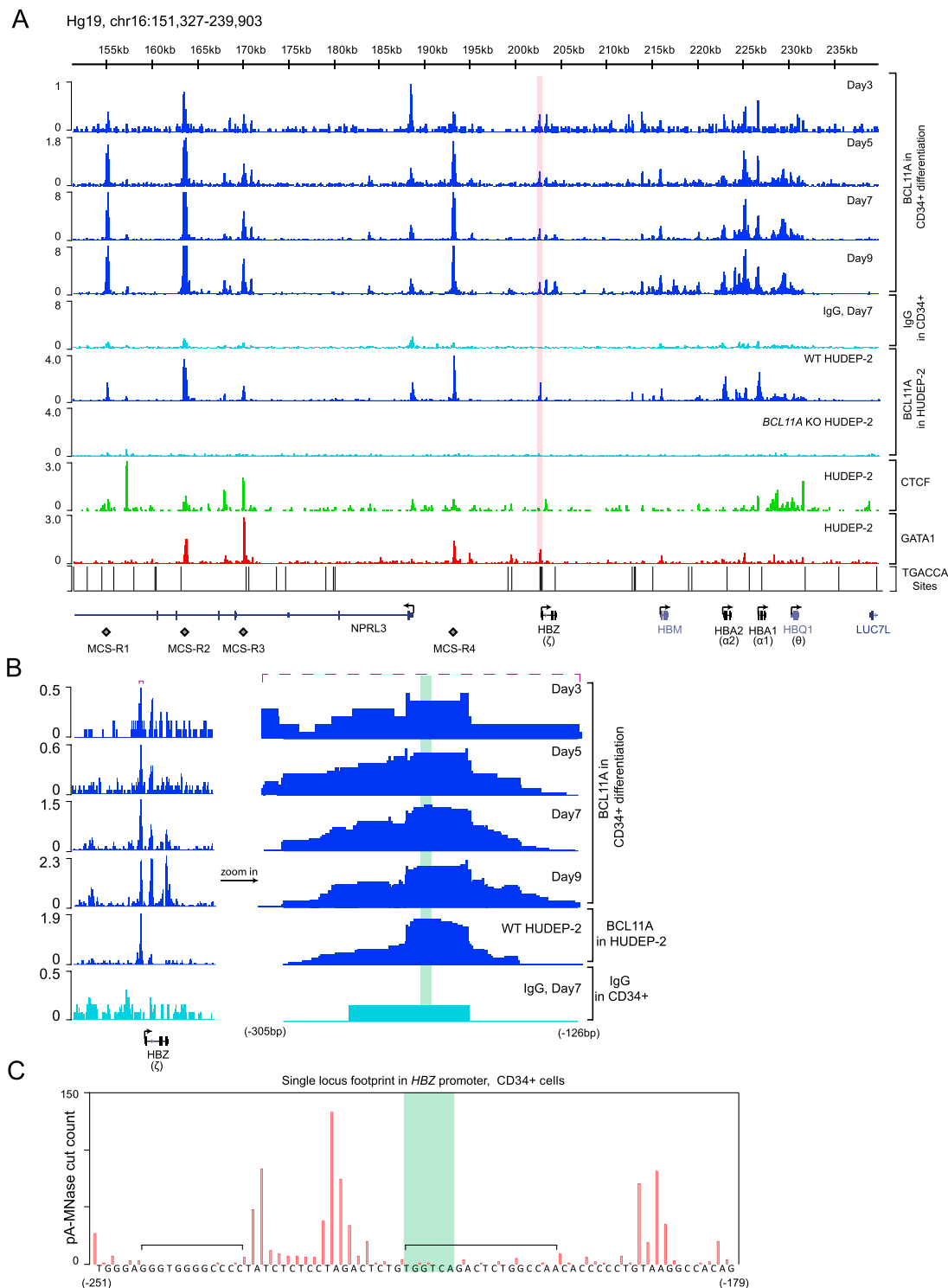
Figure S5. High-Resolution CUT&RUN Profiles in α-Globin Region, Related to Figure 5

(A) CUT&RUN profiles in α-globin cluster. Antibodies and cell types for each track are shown on the right. The promoter of ζ-globin gene (*HBZ*) is highlighted in pink.

(B) Left, BCL11A binding at *HBZ* promoter across multiple CUT&RUN experiments. Right, zoom in view of 180 bp of *HBZ* promoter region. The TGACCA motif is highlighted in green.

(C) Single locus footprint analysis shows the cut frequency at each nucleotide of the *HBZ* region (from −251 to −179 relative to TSS). 14 CUT&RUN experiments in CD34+ cells are combined for this analysis.
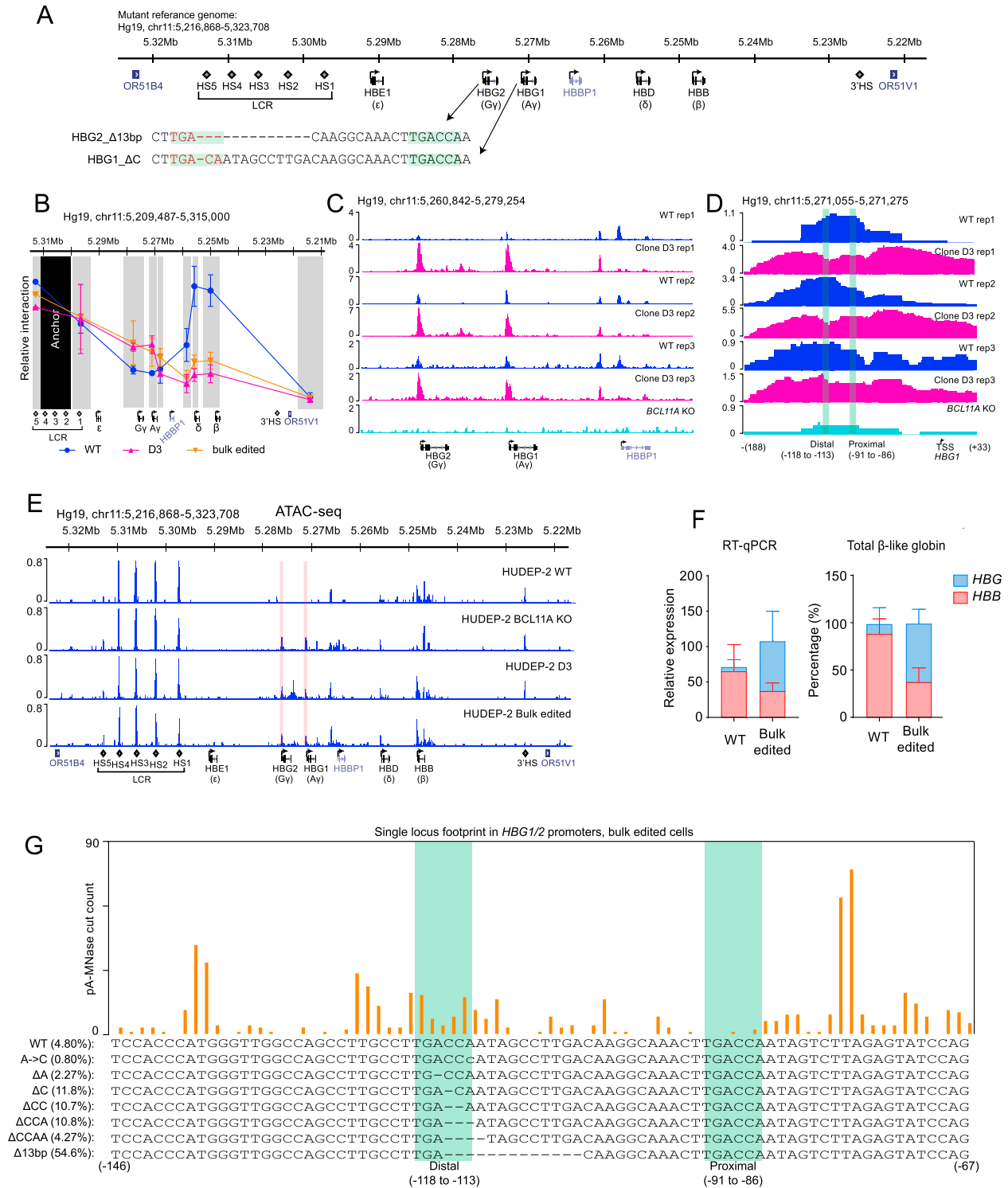
Figure S6. CUT&RUN in γ-Globin Promoter-Edited Cells, Related to Figure 6

(A) Part of the mutant reference genome containing two mutant motifs that correspond to two alleles in clone D3.

(B) Chromosome conformation capture (3C) assay in wild-type and γ-globin promoter edited cells. Results are shown as mean ± SEM of three experiments.

(C) Three biological replicates of BCL11A CUT&RUN showing the peaks in γ-globin gene region in wild-type HUDEP-2 cells and clone D3. The reads from D3 were mapped to the mutant genome. Note that in the mutant genome, *HBG1* carries the ΔC allele, and *HBG2* carries the Δ13bp allele. Thus the reads of Δ1bp alleles (ΔA+ΔC+ΔC) will be mapped to *HBG1*, and reads of Δ13bp allele will be mapped to *HBG2*.

(legend continued on next page)

(D) Zoomed in view showing the BCL11A CUT&RUN peaks on the *HBG1* promoter.

(E) ATAC-seq in wild-type, *BCL11A* knockout and γ-globin promoter edited cells.

(F) Left, RT-qPCR analysis of mRNA levels for γ-globin and β-globin in indicated cells. Right, percentages of β- or γ-globin mRNA. Results are shown as mean ± SEM of three experiments.

(G) Single locus footprint showing the γ-globin promoter region in HUDEP-2 and bulk edited cells. Bottom, sequence and percentage of each edited allele in bulk edited cells, determined by amplicon sequencing.