**Supplementary Information**

**Evaluation of human *BCL11A* enhancer in immortalized human erythroid precursors**

We utilized HUDEP-2 cells, an immortalized human CD34+ hematopoietic stem and progenitor cell (HSPC)-derived erythroid precursor cell line that expresses BCL11A and predominantly β- rather than γ-globin34. We used the CRISPR-Cas9 nuclease system to generate a clone of HUDEP-2 cells null for *BCL11A* by deleting coding sequences (Fig. 1a). These cells demonstrated elevated levels of γ-globin mRNA and HbF protein, consistent with the functional requirement of BCL11A for HbF repression (Fig. 1b, c). Deletion of the 12-kb *BCL11A* composite enhancer with a pair of sgRNAs resulted in near complete loss of BCL11A expression and induction of γ-globin and HbF protein to similar levels as cells with *BCL11A* knockout (Fig. 1a-c), analogous to the requirement of the orthologous mouse composite enhancer for erythroid BCL11A expression28. Introduction of Cas9 and an individual sgRNA targeting *BCL11A* exon-2 to HUDEP-2 cells as well as to primary human erythroid precursors produced cells with elevated HbF expression, indicating loss of BCL11A function and resultant derepression of BCL11A’s target γ-globin (Fig. 2g, h; Extended Data Fig. 2e). The level of BCL11A transcript was unaffected in these cells (Fig. 2f), suggesting that *BCL11A* transcripts with protein truncating variants (due to frameshift or nonsense mutations) escape nonsense-mediated decay.

**Pooled CRISPR enhancer saturating mutagenesis screen in the HUDEP-2 cells**

We designed all possible sgRNAs within the human *BCL11A* composite enhancer DHSs (Fig. 1d-f) as restricted only by the presence of the SpCas9 NGG protospacer adjacent motif (PAM)22,39. The NGG PAM restricted sgRNAs had a median adjacent genomic cleavage distance of 4 bp and 90th percentile of 18 bp (Fig. 1f), which suggested that this strategy could approach saturation mutagenesis *in situ* given the expected indel spectrum produced by each sgRNA to include frequent deletions up to 10 bp from the cleavage position22,23,36. The production of indels within an enhancer sequence could alter gene regulation by a variety of non-mutually exclusive mechanisms including: (1) disruption of transcription factor (TF) binding motifs, (2) creation of TF binding motifs, (3) changes in spacing of TF binding motifs without direct disruption of the binding sequences themselves, and (4) effects independent of TF binding motifs.

NAG may act as an alternate PAM for SpCas9, albeit with lower efficiency39. We also designed sgRNAs restricted by the NAG PAM (Extended Data Fig. 2a, b). The NAG PAM restricted sgRNAs had a median adjacent genomic cleavage distance of 5 bp and 90th percentile of 15 bp (Extended Data Fig. 2b).

We included 120 nontargeting sgRNAs as negative controls as well as 88 sgRNAs tiling exon-2 of *BCL11A* as positive controls. The total library included 1,338 sgRNAs (Extended Data Fig. 2c). We synthesized oligonucleotides for the NGG and NAG restricted and nontargeting sgRNAs on a microarray and cloned the sgRNAs as a pool to a lentiviral vector25. Deep sequencing of the lentiviral plasmid library (including both NGG and NAG sgRNAs) demonstrated that 1,337 of 1,338 sgRNAs (99.9%) were successfully cloned. The representation of sgRNAs within the library showed a relatively narrow distribution, with a median of 718 and the 10% and 90% percentile ranging from 337 to 1,205 normalized reads (Extended Data Fig. 2d). Since the number of viral integrants per cell follows a Poisson distribution, maximizing single integrants can be achieved by decreasing multiplicity of infection61–63. Therefore, the basic experimental schema was to transduce cells with the lentiviral library at low multiplicity such that nearly all selected cells contained a single integrant (Fig. 1d). We transduced HUDEP-2 cells stably expressing SpCas9 with the pooled library of *BCL11A* enhancer targeting sgRNAs. We initially expanded the cells for one week, and subsequently transferred them to erythroid differentiation conditions, for a total of two weeks of culture. Then we performed intracellular staining for HbF. Fluorescence activated cell sorting (FACS) was employed to isolate HbF-high and HbF-low pools (consistent with low and high BCL11A activity respectively; Fig. 1d; Extended Data Fig. 2e, g). We enumerated the representation of the library in each pool by deep sequencing. The HbF enrichment score of each sgRNA was calculated as the log2-ratio of normalized reads in the HbF-high compared to HbF-low pools. We compared the HbF enrichment of the 120 nontargeting negative control sgRNAs and 88 coding sequence targeted positive controls for both NGG and NAG PAM restricted sgRNAs. We observed equivalent representation of the nontargeting sgRNAs in the HbF-high and HbF-low pools but highly significant enrichment of the NGG sgRNAs targeting exon-2 of *BCL11A* in the HbF-high pool, consistent with a reduction of BCL11A activity (Extended Data Fig. 2h). One nontargeting sgRNA (#0548) had an HbF enrichment score of 0.803, while the remaining 119/120 nontargeting sgRNAs (99.2%) showed enrichment scores below 0.259. In contrast 40/48 sgRNAs targeting BCL11A exon 2 (83.3%) showed enrichment scores above 0.259. These results suggest that the large majority of sgRNAs in the library were competent to produce indels. However, exon-2 targeting sgRNAs with NAG PAM restriction did not show significant HbF enrichment suggesting inefficient indel production (Extended Data Fig. 2h). Therefore the NAG restricted sgRNAs were excluded from further analyses.

We compared the representation of sgRNAs in the initial plasmid pool to the representation of sgRNAs in the cells at the end of *in vitro* culture. While the majority of the library maintained neutral representation throughout the experiment, we observed a fraction of sgRNAs that were depleted, mainly among the h+62 sgRNAs (Extended Data Fig 2i). We observed that these dropout sgRNAs overlapped with repetitive elements within the genome, in particular to a SINE AluSq element that appears in the genome nearly 100,000 times64 (Extended Data Fig. 2k). Initial design of sgRNAs did not include prediction of off-target cleavage to maximize the resolution of target mutagenesis. We removed from subsequent analysis the 35 of 582 (6.0%) NGG PAM sgRNAs with cellular dropout from the plasmid pool greater than 8-fold, since these dropouts indicated apparent *BCL11A*-independent effects of genomic disruption (Extended Data Fig. 2i, k).

The majority of enhancer targeting sgRNAs showed no significant enrichment or depletion from the HbF-high pool (Extended Data Fig. 2j). We observed a number of sgRNAs with HbF enrichment at each of the DHSs as well as some with HbF depletion at h+55 (Extended Data Fig. 2j). We observed discrete sets of colocalizing sgRNAs with elevated HbF enrichment, with a particularly robust cluster at h+58 (Fig. 2a). Our screen data appear consistent with the effects of indels as produced by individual sgRNAs rather than combinations of sgRNAs for several reasons: (1) negative control sgRNAs did not show evidence of impact on BCL11A expression, arguing against the occurrence of prevalent passenger effects (Fig. 2a); (2) positive control sgRNAs targeting BCL11A coding sequences (in which the expected outcome of individual sgRNAs would be frameshifted null alleles) show that the great majority produce the expected effect of HbF derepression (Fig. 2a); (3) the enhancer targeting sgRNAs largely follow a null distribution of effect sizes with the exception of a few outliers suggesting most of the enhancer targeting sgRNAs had no impact on BCL11A expression again arguing against prevalent passenger effects (Extended Data 2j, 7b); and (4) the enhancer targeting outliers with positive enrichment map to colocalizing discrete genomic sequences (Fig. 2a, 3, 5a) which would appear unlikely if the impact on BCL11A expression were due to multifocal indels, large deletions, inversions, rearrangements, or translocations (the expected outcomes of more than one genomic cleavage) but likely if the impact were due to individual indels (the expected outcomes of single genomic cleavages).

A limitation of our primary pooled screen approach is that the measurements of the enrichment of sgRNAs come from a pooled population of cells that were transduced by the entire library. Therefore we endeavored to prospectively validate the results of individual sgRNAs as identified by the screen. We observed a strong correlation between the HbF enrichment score from the screen and the fraction of HbF+ cells in arrayed format, testing 24 sgRNAs with enrichment scores ranging from the highest to the lowest in the screen, and representing sgRNAs from all 5 mapping categories (*r* = 0.816, *p* < 0.0001; Extended Data Fig. 3a, b). These results demonstrate that a single enhancer-targeting sgRNA may mediate robust HbF induction.

**Common genetic variation in functional enhancer cores**

The h+62 Active region contains only one common SNP (MAF>1%), the variant rs1427407, which was previously identified by fine-mapping as the most highly trait-associated SNP28 (Fig. 3c; Extended Data Fig. 5c). The high-HbF T-allele is disruptive of an apparent half E-box/GATA composite motif (*P* = 9.74 x 10-4 for T-allele, *P* = 1.69 x 10-4 for G-allele, though neither met our predefined threshold for significance of *P* < 10-4) and associated with reduced GATA1 and TAL1 occupancy in primary human erythroid chromatin as previously described28. Multiple sgRNAs with cleavages mapping directly to the motif demonstrated positive enrichment scores (Extended Data Fig. 5c). Of note, there was a gap of 88 nucleotides between sgRNA cleavages at the core of the Active region due to lack of NGG PAM motifs. Despite this uncommon limitation of functional resolution by SpCas9 and NGG PAM restricted sgRNAs (Fig. 1f), the HMM model was still able to identify the region. Substantial interspecies conservation as evaluated by both PhyloP and PhastCons (which model individual nucleotide and multibase element conservation, respectively) was observed at this h+62 Active state region as compared to flanking regions (Fig. 3c, Extended Data 5c).

DHS h+55 encompasses the SNP rs7606173, which along with rs1427407 defines the most highly trait-associated haplotype (Fig. 3a; Extended Data Fig. 5a). Previous fine-mapping was unable to find additional SNPs at *BCL11A* with predictive power for the trait association beyond the rs1427407–rs7606173 haplotype based on conditional or rare-variant analyses. No common SNPs were found directly within the Active or Repressive state regions of h+55, however rs7606173 resides merely 3 bp from the Repressive region and 34 bp from the Active region (Extended Data Fig. 5a). The next closest common SNP to an Active or Repressive state within h+55 is rs62142646, which is 739 bp from an Active state. The major, ancestral G allele at rs7606173 is associated with high-HbF. The HUDEP-2 cells used in this screen are homozygous for this G variant. Given a model in which high-HbF trait is due to disruption of TF binding sequences at the *BCL11A* enhancer, sgRNA-mediated disruption of the high-HbF rs7606173-G allele might not be expected to lead to further functional impact. We did observe six motifs predicted (*P* < 10-4) to be differentially impacted by the rs7606173 genotype (Extended Data Fig. 5a). The top-scoring sgRNAs in h+55 cluster 56-58 bp from rs7606173, at a site with a predicted TAL1::GATA1 motif (*P* < 10-4). This sequence element possesses high vertebrate conservation (Extended Data Fig. 5a). The entire region encompassing the Active/Repressive h+55 states appears to have elevated sequence conservation as compared to flanking sequences (Fig. 3a).

The only common SNP within the h+58 Active region is rs6738440 just at the edge of the Active state region (chr2:60722241), 118 to 160 bp from the cluster of top-scoring sgRNAs (chr2:60722359-60722401; Fig. 4; Extended Fig. 5b); the next closest common SNP was rs62142615 (chr2:60722120), 119 bp away. Neither sgRNAs with significant adjacent enrichment nor overlying genome-scale significant motifs with either the major A- or minor G-allele were observed at rs6738440. Previous conditional analysis of the rs1427407-rs7606173 haplotype was unable to demonstrate residual significant trait association for this variant28.

Enhancers paradoxically may demonstrate both evolutionary conservation and heightened turnover. Common trait-associated enhancer variation suggests the frequent occurrence of intraspecies polymorphic sequences sufficient to modulate enhancer function and thereby produce novel phenotypes. Per above, we previously described that the trait-associated enhancer haplotype at *BCL11A* is defined by two SNPs28. Our pooled CRISPR screening revealed that each of these SNPs reside near functional enhancer states consistent with their roles as causal variants. The most potent enhancer region, within h+58, has no trait-associated variants near its functional core. This example demonstrates how fine-mapping GWAS associations to individual SNPs can substantially underestimate the biologic importance of the underlying elements to the associated trait.

**Pooled CRISPR enhancer saturating mutagenesis screen in mouse erythroid εy:mCherry reporter cells**

We generated a MEL cell reporter line with the mCherry fluorescent reporter knocked-in to the embryonic globin *Hbb-y* locus (Extended Data Fig. 6c). cells were cultured as described . Introduction of Cas9 and sgRNA targeting *Bcl11a* exon-2 resulted in the appearance of cells with elevated εy:mCherry expression, indicating derepression of the BCL11A target εy-globin (Extended Data Fig. 6h).

The mouse sgRNA library was comprised of both NGG and NAG PAM restricted sgRNAs. The library included sgRNA sets tiling the DHS m+55, m+58, and m+62 orthologs, as well as 120 nontargeting negative controls and 91 *Bcl11a* exon-2 targeting positive controls (Extended Data Fig. 6d, g). Similar to the human enhancer screen, the sgRNAs were distributed throughout the target sites, with a median distance to adjacent cleavage site of 4 bp and 90% of adjacent cleavage sites falling within 18 bp for NGG PAM restricted sgRNAs (Extended Data Fig. 6e). We successfully cloned into lentiviral plasmids all 1271 members of the library with a relatively narrow distribution of representation (median 735, 10%ile 393, 90%ile 1240 normalized reads; Extended Data Fig. 6f).

Following transduction at low multiplicity by the lentiviral library, and *in vitro* culture for two weeks, cells were sorted into high- and low-εy:mCherry pools (Extended Data Fig. 6i). Deep sequencing was performed of the genomic DNA to evaluate the representation of sgRNA libraries in the pools. The nontargeting negative control sgRNAs were evenly represented in the high- as compared to low-εy:mCherry pools whereas the positive control *Bcl11a* exon-2 targeting sgRNAs with NGG PAM were significantly overrepresented in the εy:mCherry-high pool (Extended Data Fig. 6j). Although there was slight enrichment that reached statistical significance, the NAG PAM restricted sgRNAs showed substantially reduced overrepresentation relative to the potent NGG restricted sgRNAs, so further analysis was restricted to the NGG PAM restricted sgRNAs (Extended Data Fig. 6j).

We analyzed the representation of the library in cells that had completed two weeks of *in vitro* culture (sum of the high- and low-εy:mCherry pools) as compared to the initial lentiviral plasmid pool. The large majority of sgRNAs showed equivalent representation in the initial plasmid pool and as integrants in cells at the completion of the experiment (Extended Data Fig. 7a). A small number of sgRNAs (n=8) showed substantial cellular dropout >2-3 and were removed from subsequent enrichment analysis. Similar to the human screen, these mapped to repetitive elements (Extended Data Fig. 7c).

We determined εy enrichment score as the log2-ratio between representation in the εy:mCherry-high as compared to εy:mCherry-low pools (Fig. 5a; Extended Data Fig. 7a). We noted almost all exon-2 targeting sgRNAs demonstrated both positive εy enrichment scores and negative cellular dropout scores with high correlation (Fig. 5a; Extended Data Fig. 7a, c, d).

The majority of enhancer targeting sgRNAs showed no significant εy enrichment (Extended Data Fig. 7b). We detected sgRNAs with both modest enrichment and depletion from the εy:mCherry-high pool at the m+55 ortholog, similar to h+55. We detected a set of sgRNAs with marked εy enrichment at the m+62 ortholog, exceeding the potency of those enriching at h+62. At the m+58 ortholog we did not observe any evidence of εy enriching or depleting sgRNAs (Extended Data Fig. 7b).

We applied the same HMM model to infer Active, Repressive, and Neutral states at the mouse BCL11A enhancer orthologs (Extended Data Fig. 4a, 8a-c). We identified an Active state at the m+62 ortholog and Active and Repressive states at the m+55 ortholog. Only the Neutral state was identified at the m+58 ortholog. The regions of the m+55 and m+62 DHSs with peak DNase I sensitivity were inferred as possessing Active states (Extended Data Fig 8a-c). We analyzed 108 clones in which the entire composite enhancer was first monoallelically deleted and subsequent hemizygous mutations were produced by sgRNAs targeting the m+62 ortholog on the remaining allele. We measured BCL11A expression by RT-qPCR in each of these 108 clones normalized to 25 control clones not exposed to m+62 targeting sgRNAs. This clonal analysis identified a core region of the m+62 ortholog containing functional sequences required for BCL11A expression and embryonic εy-globin repression (Extended Data Fig. 8d, 9). The region is rich with TF-binding motifs, particularly those of key factors involved in erythropoiesis and globin gene regulation, including Gata1, Klf1, and Myb (Extended Data Fig. 9). Of note, despite the presence of relatively high vertebrate conservation throughout the m+62 and h+62 Active state regions (Fig. 4c, Extended Data Fig. 5c, 8c, 9b), the impact of the m+62 ortholog on BCL11A and globin gene regulation greatly exceeded that of h+62 (Fig. 2a, c-e, 5a-c; Extended Data Fig. 8c, d, 9).

**Human and mouse sequence and functional conservation**

Sequence homology is detectable at an approximately similar distal intron-2 position with respect to the TSS for each of the mouse sequences homologous to the three human DHSs: h+55 (length 1283 bp) has 402 positions of nucleotide identity (31.3%) compared to the m+55 ortholog (length 1046 bp), h+58 (1264 bp) has 367 positions of nucleotide identity (28.6%) compared to the m+58 ortholog (length 1341 bp), and h+62 (length 1369 bp) has 281 positions of nucleotide identity (20.5%) compared to the m+62 ortholog (length 1216 bp). By comparison, of the 2508 bp in human *BCL11A* coding sequence (XL isoform), 2424 nucleotides demonstrate identity (96.7%) compared to mouse *Bcl11a* coding sequence (XL isoform).

Although enhancers are composed of TF binding motifs, the presence of motifs alone is inadequate to predict enhancers. Motif predictions can be overly sensitive, in that only a small fraction of predicted motifs tend to be corroborated by ChIP-seq occupancy studies. On the other hand, motif prediction can also be insensitive; for example, a recent report highlights the importance of low-affinity motifs for achieving specificity of enhancer function65. Previously we showed that GATA1 occupies h+58 in primary human erythroid precursors28. However the orthologous m+58 region possesses neither DNase sensitivity nor functional requirement in mouse erythroid cells. Despite this divergence, the human core GATA1 motif has a similar *P*-value in the nonfunctional mouse ortholog. These results are consistent with a model in which the motif context is critically important in enhancer activity. The sequences immediately adjacent to the GATA1 motif, where both HbF-associated sgRNAs and mutations enrich, are candidates to fulfill this contextual requirement.

Recent appreciation for the wide variation in intensity of biochemical features associated with enhancer elements has led to a renewed interest in clustered enhancer elements and so-called super-enhancers. Based on H3K27ac ChIP-seq in primary human adult erythroid precursors, the composite *BCL11A* enhancer scores as a human erythroid super-enhancer (Extended Data Fig. 1a, b). We used published H3K27ac ChIP-seq data from mouse erythroid cells and found variable results, in one dataset56 *Bcl11a* scores as a mouse erythroid super-enhancer whereas in another dataset55 it ranks below the super-enhancer threshold (Extended Data Fig. 6a, b). Here we provide an example of a super-enhancer organized as a hierarchy of constituent DHSs, with some critical and others minimally required for gene expression. Even within a critical substituent DHS such as *BCL11A* h+58, there are many dispensable and only a few critical sequences. These experiments show how a super-enhancer may be vulnerable to indels produced by single DSBs.

**Supplementary References**

51. Chen, S. *et al.* Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis Resource Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis. *Cell* **160,** 1–15 (2015).

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

53. Giarratana, M. *et al.* Proof of principle for transfusion of in vitro generated red blood cells. *Blood* **118,** 5071–5079 (2011).

54. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 1–8 (2014).

55. Kowalczyk, M. S. *et al.* Intragenic Enhancers Act as Alternative Promoters. *Mol. Cell* **45,** 447–458 (2012).

56. Dogan, N. *et al.* Occupancy by key transcription factors is a more accurate predictor of enhancer activity than histone modifications or chromatin accessibility. *Epigenetics Chromatin* **8,** 1–21 (2015).

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

58. Mathelier, A. *et al.* JASPAR 2014: An extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **42,** 142–147 (2014).

59. Weber, K., Bartsch, U., Stocking, C. & Fehse, B. A multicolor panel of novel lentiviral ‘gene ontology’ (LeGO) vectors for functional gene analysis. *Mol. Ther.* **16,** 698–706 (2008).

60. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32,** (2014).

61. Ellis, E. L. & Delbrück, M. The Growth of Bacteriophage. *J. Gen. Physiol.* **22,** 365–384 (1939).

62. Stent, G. *Molecular Biology of Bacterial Viruses*. (W. H. Freeman and Company, 1963).

63. Choi, C., Kuatsjah, E., Wu, E. & Yuan, S. The Effect of Cell Size on the Burst Size of T4 Bacteriophage Infections of Escherichia coli B23. *J. Exp. Microbiol. Immunol.* **14,** 85–91 (2010).

64. Cui, F., Sirotin, M. V & Zhurkin, V. B. Impact of Alu repeats on the evolution of human p53 binding sites. *Biol. Direct* **6,** 2 (2011).

65. Crocker, J. *et al.* Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness. *Cell* **160,** 191–203 (2015).