

BCL11A enhancer dissection by Cas9-mediated *in situ* saturating mutagenesis

Matthew C. Canver^{1*}, Elenoe C. Smith^{1*}, Falak Sher^{1*}, Luca Pinello^{2*}, Neville E. Sanjana^{3*}, Ophir Shalem³, Diane D. Chen¹, Patrick G. Schupp¹, Divya S. Vinjamur¹, Sara P. Garcia², Sidinh Luc¹, Ryo Kurita⁴, Yukio Nakamura^{4,5}, Yuko Fujiwara^{1,6}, Takahiro Maeda⁷, Guo-Cheng Yuan², Feng Zhang^{3§}, Stuart H. Orkin^{1,6§} & Daniel E. Bauer^{1§}

Enhancers, critical determinants of cellular identity, are commonly recognized by correlative chromatin marks and gain-of-function potential, although only loss-of-function studies can demonstrate their requirement in the native genomic context. Previously, we identified an erythroid enhancer of human *BCL11A*, subject to common genetic variation associated with the fetal haemoglobin level, the mouse orthologue of which is necessary for erythroid *BCL11A* expression. Here we develop pooled clustered regularly interspaced palindromic repeat (CRISPR)-Cas9 guide RNA libraries to perform *in situ* saturating mutagenesis of the human and mouse enhancers. This approach reveals critical minimal features and discrete vulnerabilities of these enhancers. Despite conserved function of the composite enhancers, their architecture diverges. The crucial human sequences appear to be primate-specific. Through editing of primary human progenitors and mouse transgenesis, we validate the *BCL11A* erythroid enhancer as a target for fetal haemoglobin reinduction. The detailed enhancer map will inform therapeutic genome editing, and the screening approach described here is generally applicable to functional interrogation of non-coding genomic elements.

Enhancers are classically described as distal genetic elements that positively regulate gene expression in an orientation-independent manner in ectopic heterologous gain-of-function expression experiments¹. These elements coordinate when, where and how genes are expressed. Enhancer sequences bind transcription factors and are correlated with specific chromatin features including reduced DNA methylation, characteristic histone modifications, heightened chromatin accessibility, long-range promoter interactions, and bidirectional transcription. Recent chromatin mapping has demonstrated the abundance of distal regulatory elements bearing an enhancer signature²⁻⁴.

The biological importance of enhancers is underscored by gene expression studies showing the predictive power of enhancer profile on lineage-specific programs⁵⁻⁷. Highly marked and clustered enhancers (for example, so-called strong, stretch, or super-enhancers) are particularly suggestive of cellular identity and may help to infer lineage-specific regulatory factors⁸⁻¹⁰. Genome-wide association studies reveal enrichment of trait-associated variants in sequences bearing lineage-restricted enhancer signatures^{4,8,11,12}. Enhancers display signs of evolutionary constraint as well as heightened turnover with evidence of positive selection¹³⁻¹⁶.

Despite their importance, enhancers are typically defined by criteria unrelated to *in situ* functional requirement. Advances in putative enhancer mapping, as well as large-scale oligonucleotide synthesis, facilitate enhancer reporter assays on a massively parallel scale, allowing a systematic evaluation of the functional significance of enhancer sequences^{17,18}. Nonetheless, ectopic heterologous enhancer assays cannot address the necessity of an element in its native chromatin

environment. The growing appreciation of the nonrandom distribution of distal elements both with respect to the linear genome and within the three-dimensional nuclear environment emphasizes the importance of studying enhancers by perturbing their endogenous condition^{10,19}.

Insightful observations have been made by mutagenizing enhancers using traditional molecular genetic approaches^{20,21}; however, the low throughput of these classical methods constrains their widespread application. Furthermore, the elevated turnover of many enhancer sequences between species may limit the ability to derive conclusions from nonhuman organisms regarding human gene regulation. Advances in genome editing technology make practical the facile modification of the human genome^{22,23}. High-throughput Cas9-mediated functional genomics studies have revealed novel genes required for various biological processes²⁴⁻²⁷. Genome editing is likewise suitable for the study of non-coding genetic elements such as enhancers, although these experiments have previously been conducted at low throughput²⁸⁻³⁰.

Human composite enhancer

Recently, we observed that common genetic variants associated with fetal haemoglobin (HbF; $\alpha_2\gamma_2$) level and β -haemoglobin disorder clinical severity mark an adult-developmental-stage- and erythroid-lineage-specific intronic enhancer of *BCL11A* (ref. 28), a validated repressor of HbF and therapeutic target for β -haemoglobin disorders^{28,31-33}. This composite human enhancer is composed of three DNase I hypersensitive sites (DHSs), termed h+55, h+58 and h+62 on the basis of distance in kilobases (kb) from the transcriptional start site (TSS)²⁸. The most highly

¹Division of Hematology/Oncology, Boston Children's Hospital, Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts 02115, USA. ³Broad Institute of MIT and Harvard, McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences and Department of Biological Engineering, MIT, Cambridge, Massachusetts 02142, USA. ⁴Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Ibaraki 305-0074, Japan. ⁵Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan. ⁶Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA. ⁷Division of Hematology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

trait-associated haplotype is defined by two single nucleotide polymorphisms (SNPs): rs1427407 within h+62 and rs7606173 within h+55 (Extended Data Fig. 1a). Previously, we showed that this enhancer possessed ectopic erythroid-restricted, adult-stage-specific enhancer activity²⁸. Moreover, the mouse orthologue of the composite enhancer—defined by primary sequence homology, shared erythroid enhancer chromatin signature and syntenic position relative to coding sequences—was shown to be required for *BCL11A* expression and embryonic globin gene repression in a mouse erythroid cell line but dispensable in a mouse B-lymphoid cell line²⁸.

To evaluate the requirement for human *BCL11A* enhancer sequences, we used HUDEP-2 cells, an immortalized human CD34⁺ haematopoietic 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 clones of HUDEP-2 cells with deletion of the 12-kb *BCL11A* composite enhancer by introduction of a pair of chimaeric single guide RNAs (sgRNAs). Enhancer deletion 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), consistent with the possibility that these sequences could serve as targets for therapeutic genome editing for HbF reinduction for the β -haemoglobinopathies³⁵. Although targeted deletions by paired double strand breaks (DSBs) may be achieved by genome editing, competing genomic outcomes include local insertion/deletion (indel) production at each cleavage site as well as inversion of the intervening segment^{22,23,36–38}.

Tiled pooled enhancer editing *in situ*

We hypothesized that composite enhancers may be composed of a functional hierarchy with essential and dispensable constituent components. A functional hierarchy might enable enhancer disruption by a single DSB at a critical region followed by non-homologous end joining (NHEJ) repair with indels. In fact, the hypothesis that a prevalent mechanism of trait associations is enhancer variation rests on the premise that single nucleotide changes themselves may substantively modulate enhancer function. Therefore, we reasoned that a tiling set of sgRNAs could uncover critical enhancer regions by disruption of nearly all sequences within an enhancer on the basis of the typical outcome of Cas9 cleavage and NHEJ repair, an indel spectrum with frequent deletions of up to 10 base pairs (bp) from the cleavage position^{22,23,36,38,39}.

We designed all possible sgRNAs within the human *BCL11A* composite enhancer DHSs (Fig. 1d, e) as restricted only by the presence of the *Streptococcus pyogenes* Cas9 (SpCas9) NGG protospacer adjacent motif (PAM) sequence, which restricts cleavage at an average 1/8 frequency at each genomic position^{22,39}. The NGG-PAM-restricted sgRNAs had a median gap between adjacent genomic cleavages of 4 bp and 90th percentile of 18 bp (Fig. 1f), which suggested that this strategy could approach saturation mutagenesis *in situ*. We included non-targeting sgRNAs as negative controls as well as sgRNAs tiling exon 2 of *BCL11A* as positive controls (Fig. 1e). The library was successfully cloned to a lentiviral vector. The basic experimental schema was to transduce HUDEP-2 cells with the lentiviral library at low multiplicity such that nearly all selected cells contained a single integrant (Fig. 1d). After expansion, differentiation, sorting by HbF level, genomic DNA isolation and deep sequencing of integrated sgRNAs, an HbF enrichment score was calculated for each sgRNA by comparing its representation in HbF-high and HbF-low pools (see Supplementary Information and Extended Data Fig. 2 for additional technical details).

We mapped the HbF enrichment score of each sgRNA to its predicted position of genomic cleavage (Fig. 2a). Most enhancer targeting sgRNAs showed no significant enrichment or depletion from the HbF-high pool. The enriching sgRNAs co-localized to discrete genomic positions. For example, we observed a cluster of sgRNAs at h+62 with modest enrichment, a cluster at h+55 with moderate

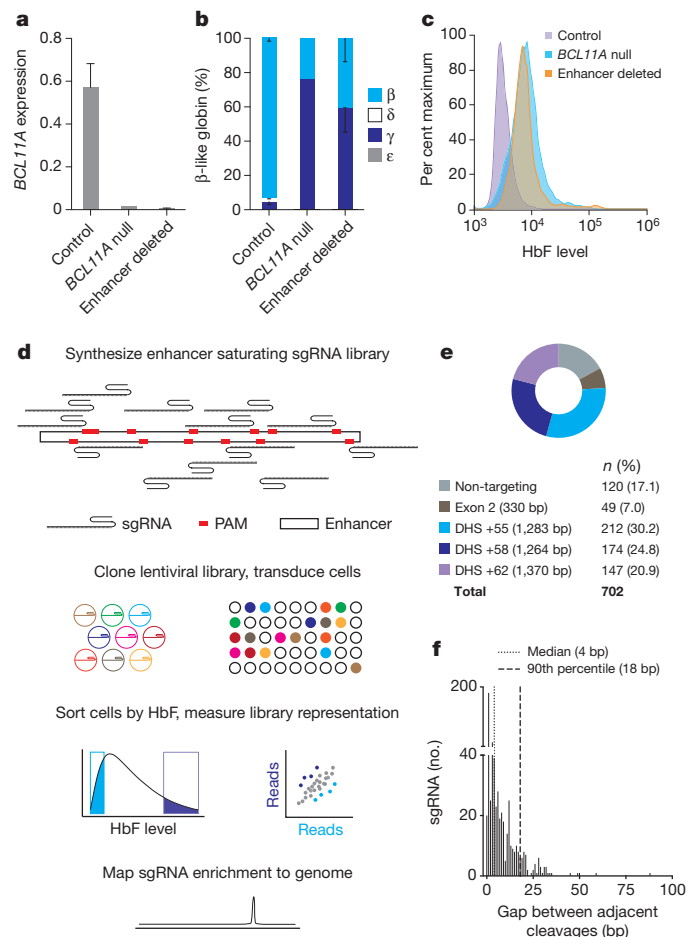


Figure 1 | Tiled pooled *in situ* CRISPR-Cas9 *BCL11A* enhancer screen. a–c, Deletion of the human composite *BCL11A* enhancer in HUDEP-2 cells demonstrates its necessity for *BCL11A* expression (normalized to GAPDH), repression of γ -globin mRNA, and repression of HbF; control clones, $n = 4$; *BCL11A* null, $n = 1$; enhancer deleted, $n = 3$; error bars show s.e.m. d, Workflow of CRISPR-Cas9 enhancer screen showing library synthesis, delivery and analysis. e, Human NGG PAM sgRNA library distribution. f, Gaps between adjacent genomic cleavages for NGG PAM sgRNAs targeting *BCL11A* exon 2, h+55, h+58 and h+62.

enrichment (as well as adjacent clusters with depletion), and a cluster at h+58 with marked enrichment. Of note, we observed ten sgRNAs at h+58 with cleavage positions within 42 bp each with HbF enrichment scores exceeding 0.99, the median enrichment score of *BCL11A* exon-2-targeting sgRNAs.

Exon-2-targeted sgRNAs showed a linear correlation between HbF enrichment and cellular dropout, suggesting that sgRNAs that result in complete knockout of *BCL11A* lead to a reduced rate of cell accumulation inseparable from HbF derepression (Fig. 2b). In contrast, the sgRNAs at h+58 associated with marked HbF enrichment showed blunted impact on dropout (Fig. 2b). This finding could be consistent with a low residual level of *BCL11A* adequate to promote cellular accumulation but inadequate to suppress HbF.

To corroborate these findings, we introduced two sgRNAs to the HUDEP-2/Cas9 cells to produce targeted deletion or inversion clones³⁶. Deletion of h+58 phenocopied deletion of the composite enhancer and deletion of h+55 had moderate effect (while deletion of h+62 showed a non-significant trend towards a modest effect), consistent with the magnitude of top-scoring and co-localizing sgRNAs from the screen (Fig. 2a, c–e). Inversion of the h+58 or h+55 sites had no significant effect on gene expression, demonstrating that the *BCL11A* enhancer functions in an orientation-independent manner *in situ*, consistent with the classic enhancer definition¹ (Fig. 2c–e).

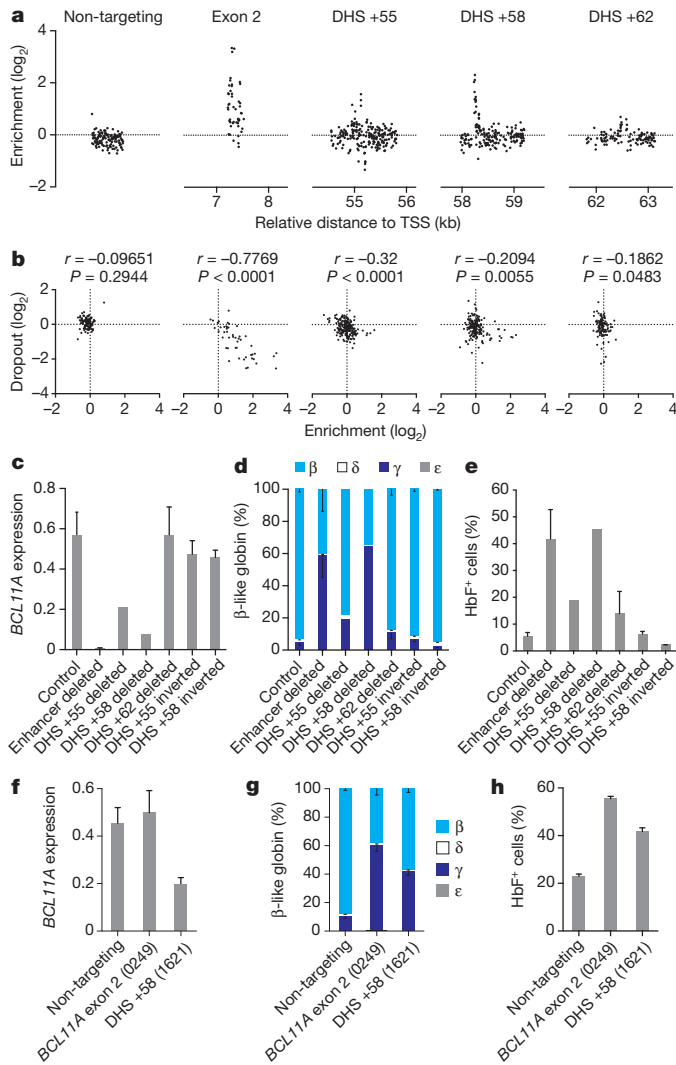


Figure 2 | Functional mapping of the *BCL11A* enhancer. **a**, Mapping sgRNA HbF enrichment scores relative to genomic cleavage positions. Non-targeting sgRNAs are pseudo-mapped with 5-bp spacing. **b**, Correlation between cellular dropout and HbF enrichment scores. **c–e**, *BCL11A* expression normalized to GAPDH (**c**), β -like globin expression (**d**), and HbF^+ fraction (**e**) in HUDEP-2 cells with deletion or inversion of individual DHSs; control clones, $n = 4$; enhancer deleted, $n = 3$; +55 deleted, $n = 1$; +58 deleted, $n = 1$; +62 deleted, $n = 5$; +55 inverted, $n = 3$; +58 inverted, $n = 2$. **f–h**, *BCL11A* expression normalized to GAPDH (**f**), β -like globin expression (**g**), and HbF^+ fraction (**h**) in primary human erythroid precursors transduced with Cas9 and individual sgRNAs; $n = 3$. Error bars represent s.e.m. (**c, d, f, g**) or s.d. (**e, h**).

To validate the findings from the HUDEP-2 cells, the top-scoring enhancer-targeting sgRNA from the screen (number 1621 at h+58) was tested in primary human erythroblasts by lentiviral transduction of human $CD34^+$ HSPCs exposed to *ex vivo* erythroid culture conditions. Consistent with the screen results, sgRNA-1621 resulted in downregulation of *BCL11A* expression and corresponding upregulation of γ -globin expression and increase in HbF^+ cells (Fig. 2f–h). Notably, sgRNA-1621 did not alter surface marker profile, enucleation frequency, or cellular morphology (Extended Data Fig. 3c). Together, these results suggest proof-of-principle of an individual sgRNA targeting a non-coding element for therapeutic genome editing of β -haemoglobin disorders.

Primate-specific enhancer sequences

We applied a hidden Markov model (HMM) to the sgRNA enrichment score data to infer functionally important sequences within each

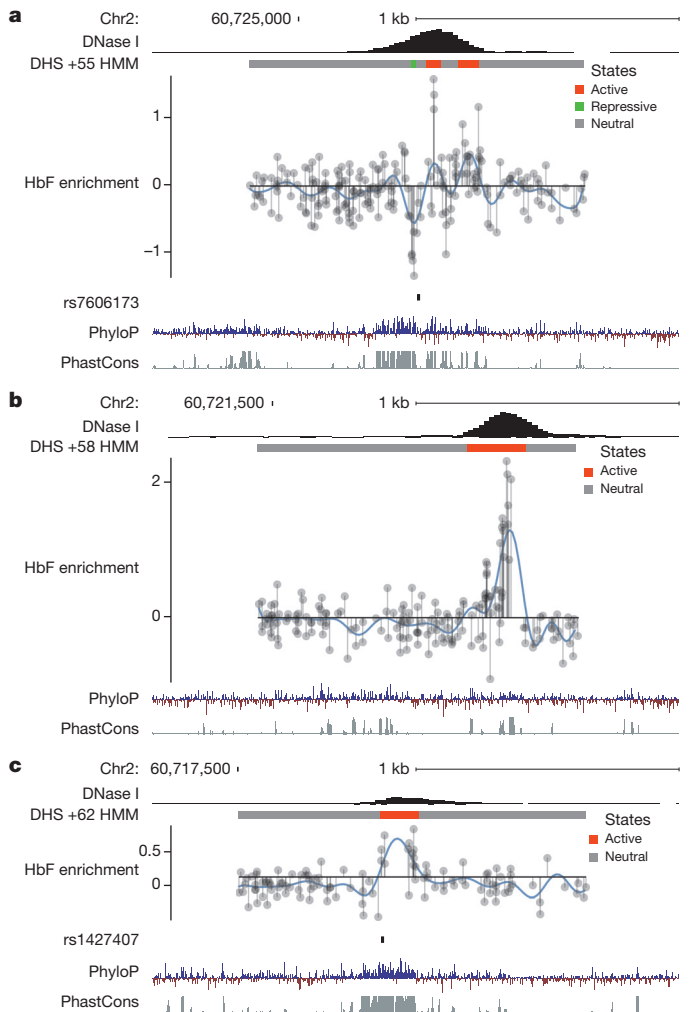
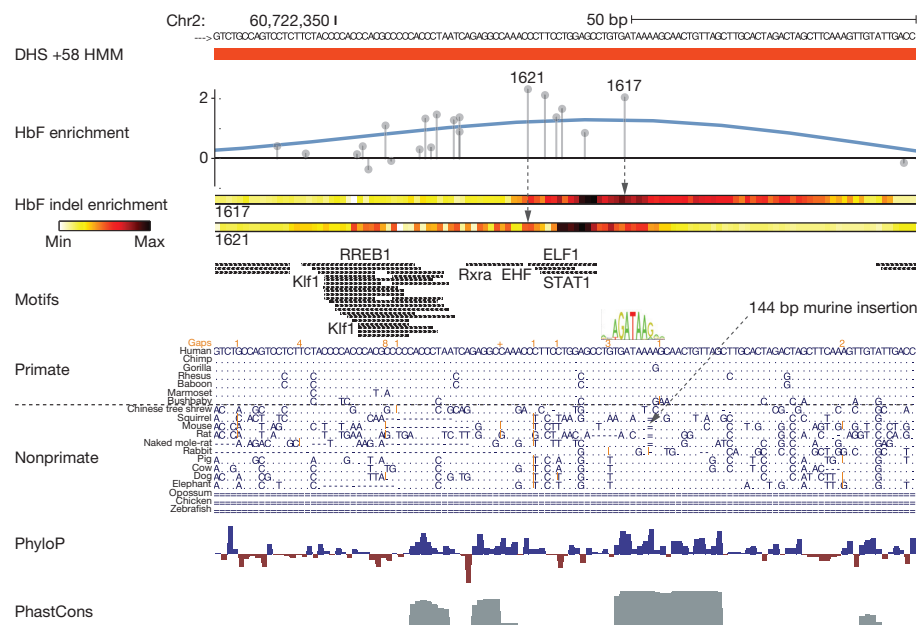


Figure 3 | Inferred functional enhancer states relative to genomic features. **a–c**, HMM segmentation of functional enhancer states. HbF enrichment scores are shown throughout DHSs h+55, h+58 and h+62 by grey lines and circles, with the blue line representing smoothed enrichment score. DNase I sequencing from primary human erythroblasts²⁸. PhyloP (scale from -4.5 to 4.88) and PhastCons (from 0 to 1) estimates of evolutionary conservation among 100 vertebrates are shown. Positions of SNPs rs7606173 and rs1427407 are denoted, which together define the haplotype most highly associated with HbF level²⁸.

DHS (Extended Data Fig. 4a). This model defined three functional states—active, repressive and neutral—based on likelihood to encompass sequences that positively, negatively and neutrally regulate target gene expression, respectively. The model identified functional states within each DHS (Fig. 3a–c). At each of the three DHSs, the active states were precisely located at regions with the highest degree of DNase I sensitivity.

The overall sequence conservation at the h+58 active region appears both less intense and less distinct from flanking sequences as compared to those of h+62 and h+55 (Fig. 3a–c). The top-scoring sgRNAs in the screen co-localize to 42 bp within h+58 (Fig. 4 and Extended Data Fig. 5b). The third-highest-scoring enhancer-targeted sgRNA (sgRNA-1617) mapped directly onto an apparent GATA1 motif, although below a genome-scale significance threshold ($P = 3.74 \times 10^{-4}$). The mouse orthologous sequence has a GATA1 motif P value only modestly higher than has the human ($P = 4.33 \times 10^{-4}$). This GATA1 motif appears to have relatively high vertebrate conservation, with exact human sequence identity in rabbits, pigs, dogs and elephants. The top-scoring sgRNA (sgRNA-1621) mapped to a position 15 bp from this GATA1 motif (Fig. 4). An additional four sgRNAs mapping between sgRNA-1621

Figure 4 | Primate-specific *BCL11A* enhancer functional core. DHS h+58 functional core defined by maximal HbF enrichment score and active HMM state. HbF enrichment scores are shown by grey lines and circles. HbF indel enrichment per nucleotide is based on amplicon genomic sequencing of sorted cells exposed to either sgRNA-1617 or -1621. No common SNPs (minor allele frequency >1%) are present at this region. JASPAR motifs ($P < 10^{-4}$) are depicted in black with selected motifs annotated by transcription factor based on known erythroid-specific function or genomic position. Gata1 motif LOGO at sgRNA-1617 cleavage position as described in text. Orthologous sequences are listed from representative primates and nonprimates of distributed phylogeny. PhyloP (scale from -4.5 to 4.88) and PhastCons (from 0 to 1) estimates of evolutionary conservation among 100 vertebrates are shown.



and sgRNA-1617 each had substantially elevated HbF enrichment scores. Underlying these sgRNAs were additional predicted motifs (that is, RXRA, EHF, ELF1 and STAT1). Although these sequences showed a high level of conservation among primates, they showed high degeneracy among nonprimate vertebrates (Fig. 4).

We tested the pattern of mutations observed upon treatment of cells with either sgRNA-1621 or sgRNA-1617 by deep sequencing. Each of these sgRNAs is sufficient to substantially induce HbF in human erythroid cells (Fig. 2h and Extended Data Fig. 3a, b). We sorted cells exposed to Cas9 and these sgRNAs into HbF-high and HbF-low pools. We determined the indel spectrum in each population by deep sequencing (Extended Data Fig. 4b). As expected, we observed indels clustering around the predicted cleavage positions. By comparing the per nucleotide indel ratio between cells from the HbF-high and HbF-low pools, we calculated a relative indel enrichment across the sequencing amplicon. Notably, both sgRNAs yielded maximal HbF indel enrichment not precisely at the expected cleavage position but offset at shared intervening sequences (Fig. 4). These sites of maximal HbF mutation enrichment mapped to 7 bp directly overlapping predicted motifs (Fig. 4). Taken together, these data suggest that a conserved GATA1 motif scoring below the prediction threshold adjacent to primate-specific sequences form the core of an enhancer essential for human erythroid *BCL11A* expression and HbF repression.

Mouse enhancer dissection

To test functional conservation of the *BCL11A* enhancer, we examined the orthologous mouse *Bcl11a* enhancer in greater detail. Erythroid DNase I sensitivity is observed at those sequences homologous to h+55 and h+62 but not h+58 (Extended Data Fig. 6a), consistent with the reduced sequence homology within the h+58 active region (Fig. 3a–c). We performed a pooled CRISPR enhancer saturating mutagenesis screen in MEL $\epsilon\gamma$:mCherry reporter cells, similar to the human screen described above (Extended Data Figs 6 and 7, and Supplementary Information).

Upon mapping the sgRNA cleavage positions to the genome, we again observed that the majority of enhancer-targeting sgRNAs demonstrated no significant $\epsilon\gamma$ enrichment or depletion. We observed co-localization of sets of sgRNAs with $\epsilon\gamma$ enrichment (Fig. 5a). There was a similar complex pattern at the m+55 orthologue as at h+55, with adjacent regions with enriching and depleting sgRNAs from the high- $\epsilon\gamma$:mCherry pool at the DHS core. At the m+58 orthologue we

did not observe any evidence of $\epsilon\gamma$ enriching or depleting sgRNAs. At the m+62 orthologue there was a marked peak, with five sgRNAs with $\epsilon\gamma$ enrichment scores exceeding 1.30, the median enrichment score of *Bcl11a* exon-2-targeting sgRNAs (Fig. 5a). This potent impact of the m+62 orthologue was in contrast to the modest impact of individual sgRNAs or DHS deletion at h+62.

We used pairs of sgRNAs in the presence of Cas9 to produce MEL clones with deletions of various substituent elements at the *Bcl11a* enhancer (Fig. 5b). Deletion of the DNase-insensitive m+58 orthologue had no apparent effect on *Bcl11a* expression, consistent with the pooled screen result. Deletion of the m+55 orthologue led to an approximately twofold reduction in *Bcl11a* expression (mean

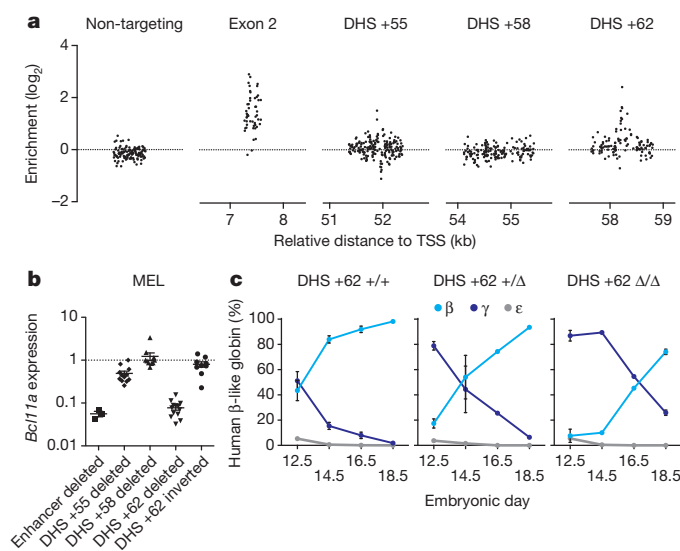


Figure 5 | Functional sequence requirement at the mouse *Bcl11a* erythroid enhancer for *in vivo* haemoglobin switching. **a**, Mapping sgRNA $\epsilon\gamma$ enrichment scores to genomic cleavage positions. Non-targeting sgRNAs were pseudo-mapped with 5 bp spacing. **b**, *Bcl11a* expression in mouse erythroid clones with deletion or inversion of individual DHSs relative to non-deleted controls. **c**, Transgenic human β -like globin expression in β -YAC/+62 deletion mice. For +/+ , +/ Δ and Δ/Δ : at E12.5, $n = 5, 11$ and 3 embryos, respectively; at E14.5, $n = 2, 3$ and 4; at E16.5, $n = 2, 4$ and 1; at E18.5, $n = 3, 1$ and 3. Error bars represent s.e.m.

residual level 49%, $P < 0.0001$), whereas deletion of the m+62 orthologue approached deletion of the entire composite enhancer in terms of reduction in *Bcl11a* expression (mean residual levels of 8% ($P < 0.0001$) and 6% ($P < 0.0001$), respectively, Fig. 5b; also see Supplementary Information and Extended Data Figs 8 and 9). In addition, clones in which the m+62 orthologue was inverted showed no change in *Bcl11a* expression, suggesting that the mouse enhancer, like the human enhancer, functions independently of orientation *in situ* (Figs 2c–e and 5b).

Erythroid-restricted function *in vivo*

To substantiate the importance of the m+62 orthologue for *Bcl11a* expression and to validate *BCL11A* enhancer disruption as a therapeutic strategy, we generated mice deficient for the *Bcl11a* m+62 orthologue. We used the same Cas9 and paired sgRNA deletion strategy in mouse embryonic stem cells, from which we derived mice with germline transmission of the enhancer deletion. Previous studies have demonstrated essential roles for *Bcl11a* in structural development of the central nervous system and B lymphopoiesis^{40–42}. Notably, unlike conventional *Bcl11a* knockouts, which die hours after birth, m+62 orthologue deletion mice were born healthy at expected Mendelian ratios (Extended Data Fig. 10a). The m+62 orthologue deletion mice also showed normal frequencies of B-cell progenitors in the fetal liver and mature B lymphocytes in the adult peripheral blood (Extended Data Fig. 10b, c). Other haematopoietic lineages were also observed at wild-type frequencies (Extended Data Fig. 10c). *Bcl11a* expression was unperturbed in the brain or sorted B cell precursors from embryonic day 16.5 (E16.5) embryos (Extended Data Fig. 10d). In contrast, there was substantial reduction in *Bcl11a* levels in sorted E16.5 erythroid precursors (26% residual, $P < 0.05$; Extended Data Fig. 10d).

The m+62 orthologue deletion mice were bred to mice transgenic for the human β -globin cluster (β -YAC) to model the role of *BCL11A* in haemoglobin switching⁴³. Unlike its fetal-stage expression in humans, in the mouse fetal liver transgenic human γ -globin is subject to intense repression (like an embryonic globin). *Bcl11a* is required for this early murine silencing of transgenic γ -globin at E14.5, although even in the absence of *Bcl11a*, γ -globin is ultimately repressed^{32,33}. Fetal livers were evaluated between days E12.5 and E18.5 to monitor haemoglobin switching. Repression of human γ -globin and activation of human β -globin was markedly delayed in the m+62 orthologue deleted mice (Fig. 5c). Heterozygous mice showed an intermediate γ -globin derepression phenotype, underscoring the dose-dependent inverse relationship between *BCL11A* and HbF level. These results indicate that targeting the erythroid enhancer of *Bcl11a* *in vivo* results in erythroid-specific disruption of *Bcl11a* expression and relaxed repression of γ -globin, unaccompanied by the obvious neurological and immunological toxicities seen in the *Bcl11a* conventional knockout context.

Discussion

We employed a novel application of CRISPR-Cas9 genome editing, saturating mutagenesis of non-coding elements *in situ*, to provide an important insight into the organization and function of the *BCL11A* erythroid enhancer. Traditional tests of enhancer function rely on ectopic heterologous reporter assays and/or correlative biochemical features. Genome editing allows facile evaluation of the requirement of enhancer sequences within their endogenous chromatin context for appropriate gene regulation. As shown here, high-resolution, high-throughput pooled tiling sgRNA screening reveals underlying enhancer sequence requirements approaching nucleotide resolution. A limitation to the resolution of this approach is the availability of NGG PAM sequences in a given region. We did not observe efficient editing by SpCas9 with NAG-restricted sgRNAs (Extended Data Figs 2h and 6j). Recent studies have identified Cas9 orthologues and variants restricted by alternative PAM sequences, each capable of efficient genome editing^{44–46}. This increased targeting range of Cas9

could allow increased resolution of *in situ* mutagenesis, particularly at sequences with paucity of NGG motifs. Alternatively, approaches reliant on homology-directed repair⁴⁷ could offer nucleotide-resolution functional mutagenesis of non-coding sequences, although issues of efficiency, fidelity and quantitative sensitivity would need to be considered. We suggest that our tiled pooled CRISPR screening approach could be readily adapted to the functional interrogation of numerous non-coding genomic elements.

In addition, these data demonstrate that apparent sequence conservation at the *BCL11A* enhancer masks underlying functional divergence. The mouse and human *BCL11A* erythroid composite enhancers share primary sequence homology, an erythroid enhancer chromatin signature, and syntenic intronic position relative to coding sequences. Moreover, both are required for erythroid expression of *BCL11A* and repression of embryonic/fetal globin genes. However, our high-resolution CRISPR mutagenesis analysis reveals divergence in the architecture of these enhancers. Of note, human *BCL11A* enforces the γ - to β -globin developmental switch around the time of birth. The timing and nature of these switches and the globin genes themselves are distinct in primates as compared to nonprimate vertebrates that only exhibit a mid-gestation embryonic to adult switch⁴⁸. Therefore, it would seem plausible that critical regulatory mechanisms at *BCL11A* might differ between species (also see Supplementary Information).

The haemoglobin disorders are among the most common Mendelian inherited human conditions. The level of HbF is a key modifier of clinical severity of these diseases and *BCL11A* is the chief regulator of HbF level⁴⁸. Naturally occurring genetic variation at the *BCL11A* enhancer is well-tolerated and associated with HbF level and β -haemoglobin disorder clinical severity. The work presented here offers a framework for therapeutic genome editing of the *BCL11A* enhancer for β -haemoglobin disorders. Enhancer disruption by individual sgRNAs in primary erythroid precursors results in substantial HbF induction. This approach may mitigate erythroid-specific growth disadvantages of complete *BCL11A* loss (Fig. 2b). Furthermore, erythroid enhancer disruption may spare *BCL11A* expression and function in non-erythroid contexts, such as B lymphopoiesis (Extended Data Fig. 10b–d). A challenge for the field is that it is not yet possible to accurately model HbF repression experimentally. However, individuals haploinsufficient for *BCL11A* due to microdeletions exhibit marked neurological deficits, and elevated HbF beyond that seen in homozygotes for high-HbF common enhancer haplotypes^{49,50}. Taken together, these data suggest that perturbation of critical sequences within the *BCL11A* enhancer defined here may result in HbF levels exceeding a clinical threshold required to ameliorate the β -haemoglobin disorders.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 April; accepted 25 August 2015.

Published online 16 September 2015.

- Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308 (1981).
- Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
- Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **488**, 75–82 (2012).
- Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
- Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
- Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
- Xu, J. *et al.* Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. *Dev. Cell* **23**, 796–811 (2012).
- Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).

9. Parker, S. C. J. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl Acad. Sci. USA* **110**, 17921–17926 (2013).
10. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
11. Paul, D. S. *et al.* Maps of open chromatin guide the functional follow-up of genome-wide association signals: Application to hematological traits. *PLoS Genet.* **7**, e1002139 (2011).
12. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
13. Hardison, R. C. Variable evolutionary signatures at the heart of enhancers. *Nature Genet.* **42**, 734–735 (2010).
14. Vierstra, J. *et al.* Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346**, 1007–1012 (2014).
15. Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015).
16. Pennacchio, L. A. *et al.* *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
17. Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature Biotechnol.* **30**, 271–277 (2012).
18. Patwardhan, R. P. *et al.* Massively parallel functional dissection of mammalian enhancers *in vivo*. *Nature Biotechnol.* **30**, 265–270 (2012).
19. Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049–1059 (2015).
20. Bender, M., Bulger, M., Close, J. & Groudine, M. β -globin gene switching and DNase I sensitivity of the endogenous β -globin locus in mice do not require the locus control region. *Mol. Cell* **5**, 387–393 (2000).
21. Johnson, K. D. *et al.* Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity. *J. Clin. Invest.* **122**, 3692–3704 (2012).
22. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
23. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
24. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
25. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
26. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotechnol.* **32**, 267–273 (2014).
27. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).
28. Bauer, D. E. *et al.* An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level. *Science* **342**, 253–257 (2013).
29. Gröschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in Leukemia. *Cell* **157**, 369–381 (2014).
30. Mansour, M. R. *et al.* An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014).
31. Sankaran, V. G. *et al.* Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. *Science* **322**, 1839–1842 (2008).
32. Sankaran, V. G. *et al.* Developmental and species-divergent globin switching are driven by BCL11A. *Nature* **460**, 1093–1097 (2009).
33. Xu, J. *et al.* Correction of sickle cell disease in adult mice by interference with fetal hemoglobin silencing. *Science* **334**, 993–996 (2011).
34. Kurita, R. *et al.* Establishment of immortalized human erythroid progenitor cell lines able to produce enucleated red blood cells. *PLoS ONE* **8**, e59890 (2013).
35. Hardison, R. C. & Blobel, G. A GWAS to therapy by genome edits? *Science* **342**, 206–207 (2013).
36. Canver, M. C. *et al.* Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289**, 21312–21324 (2014).
37. Mandal, P. K. *et al.* Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. *Cell Stem Cell* **15**, 643–652 (2014).
38. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
39. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnol.* **31**, 827–832 (2013).
40. Liu, P. *et al.* Bcl11a is essential for normal lymphoid development. *Nature Immunol.* **4**, 525–532 (2003).
41. John, A. *et al.* Bcl11a is required for neuronal morphogenesis and sensory circuit formation in dorsal spinal cord development. *Development* **139**, 1831–1841 (2012).
42. Yu, Y. *et al.* Bcl11a is essential for lymphoid development and negatively regulates p53. *J. Exp. Med.* **209**, 2467–2483 (2012).
43. Porcu, B. S. *et al.* The human β globin locus introduced by YAC transfer exhibits a specific and reproducible pattern of developmental regulation in transgenic mice. *Blood* **90**, 4602–4609 (1997).
44. Ran, F. A. *et al.* *In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
45. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered and improved PAM specificities. *Nature* **523**, 481–485 (2015).
46. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods* **10**, 1116–1121 (2013).
47. Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120–123 (2014).
48. Bauer, D. E., Kamran, S. C. & Orkin, S. H. Reawakening fetal hemoglobin: Prospects for new therapies for the beta-globin disorders. *Blood* **120**, 2945–2953 (2012).
49. Basak, A. *et al.* Persistence of fetal hemoglobin and altered neurodevelopment due to BCL11A deletions. *JCI* **125**, 2363–2368 (2015).
50. Funnell, P. W. *et al.* 2p15-p16.1 microdeletions encompassing and proximal to BCL11A are associated with elevated HbF in addition to neurologic impairment. *Blood* **126**, 89–93 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Hughes and D. Higgs for assistance with analysis of ChIP-seq; R. Mathieu and the Boston Children's Hospital Hematology/Oncology-HSCI Flow Cytometry Research Facility for cell sorting; Z. Herbert and F. Abderazzaq at the Dana-Farber Cancer Institute Molecular Biology Core Facility and Center for Cancer Computational Biology, respectively, for sequencing; J. Doench for providing TALENs; C. Peng for advice with MEL reporter cell generation; F. Godinho and M. Nguyen for technical help with ESCs and transgenic mice; A. Dass, C. Lin and S. Kamran for general technical assistance; C. Brendel and D. Williams for input regarding lentiviral transduction of HSPCs; J. Desimini for graphical assistance; and J. Xu and G. Lettre for insightful discussions. M.C.C. is supported by F30DK103359-01A1. E.C.S. is supported by a Jane Coffin Childs Memorial Fund for Medical Research Fellowship. L.P. is supported by NHGRI Career Development Award K99HG008399. N.E.S. is supported by a Simons Center for the Social Brain Postdoctoral Fellowship and NIH NHGRI award K99-HG008171. O.S. is supported by a fellowship from the Klarman Family Foundation. S.L. is supported by a Leukemia & Lymphoma Society Fellow Award. T.M. is supported by NIH R01 A1084905. G.-C.Y. is supported by NIH R01HL119099 and R01HG005085. F.Z. is supported by the NIMH (5DP1-MH100706) and NIDDK (5R01-DK097768), a Waterman award from the National Science Foundation, the Keck, McKnight, Damon Runyon, Searle Scholars, Merkin, Vallee, and Simons Foundations, and Bob Metcalfe. S.H.O. is supported by P01HL032262 and P30DK049216 (Center of Excellence in Molecular Hematology). D.E.B. is supported by an NIDDK Career Development Award K08DK093705, Doris Duke Charitable Foundation Innovations in Clinical Research Award (2013137), and Charles H. Hood Foundation Child Health Research Award. Computational tools and instructions for designing CRISPR-Cas9 sgRNA libraries for conducting non-coding screening can be found at the Zhang laboratory website <http://www.genome-engineering.org>.

Author Contributions D.E.B. conceived this study. N.E.S., O.S. and F.Z. conceived the pooled non-coding screening strategy using CRISPR-Cas9. M.C.C., N.E.S., O.S., F.Z., S.H.O. and D.E.B. designed and executed the pooled CRISPR screening strategy. E.C.S., F.S., Y.F., S.L., S.H.O. and D.E.B. designed, produced and analysed the transgenic mice. R.K. and Y.N. provided the HUDEP-2 cell line. M.C.C., F.S., T.M., S.H.O. and D.E.B. adapted the HUDEP-2 cell line as a model of globin gene regulation. M.C.C., F.S., D.D.C., P.G.S., D.S.V. and D.E.B. performed all experiments in cell lines. M.C.C., L.P., N.E.S., S.P.G., G.-C.Y., F.Z., S.H.O. and D.E.B. analysed the data. L.P., S.P.G. and G.-C.Y. developed the HMM. M.C.C., S.H.O., and D.E.B. wrote the manuscript with input from all authors.

Author Information All reagents described in this manuscript have been deposited with Addgene (<http://www.addgene.org>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E.B. (daniel.bauer@childrens.harvard.edu), S.H.O. (stuart_orkin@dfci.harvard.edu), or F.Z. (zhang@broadinstitute.org).

METHODS

No statistical methods were used to predetermine sample size.

Generation of genomic deletions in HUDEP-2 cells. HUDEP clone 2 (HUDEP-2) was used as previously described³⁴. HUDEP-2 cells were expanded in StemSpan SFEM (Stem Cell Technologies) supplemented with 10^{-6} M dexamethasone (Sigma), 100 ng ml^{-1} human stem cell factor (SCF) (R&D), 3 international units (IU) ml^{-1} erythropoietin (Amgen), 1% L-glutamine (Life Technologies), and 2% penicillin/streptomycin. $1 \mu\text{g ml}^{-1}$ doxycycline (Sigma) was included in the culture to induce expression of the human papilloma virus type 16 E6/E7 genes³⁴. HUDEP-2 cells were differentiated in Iscove's Modified Dulbecco's Medium (IMDM) (Life Technologies) supplemented with $330 \mu\text{g ml}^{-1}$ holo-transferrin (Sigma), $10 \mu\text{g ml}^{-1}$ recombinant human insulin (Sigma), 2 IU ml^{-1} heparin (Sigma), 5% human solvent detergent pooled plasma AB (Rhode Island Blood Center), 3 IU ml^{-1} erythropoietin, 100 ng ml^{-1} human SCF, $1 \mu\text{g ml}^{-1}$ doxycycline, 1% L-glutamine, and 2% penicillin/streptomycin.

Tandem sgRNA lentiviruses were transduced into HUDEP-2 with stable Cas9 expression (Supplementary Table 1). Bulk cultures were incubated for 7–10 days with $10 \mu\text{g ml}^{-1}$ blasticidin and $1 \mu\text{g ml}^{-1}$ puromycin selection to allow for editing. Then bulk cultures were plated clonally at limiting dilution. 96 well plates with greater than 30 clones per plate were excluded to avoid mixed clones. After approximately 14 days of clonal expansion, genomic DNA was extracted using 50 μl QuickExtract DNA Extraction Solution per well (Epicentre). Clones were screened for deletion by conventional PCR with one PCR reaction internal to segment to be deleted (non-deletion band) and one gap-PCR reaction across the deletion junction (deletion band) that would only amplify in the presence of deletion³⁶. Biallelic deletion clones were identified as the absence of the non-deletion PCR band and the presence of the deletion PCR band (Supplementary Table 2). Inversion clones were identified as previously described by PCR³⁶ (Supplementary Table 3). Briefly, inversion clones had one inverted allele and one deleted allele without the presence of non-deletion alleles. In our experience biallelic inversion clones are very rare events³⁶. PCR was performed using the Qiagen HotStarTaq 2 \times master mix and the following cycling conditions: 95 °C for 15 min; 35 cycles of 95 °C for 15 s, 60 °C for 1 min, 72 °C for 1 min; 72 °C for 10 min. Alternatively, PCR was also performed using 2 \times Accuprime Supermix II (Life Technologies) with the following cycling conditions: 94 °C for 2 min; 35 cycles of 94 °C for 20 s, 60 °C for 20 s, 68 °C for 1 min kb^{-1} of PCR product; 68 °C for 5 min. RNA was extracted from each positive clone using a kit (Qiagen) and quantitative real-time RT-qPCR was performed using iQ SYBR Green Supermix (Bio-Rad). Primers used are found in Supplementary Table 5. Gene expression was normalized to that of *GAPDH*. We isolated four control, one *BCL11A* null, three composite enhancer deleted, one h+55 deleted, one h+58 deleted, five h+62 deleted, three h+55 inverted, and two h+58 inverted clones. The *BCL11A* null clone had a 216 bp interstitial deletion of exon 2, preventing binding of the RT-qPCR primers. All gene expression data reported from these clones represents the mean of at least three technical replicates.

Design and synthesis of human and mouse lentiviral sgRNA libraries. Every 20-mer sequence upstream of an NGG or NAG PAM sequence on the plus or minus strand was identified for both the human and mouse orthologous +55, +58 and +62 DHS as well as *BCL11A/Bcl11a* exon 2 (Fig. 1 and Extended Data Figs 2, 6). Relative to the human hg19 reference genome, a reference was used with the following substitutions to approximate a common low-HbF-associated haplotype: rs1427407-G, rs1896293-T, rs6706648-T, rs6738440-G, rs7606173-C. The mouse orthologous sequences to each of the human DHSs were defined by using the liftOver tool of UCSC Genome Browser as previously described²⁸. Each of the sgRNA oligos were synthesized as previously described^{25,51,52} and cloned using a Gibson Assembly master mix (New England Biolabs) into lentiGuide-Puro (Addgene plasmid ID 52963) which had been BsmBI digested, gel purified, and dephosphorylated. Gibson Assembly products were transformed to electrocompetent cells (*E. coli*, Lucigen). Sufficient colonies were isolated to ensure $\sim 90\times$ library coverage for both human and mouse libraries. Plasmid libraries were deep sequenced to $533\times$ and $813\times$ coverage for human and mouse libraries, respectively, to confirm representation.

To produce lentivirus, HEK293T cells were cultured with Dulbecco's Modified Eagle's Medium (DMEM) (Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Omega Scientific) and 2% penicillin-streptomycin (Life Technologies) in 15 cm tissue culture treated Petri dishes. HEK293T cells were transfected at 80% confluence in 12 ml of media with 13.3 μg psPAX2, 6.7 μg VSV-G, and 20 μg of the lentiviral construct plasmid of interest using 180 μg of linear polyethylenimine (Polysciences). Medium was changed 16–24 h after transfection. Lentiviral supernatant was collected at 48 and 72 h post-transfection and subsequently concentrated by ultracentrifugation (24,000 rpm for 2 h at 4 °C with Beckman Coulter SW 32 Ti rotor).

Tiled pooled CRISPR-Cas9 screen for *in situ* functional mapping the human *BCL11A* erythroid enhancer. HUDEP-2 cells with stable Cas9 expression were transduced at low multiplicity with the human sgRNA library lentivirus pool while in expansion medium. Control transductions were performed to ensure transduction rate did not exceed 50%. Cell numbers were maintained throughout the experiment at levels adequate to exceed $1,000\times$ representation of the library. $10 \mu\text{g ml}^{-1}$ blasticidin (Sigma) and $1 \mu\text{g ml}^{-1}$ puromycin (Sigma) were added 24 h after transduction to select for lentiviral library integrants in cells with Cas9. Cells were cultured in expansion media for one week followed by differentiation media for an additional week.

Intracellular staining was performed by fixing cells with 0.05% glutaraldehyde (grade II) (Sigma) for 10 min at room temperature. Cells were centrifuged for 5 min at 600g and then resuspended in 0.1% Triton X-100 (Life Technologies) for 5 min at room temperature for permeabilization. Triton X-100 was diluted with phosphate buffered saline (PBS) with 0.1% BSA and then centrifuged at 600g for 15 min. Cells were stained with anti-human antibodies for HbF (clone HbF-1 with FITC or APC conjugation; Life Technologies) and β -haemoglobin antibody (clone 37-8 with PerCP-Cy5 or PE conjugation; Santa Cruz) for 20 min in the dark. Cells were washed to remove unbound antibody before FACS analysis. 0.2 μg HbF and 2 μg of adult haemoglobin (HbA) (β -haemoglobin) antibodies were used per 5 million cells. Control cells exposed to a non-targeting sgRNA sample and *BCL11A* exon 2 were used as negative and positive controls, respectively, to establish flow cytometry conditions. Populations of cells with the top and bottom 10% of expression of HbF were sorted by FACS.

After sorting the HbF-high and HbF-low pools, library preparation and deep sequencing was performed as previously described²⁵. Briefly, genomic DNA was extracted using the Qiagen Blood and Tissue kit. Hercules PCR reaction (Agilent) using lentiGuide-Puro specific primers (5'-AATGGACTATCATA TGCTTACCGTAACCTGAAAGTATTTTCG-3' and 5'-CTTTAGTTGTAT GTCTGTTGCTATTATGCTACTATTCTTTCCC-3') including a handle sequence was performed as follows: Hercules II reaction buffer (1 \times), forward and reverse primers (0.5 μM each), dimethyl sulfoxide (DMSO) (8%), deoxynucleotide triphosphates (dNTPs) (0.25 mM each), Hercules II Fusion DNA Polymerase (0.5 reactions) using the following cycling conditions: 95 °C for 2 min; 20 cycles of 95 °C for 15 s, 60 °C for 20 s, 72 °C for 30 s; 72 °C for 5 min. Multiple reactions of no more than 200 ng each were used to amplify from 6.6 μg gDNA ($\sim 1\times 10^6$ cell genomes) per pool. Samples were subjected to a second PCR using handle-specific primers²⁵ to add adaptors and indexes to each sample using the following conditions: Hercules II reaction buffer (1 \times), forward and reverse primers (0.5 μM each), dNTPs (0.25 mM each), Hercules II Fusion DNA Polymerase (0.5 reactions) with the following cycling conditions: 95 °C for 2 min; 25 cycles of 95 °C for 15 s, 60 °C for 20 s, 72 °C for 30 s; 72 °C for 5 min. PCR products were run on an agarose gel and the band of expected size was gel purified. Illumina MiSeq 150 bp paired end sequencing was performed.

sgRNA sequences present in the plasmid pool as well as in the HbF-high and HbF-low pools were enumerated. Guide sequences were mapped to the guides comprising the sgRNA library without allowing mismatches. Total reads were normalized to library sequencing depth. Cellular dropout score was determined by calculating (1) the ratio of normalized reads in the cells at end of experiment (average of reads in the HbF-high and HbF-low pools) to reads in the plasmid pool; (2) \log_2 transformation; and (3) median of biological replicates. HbF enrichment score was determined by calculating (1) the ratio of normalized reads in the HbF-high compared to reads in the HbF-low pools; (2) \log_2 transformation; and (3) median of biological replicates. After exclusion of sgRNAs with dropout scores $< 2^{-3}$ and NAG PAM sgRNAs, a quantile-quantile plot was made with a line fitted through the first and third quantiles using R software. HbF enrichment scores and cellular dropout scores were compared by Spearman rank correlation. sgRNA sequences were mapped to the human genome (hg19) with cleavage positions set to between positions 17 and 18 given PAM positions 21–23. For visual comparisons to targeting sgRNAs, non-targeting sgRNAs were pseudomapped each separated by 5 bp.

Validation in primary human CD34⁺ HSPCs. Primary human CD34⁺ HSPCs from G-CSF mobilized healthy adult donors were obtained from the Center of Excellence in Molecular Hematology at the Fred Hutchinson Cancer Research Center, Seattle, Washington. CD34⁺ HSPCs were subject to erythroid differentiation liquid culture as previously described⁵³. Briefly, HSPCs were thawed on day 0 into erythroid differentiation medium (EDM) consisting of IMDM supplemented with $330 \mu\text{g ml}^{-1}$ holo-human transferrin, $10 \mu\text{g ml}^{-1}$ recombinant human insulin, 2 IU ml^{-1} heparin, 5% human solvent detergent pooled plasma AB, 3 IU ml^{-1} erythropoietin, 1% L-glutamine, and 2% penicillin/streptomycin. During days 0–7 of culture, EDM was further supplemented with 10^{-6} M hydrocortisone (Sigma), 100 ng ml^{-1} human SCF, and 5 ng ml^{-1} human IL-3 (R&D).

During days 7–11 of culture, EDM was supplemented with 100 ng ml^{-1} human SCF only. During days 11–18 of culture, EDM had no additional supplements.

HSPCs were transduced with lentiCas9-Blast (Addgene plasmid ID 52962) 24 h after thawing in the presence of $10 \mu\text{M}$ 16,16-dimethylprostaglandin E2 (PGE2; Cayman Chemical). At 48 h after thawing, medium was changed and cells were transduced with lentiGuide-Puro or lentiGuide-Crimson cloned with relevant sgRNA sequence in the presence of $10 \mu\text{M}$ PGE2. Three independent transductions were performed per sgRNA. At 72 h after thawing, medium was changed and HSPCs were selected with $10 \mu\text{g ml}^{-1}$ blasticidin and $1 \mu\text{g ml}^{-1}$ puromycin or $10 \mu\text{g ml}^{-1}$ blasticidin followed by sorting for lentiGuide-Crimson⁺ cells on day 16 of culture. Blasticidin and/or puromycin selection occurred from days 3 to 8 of culture.

Differentiation was assessed on day 18 of culture using anti-human antibodies against the transferrin receptor (CD71) (Clone OKT9 with FITC conjugation; eBioscience) and glycophorin A (CD235a) (Clone HIR2 with PE conjugation; eBioscience). Enucleation was assessed using $2 \mu\text{g ml}^{-1}$ of the cell-permeable DNA dye Hoescht 33342 (Life Technologies). CD235a⁺Hoescht 33342⁻ cells were determined to be enucleated erythroid cells. Cells were intracellularly stained for HbF and HbA on day 18 of culture as described above. 50,000–100,000 cells were centrifuged onto microscope slides at 350 rpm for 4 min. Slides were stained with Harleco May-Grünwald stain (Millipore) for 2 min, Giemsa stain (Sigma) for 12 min, and two water washes for 30 s each. Slides were air dried and then coverslipped using Fisher Chemical Permount Mounting Medium (Fisher). RNA isolation and RT-qPCR was performed as above. Gene expression was normalized to that of *GAPDH*. All gene expression data represent the mean of at least three technical replicates.

PCR primers were designed to amplify the genomic cleavage site for a given sgRNA. Resulting PCR products were subjected to Sanger sequencing. Sequencing traces were used for editing quantification using a previously described publically available tool⁵⁴.

Computational analysis. Human erythroid H3K27ac ChIP-seq was obtained from Xu *et al.*⁷ and mouse erythroid H3K27ac ChIP-seq was obtained from Kowalczyk *et al.*⁵⁵ and Dogan *et al.*⁵⁶. We uniformly processed all the data sets using the same pipeline with the same criteria to call super-enhancers. Specifically, we started from raw reads and realigned each data set with Bowtie2 with the default parameters. We then removed duplicate reads with the Picard Suite. To call the peaks we used MACS2 in the narrow mode. Finally, to call the super-enhancers we used the ROSE algorithm with the default parameters¹⁰. Using these settings, peaks closer than 12.5 kb are stitched together and then ranked based on the H3K27ac intensity. To assign super-enhancers to genes we used again ROSE with default settings. In particular, the tool reports three categories of genes for each super-enhancer: (1) overlapping genes (genes for which the gene body region overlaps a super-enhancer); (2) proximal genes (genes close to a super-enhancer considering a window of 50 kb); (3) closest gene (closest gene considering its TSS and the centre of the super-enhancer). To generate a Venn diagram of genes for super-enhancer data sets, we used the union of these three gene categories.

HMM segmentation was performed to automatically segment the enrichment score signals into enhancer regions with active, repressive and neutral effect. We designed a HMM with 3 states using the GHMM package (<http://ghmm.sourceforge.net/>). To learn the HMM parameters we used the Baum–Welch algorithm. To find the best segmentation for each region we used the Viterbi algorithm. The emission probability for each state was modelled as a Gaussian distribution and all the possible transitions between states were allowed as shown in Extended Data Fig. 4a. Since the signal was not obtained with a constant genomic resolution, we interpolated and smoothed the signal using a Gaussian kernel over 12 bp and applied the HMM to the smoothed signal. To set the initial parameters, we used the 1%, 50% and 99% percentile of the smoothed signal for the prior of the means of the repressive, neutral and active states, respectively, while the prior for the standard deviation was set to 0.001 for all the three states.

Motif analysis was performed to evaluate the human and mouse enhancer regions for potential binding sites for known transcription factors. We used the FIMO software⁵⁷ with a *P*-value threshold of $<10^{-4}$. For each region we extracted sequences using the hg19 and mm9 assemblies respectively for human and mouse. The motif database was the latest version of the JASPAR database⁵⁸.

Deep sequencing paired-end reads of genomic amplicons from genome editing target sites were first filtered for reads with PHRED quality score <30 , merged with the FLASH (Fast Length Adjustment of SHort reads) software, and subsequently aligned to a reference amplicon using the needle aligner from the EMBOSS suite (<http://emboss.sourceforge.net/>) to quantify insertions and deletions. Per nucleotide frequency of deletion of a position, insertion directly adjacent to the position, or no mutation at the position was quantitated using CRISPResso (<https://github.com/lucapinello/CRISPResso>).

Pooled CRISPR-Cas9 screen for high-resolution functional mapping of mouse *Bcl11a* enhancer. Murine erythroleukaemia (MEL, MEL-745A cl. DS19) cells were cultured in DMEM supplemented with 10% FBS, 1% L-glutamine, and 2% penicillin/streptomycin as previously described^{28,36}. Cell lines tested negative for mycoplasma contamination. $\epsilon\text{y:mCherry}$ reporter MEL cells with stable Cas9 expression were transduced at low multiplicity with the mouse sgRNA library lentivirus pool (see Supplementary Information and Extended Data Fig. 6 for additional technical details). Control transductions were performed to ensure that the transduction rate did not exceed 50%. Cell numbers were maintained throughout the experiment at levels adequate to exceed $1,000\times$ representation of the library. $10 \mu\text{g ml}^{-1}$ blasticidin and $1 \mu\text{g ml}^{-1}$ puromycin were added 24 h after transduction to select for lentiviral library integrants in cells with Cas9. Subsequently cells were cultured for 2 weeks. The top and bottom 5% of $\epsilon\text{y-mCherry}$ -expressing cells exposed to the library were sorted by FACS. A non-targeting sgRNA sample was used as a negative control and *Bcl11a* exon 2 as a positive control to establish flow cytometry conditions. After sorting, library preparation and deep sequencing were performed as described for the human library²⁵.

sgRNA sequences present in the Hbb- $\epsilon\text{y:mCherry}$ -high and Hbb- $\epsilon\text{y:mCherry}$ -low pools were enumerated. Cellular dropout and ϵy enrichment scores were calculated analogously to the human screen. sgRNA sequences were then mapped to the mouse genome (mm9).

Generation of genomic deletions in MEL cells. Deletions in MEL cells were generated using two sgRNA as previously described³⁶. Briefly, sgRNA sequences were cloned into pX330 (Addgene plasmid ID 42230) using a Golden Gate assembly cloning strategy (Supplementary Table 1). MEL cells were electroporated with 5 μg of each pX330-sgRNA plasmid and 0.5 μg pmax-GFP (Lonza) in BTX electroporation buffer using a BTX electroporator (Harvard Apparatus). Approximately 48 h post-electroporation, the top 1–3% of GFP⁺ cells were sorted and plated clonally at limiting dilution. Clones were allowed to grow for 7–10 days. Clones were screened for deletion by conventional PCR using the same strategy as with the HUDEP-2 cells (Supplementary Tables 2 and 4). Inversion clones were identified by PCR as previously described³⁶ (Supplementary Table 3).

Generation of genomic deletions in mouse embryonic stem cells (mESCs). mESCs were maintained on irradiated mouse embryonic fibroblasts (GlobalStem) and cultured in high glucose DMEM supplemented with 15% FBS, 1% L-glutamine, 2% penicillin/streptomycin (Life Technologies), 100 μM non-essential amino acids (Life Technologies), 1% nucleosides (Sigma), 10^{-4} M β -mercaptoethanol (Sigma), and 10^3 U ml⁻¹ leukaemia inhibitory factor (Millipore). Cells were passaged using 0.25% trypsin (Life Technologies).

The *Bcl11a* +62 deletion mice were derived from CRISPR-Cas9 modified C9J ES cells. Using Amara ES Cell transfection reagent (Lonza), two million mESCs were electroporated with 2 μg of each pX330 plasmid vector containing individual target sequences flanking the +62 site along with 0.5 μg of a GFP plasmid. After 48 h, the top 5% of GFP expressing cells were sorted, plated on irradiated fibroblasts and maintained. Individual ES cell colonies were then picked and screened for biallelic deletion using the same strategy as HUDEP-2 and MEL cells³⁶. DNA for screening CRISPR-Cas9 modified clones was obtained from gelatin adapted ES cell clones to avoid genomic contamination from the fibroblasts. Correctly targeted clones with greater than 80% normal karyotype were used to generate mice. Clones were injected into embryonic day 3.5 (E3.5) C57BL/6 blastocysts and implanted into pseudo-pregnant females.

The β -YAC mouse line (A20), previously described as containing a transgene encompassing ~ 150 kb of the human β -globin locus⁴³, was used to analyse human globin expression. The mouse line was maintained in a hemizygous state and bred with *Bcl11a* +62 deletion mice. Sufficient matings were established to ensure adequate homozygotes for analysis.

Mouse cell and tissue analysis. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. For developmental haematopoiesis, fetal liver cells were taken at E12.5, E14.5, E16.5 and E18.5 and mechanically dissociated to form single cell suspensions from which RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) and analysed. At E16.5, fetal livers were also stained with CD19-PerCP-Cy5.5 (Clone 1D3; eBioscience), B220-APC (RA3-6B2; Biolegend), CD71-PE (Clone C2; BD Biosciences), and Ter119-FITC (Clone Ter119; BD Biosciences) to isolate B cells (B220⁺CD19⁺) and erythroid cells (Ter119⁺CD71⁺) by FACS for RNA extraction and *Bcl11a* quantification. Additionally, flow cytometry was used to analyse fetal liver from E18.5 embryos. Single cell suspensions were stained with IgM-FITC (Clone Il-41; eBioscience), CD19-PerCP-Cy5.5, (Clone 1D3; eBioscience), CD43-PE (Clone S7; eBioscience), AA4.1-PE-Cy7 (Clone AA4.1; BD Biosciences), B220-APC, (RA3-6B2; Biolegend), and DAPI (Invitrogen). For adult haematopoietic assays, peripheral blood was obtained from the tail vein of 4-week-old male and female mice. Blood was collected in EDTA-coated tubes, red

cells removed by 2% dextran (Sigma), residual red cells lysed with ammonium chloride solution (Stem Cell Technologies) and stained with the following anti-mouse antibodies: CD3e-FITC (Clone 145-2C11; Biolegend), CD19-PerCP-Cy5.5 (Clone 1D3; eBioscience), CD71-PE (Clone C2; BD Biosciences), NK1.1-PE-Cy5 (Clone PK136; Biolegend), Ter119-APC (Clone TER-119; Biolegend), Gr-1-eF450 (Clone RB6-8C5; eBioscience), B220-BV605 (RA3-6B2; Biolegend), Mac-1-BV510 (Clone M1/70; Biolegend), and 7-AAD (BD Biosciences). Fetal brain analysis was conducted on whole brains from E16.5 mouse embryos on ice-cold PBS. Tissue was directly lysed into the RLT plus buffer (Qiagen) and total RNA extracted according to manufacturer's instructions provided in the RNeasy Plus Mini Kit. RT-qPCR was performed as above, with gene expression normalized to *Gapdh*. All gene expression data represent the mean of at least three technical replicates. All animal experiments were conducted under the approval of the local Institutional Animal Care and Use Committee.

Cloning lentiCas9-Venus. Venus template⁵⁹ was PCR amplified to add BamHI (5') and EcoRI (3') restriction sites (lowercase font) for cloning purposes using the following conditions: KOD buffer (1×), MgSO₄ (1.5 mM), dNTPs (0.2 mM each), forward primer (0.3 μM; [GGCCGGCCggatccGGCGCAACA AACTTCTCTCTGCTGAAACAAGCCGGAGATGTGCAAGAGAATCCTGG ACCGATGGTGAGCAAGGGCGAGGA), reverse primer (0.3 μM; GGCCGGC CgaattcTTACTGTACAGCTCGTCCA), and KOD Hot Start DNA Polymerase (0.02 U μl⁻¹) (Millipore). KOD PCR reaction used the following cycling conditions: 95 °C for 2 min; 50 cycles of 95 °C for 20 s, 60 °C for 20 s, and 70 °C for 30 s; 60 °C for 5 min. PCR products were purified (QIAquick PCR Purification Kit, Qiagen) and blunt end cloned with Zero Blunt PCR cloning kit (Invitrogen). PCR-blunt cloned products and lentiCas9-Blast (Addgene plasmid ID 52962) were separately digested with BamHI-HF (New England Biolabs) and EcoRI-HF (New England Biolabs) in 1× Buffer CutSmart at 37 °C (New England Biolabs). Digest of lentiCas9-Blast was performed to remove the blasticidin cassette. Then digested PCR product was ligated into the lentiCas9 backbone.

Cloning lentiGuide-Crimson. E2-Crimson template (Clontech) was PCR amplified to add BsiWI (5') and MluI (3') restriction sites for cloning purposes using the following conditions: KOD buffer (1×), MgSO₄ (1.5 mM), dNTPs (0.2 mM each), forward primer (0.3 μM; GGCCGGCCCGTACGctagcGCCACCATG GATAGCACTGAGAACGTCATCAAGCCCTT), reverse primer (0.3 μM; GG CCGGCCacgctTACTGGAACAGGTGGTGGCGGGCCT), and KOD Hot Start DNA Polymerase (0.02 U μl⁻¹). KOD PCR reaction used the following cycling conditions: 95 °C for 2 min; 50 cycles of 95 °C for 20 s, 60 °C for 20 s, and 70 °C for 30 s; 60 °C for 5 min. PCR products were purified (QIAquick PCR Purification Kit) and cloned with Zero Blunt PCR cloning kit. Cloned products and lentiGuide-puro were separately digested with BsiWI (New England Biolabs) and MluI (New England Biolabs) in 1× buffer 3.1 at 37 °C (New England Biolabs). Digest of lentiGuide-Puro (Addgene plasmid ID 52963) was performed to remove the puromycin cassette. Then digested PCR product was ligated into the lentiGuide backbone.

Cloning sgRNAs. lentiGuide-Puro (Addgene plasmid ID 52963) was digested with BsmBI in 1× buffer 3.1 at 55 °C (New England Biolabs) for linearization. One unit of TSAP thermosensitive alkaline phosphatase (Promega) was added for 1 h at 37 °C to dephosphorylate the linearized lentiGuide and then TSAP was heat inactivated at 74 °C for 15 min. Linearized and dephosphorylated lentiGuide was run on an agarose gel and gel purified. sgRNA-specifying oligos were phosphorylated and annealed using the following conditions: sgRNA sequence oligo (10 μM); sgRNA sequence reverse complement oligo (10 μM); T4 ligation buffer (1×) (New England Biolabs); and T4 polynucleotide kinase (5 units) (New England Biolabs) with the following temperature conditions: 37 °C for 30 min; 95 °C for 5 min; and then ramp down to 25 °C at 5 °C min⁻¹. Annealed oligos were ligated into lentiGuide in a 1:3 ratio (vector:insert) using T4 ligation buffer (1×) and T4 DNA Ligase (750 units) (New England Biolabs). Plasmids were verified by sequencing using a U6 promoter forward primer CGTAACTTG AAAGTATTCGATTCTTGGC.

sgRNA-specifying oligos using sgRNA sequences from the screen library (see Source Data associated with Figs 2 and 5) were obtained and cloned as described into either lentiGuide-Puro or lentiGuide-Crimson. sgRNA constructs were used to produce lentivirus and transduce HUDEP-2 with stable Cas9 expression. Bulk cultures were incubated for 7–10 days with 10 μg ml⁻¹ blasticidin and 1 μg ml⁻¹ puromycin selection to allow for editing. Then bulk cultures were plated clonally

at limiting dilution. Clones were allowed to grow for approximately 14 days and then genomic DNA was extracted using 50 μl QuickExtract DNA Extraction Solution per well.

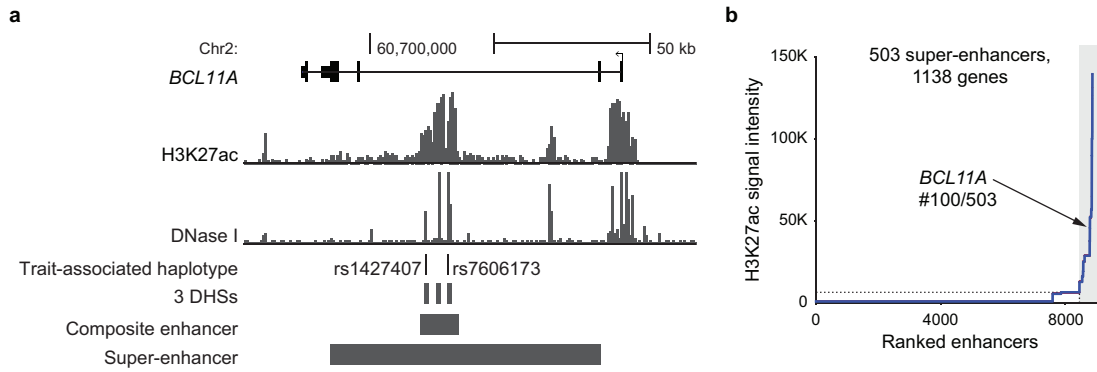
lentiTandemGuide cloning. lentiGuide-sgRNA1 was digested with PspXI and XmaI at 37 °C for 4 h (New England Biolabs). Digests were run on an agarose gel and gel purified. lentiGuide-sgRNA2 was linearized using NotI (New England Biolabs). The hU6 promoter and sgRNA chimaeric backbone for lentiGuide-sgRNA2 was PCR amplified using the following conditions: KOD buffer (1×), MgSO₄ (1.5 mM), dNTPs (0.2 mM each), forward primer (0.3 μM; GGCCGGCC gctcgggGAGGGCCTATTTC), reverse primer (0.3 μM; CCGCCGGCccgggT TGTGGATGAATACTGCCATTT), and KOD Hot Start DNA Polymerase (0.02 U μl⁻¹) (Millipore). KOD PCR reaction used the following cycling conditions: 95 °C for 2 min; 50 cycles of 95 °C for 20 s, 60 °C for 20 s, and 70 °C for 30 s; 60 °C for 5 min. PCR products were purified (QIAquick PCR Purification Kit), blunt-ended cloned with Zero Blunt PCR cloning kit, transformed, and plated. Colonies were screened by digesting minipreps with EcoRI. Mini-preps were then digested with PspXI and XmaI as described above followed by PCR purification. After PCR purification, sgRNA2 was ligated into digested lentiGuide-sgRNA1. Sequence was verified with following primers: GGAGGCTTGGTAGGTTTA AGAA and CCAATTCCTACTCCTTTCAA.

Generation of HUDEP-2 with stable Cas9. lentiCas9-Blast (Addgene plasmid ID 52962) or lentiCas9-Venus were produced as described above and used to transduce HUDEP-2 cells. Transduced cells were selected with 10 μg ml⁻¹ blasticidin or Venus⁺ cells were sorted. Functional Cas9 was confirmed using the pXPR-011 (Addgene plasmid ID 59702) GFP reporter assay as previously described⁶⁰.

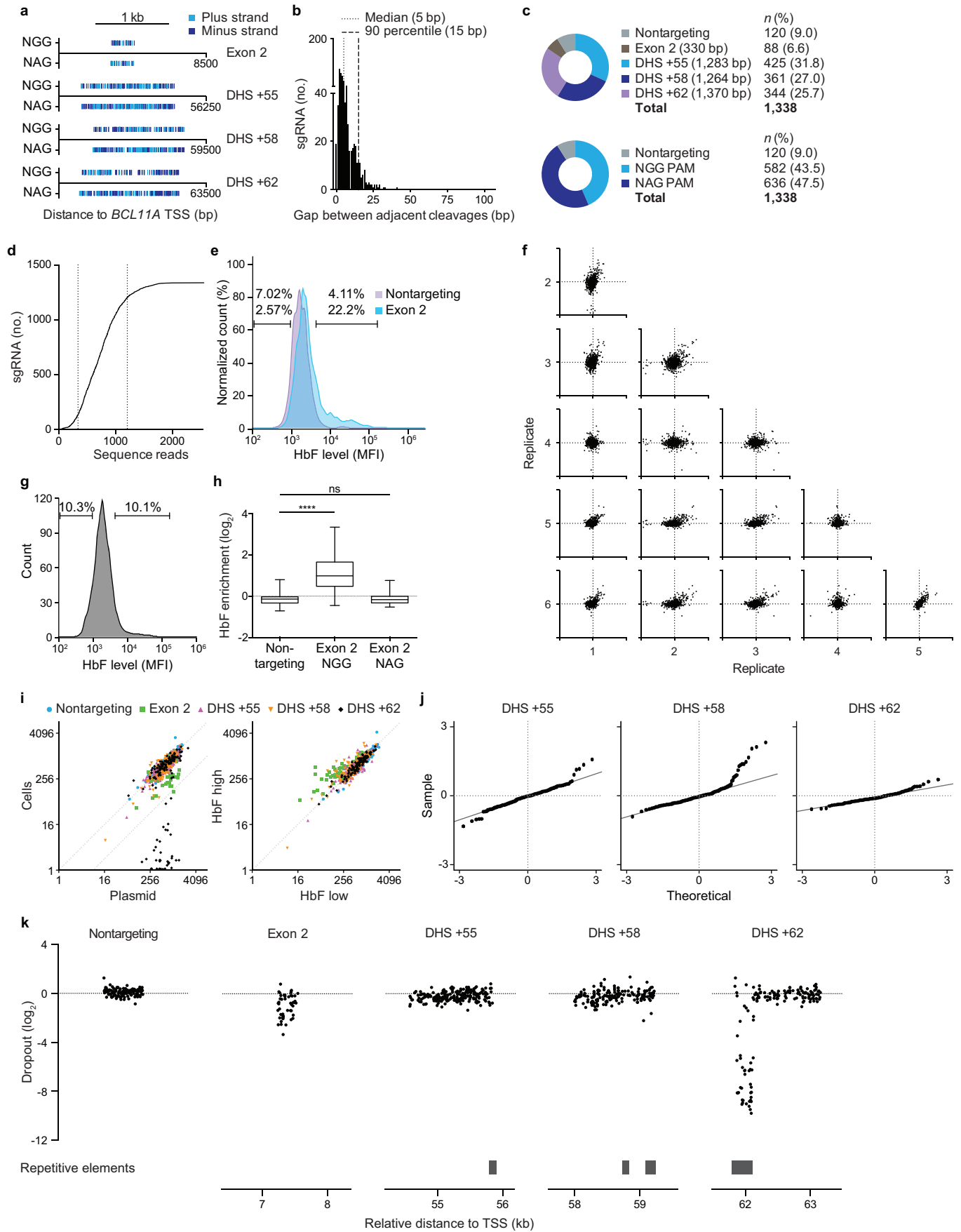
Generation of *ey:mCherry* reporter MEL cells. A reporter MEL line in which mCherry was knocked into the *Hbb-y* locus was created (Extended Data Fig. 6c). Briefly, a TALEN-induced DSB was created adjacent to the *Hbb-y* transcriptional start site. A targeting vector with mCherry and a neomycin cassette was introduced through homology-directed repair. Homology arms included mm9 sequences from Chr7:111,001,667–111,002,675 and Chr7:111,000,661–111,001,666. Cre-mediated recombination was used to remove the neomycin cassette. Long-range PCR spanning each homology arm was used to ensure appropriate targeted integration. Cells were tested upon *Bcl11a* disruption by RT-qPCR and flow cytometry to confirm expected effects on *ey:mCherry* derepression. Subsequently, CRISPR-Cas9 was used as described above to produce cells with monoallelic composite enhancer deletion to maximize screening sensitivity for enhancer disruption.

Generation of MEL cells with stable Cas9 expression. lentiCas9-Blast (Addgene plasmid ID 52962) lentivirus was produced as described above and used to transduce MEL cells. Transduced cells were selected with 10 μg ml⁻¹ blasticidin. Functional Cas9 was confirmed using the pXPR-011 (Addgene plasmid ID 59702) GFP reporter assay as previously described⁶⁰.

- Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods* **11**, 783–784 (2014).
- Chen, S. *et al.* Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246–1260 (2015).
- Giarratana, M. *et al.* Proof of principle for transfusion of *in vitro* generated red blood cells. *Blood* **118**, 5071–5079 (2011).
- Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).
- Kowalczyk, M. S. *et al.* Intragenic enhancers act as alternative promoters. *Mol. Cell* **45**, 447–458 (2012).
- 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**, 16 (2015).
- Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
- 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).
- 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).
- Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnol.* **32**, 1262–1267 (2014).

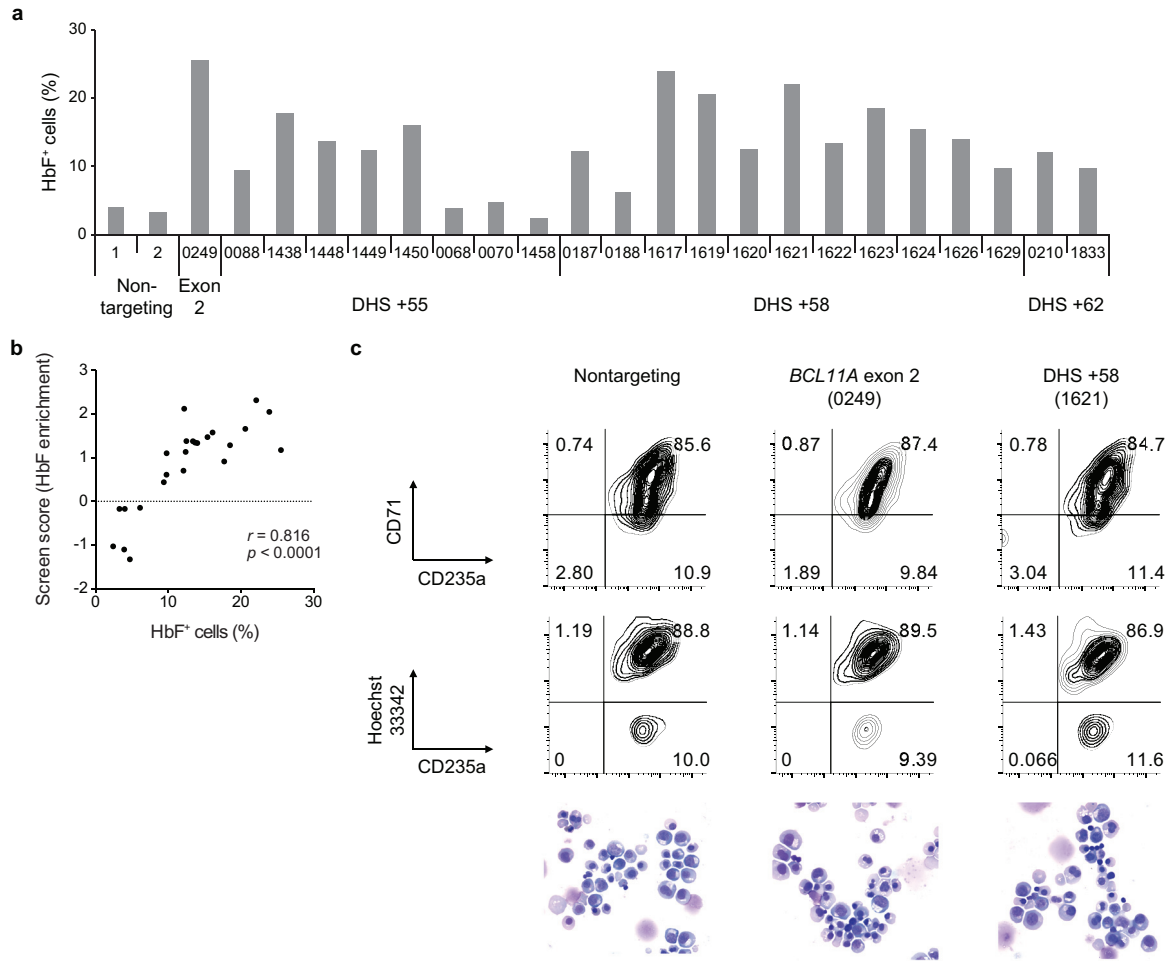


Extended Data Figure 1 | Human *BCL11A* locus. **a**, Schematic of the human *BCL11A* locus (hg19, transcription from right to left) with erythroid chromatin marks and trait-associated haplotype denoted, and composite enhancer as previously defined²⁸. **b**, Ranked enhancers in primary human adult erythroid precursors by H3K27ac signal intensity, with super-enhancers shaded, and super-enhancer-associated genes indicated.



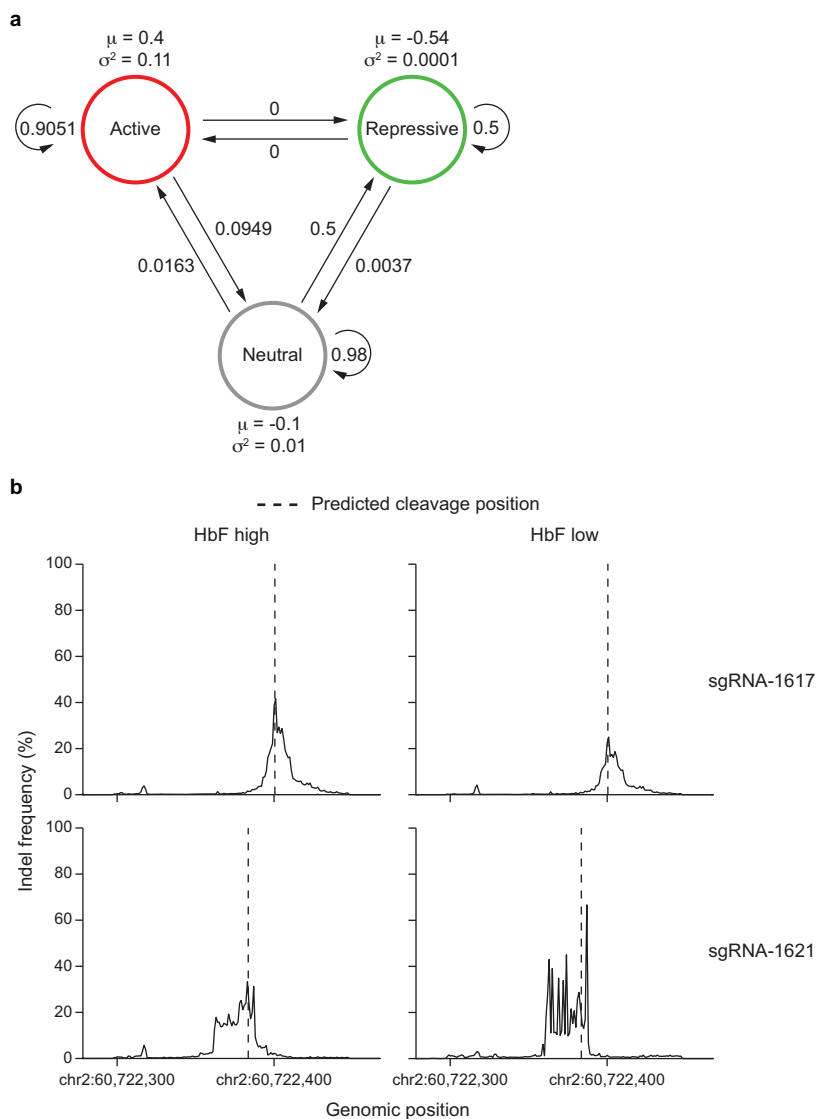
Extended Data Figure 2 | Tiled pooled *in situ* CRISPR-Cas9 *BCL11A* enhancer screen. **a**, Distribution of NGG and NAG PAM sgRNAs mapped to genomic cleavage position. The vertical lines represent cleavage sites for sgRNAs mapped to plus and minus strands. **b**, Gap distance between adjacent genomic cleavage position for NAG PAM sgRNAs. **c**, Library composition by target sequence and PAM restriction. **d**, Representation of both NGG and NAG sgRNA (1,338 sgRNAs in total) within the plasmid pool by deep sequencing. The median was 718 normalized reads and the 10th and 90th percentiles (indicated by the vertical dotted lines) ranged from 337 to 1,205 normalized reads. **e**, HbF distribution in HUDEP-2 cells transduced with Cas9 and individual sgRNAs, either non-targeting or targeting *BCL11A* exon 2.

f, HbF enrichment scores of NGG sgRNAs in six biological replicates. **g**, Sort of library-transduced cells into HbF-high and HbF-low pools. **h**, Control sgRNA enrichment. Boxes demonstrate 25th, median, and 75th percentiles and whiskers minimum and maximum values. **** $P < 0.0001$, NS, non-significant. **i**, NGG sgRNA representation in plasmid pool and cells at conclusion of experiment (left), and in HbF-high and HbF-low pools (right), with dotted lines at $x = y$ and $x = 8y$. **j**, Quantile–quantile plots of NGG sgRNA enrichment scores. **k**, Cellular dropout scores of NGG sgRNAs relative to genomic cleavage position and repetitive elements. Non-targeting sgRNAs pseudo-mapped with 5-bp spacing.



Extended Data Figure 3 | Validation of enhancer screen. **a**, HbF⁺ fraction in HUDEP-2 cells transduced in arrayed format with 24 sgRNAs from all 5 mapping categories with enrichment scores ranging from the highest to the lowest in the screen. **b**, Correlation between HbF enrichment score from pooled sgRNA screen and HbF⁺ fraction by arrayed validation of individual

sgRNAs in HUDEP-2 cells. **c**, Erythroid differentiation of primary human erythroid precursors evaluated by CD71 and CD235a surface markers, enucleation frequency (CD235a⁺ Hoescht33342⁻), and morphology by May-Grünwald Giemsa staining.

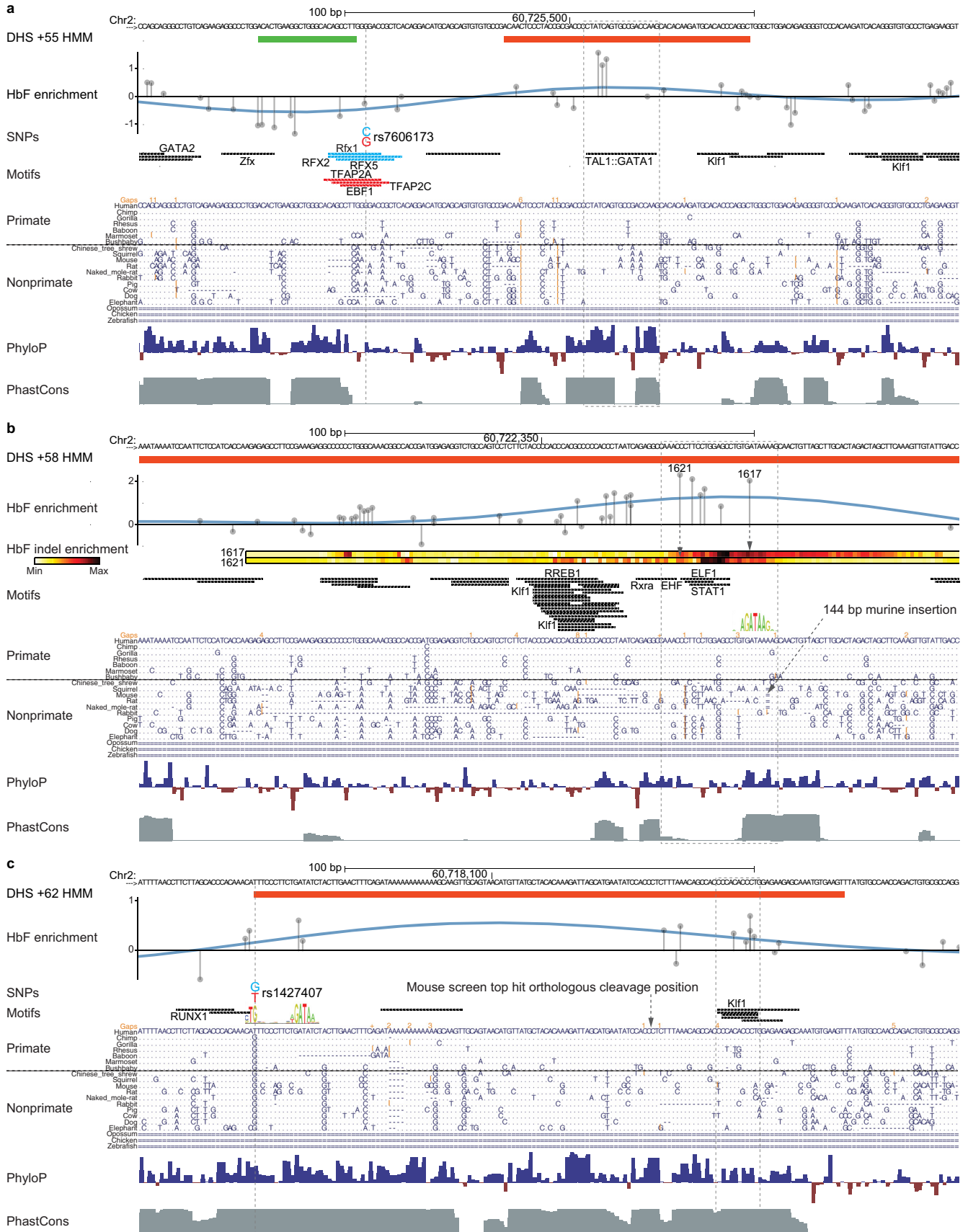


Extended Data Figure 4 | Functional assessment of enhancer sequences.

a, Topology of the HMM used to infer the three functional enhancer states (active, repressive and neutral). The emission probabilities of HbF enrichment scores from each state were modelled as Gaussian distributions (the values of μ and σ^2 are shown). The transition probabilities (arrows) are displayed.

b, Frequency distribution of indels from HUDEP-2 cells exposed to Cas9 and

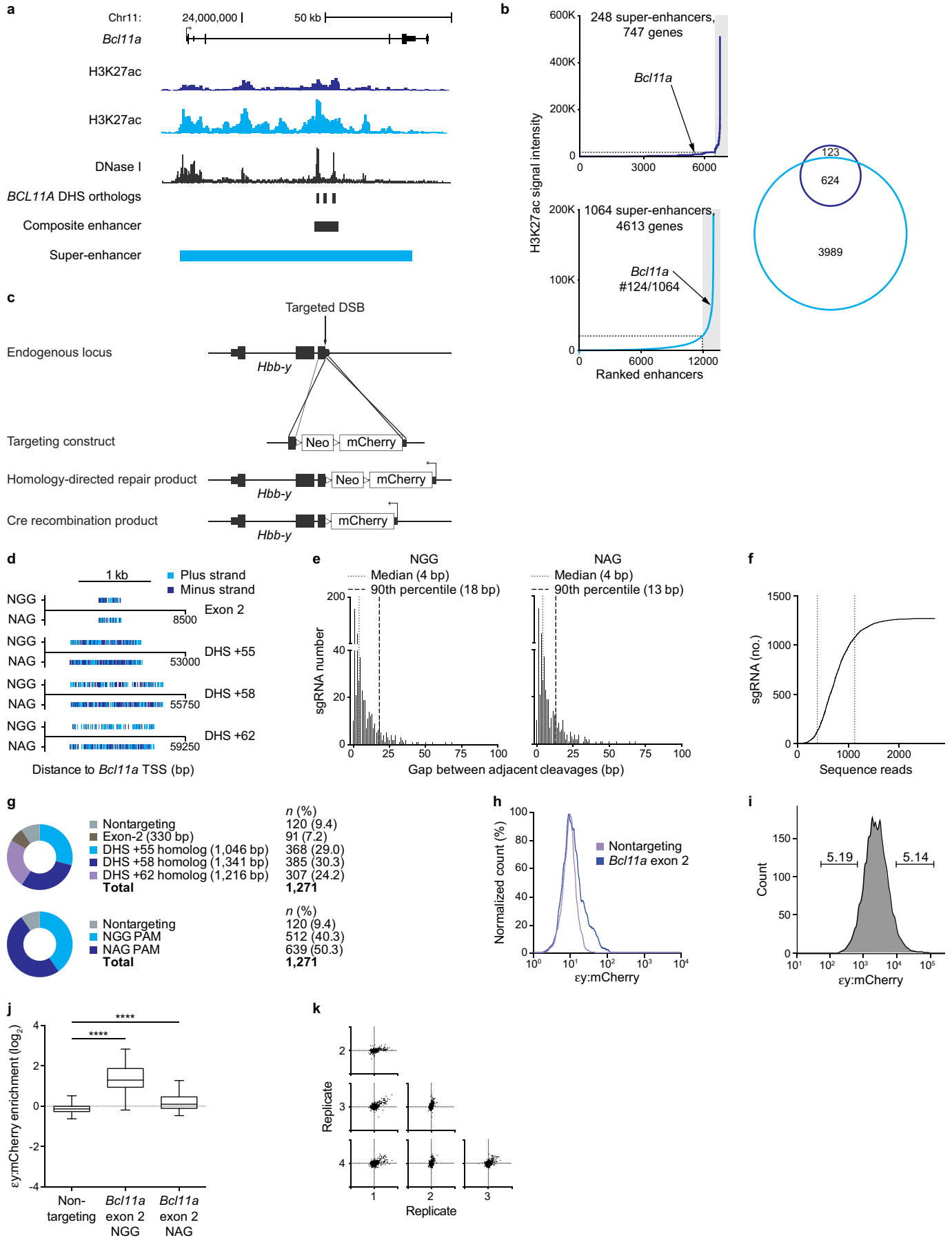
individual sgRNAs, sorted into HbF-high and HbF-low pools, and subjected to deep sequencing of the target site. Indels calculated on a per nucleotide basis throughout an amplicon surrounding the sgRNA-1617 and -1621 cleavage sites (dotted lines). An indel enrichment ratio was calculated by dividing normalized indel frequencies in the HbF-high pool by those in the HbF-low pool.



Extended Data Figure 5 | Functional cores of the *BCL11A* enhancer.

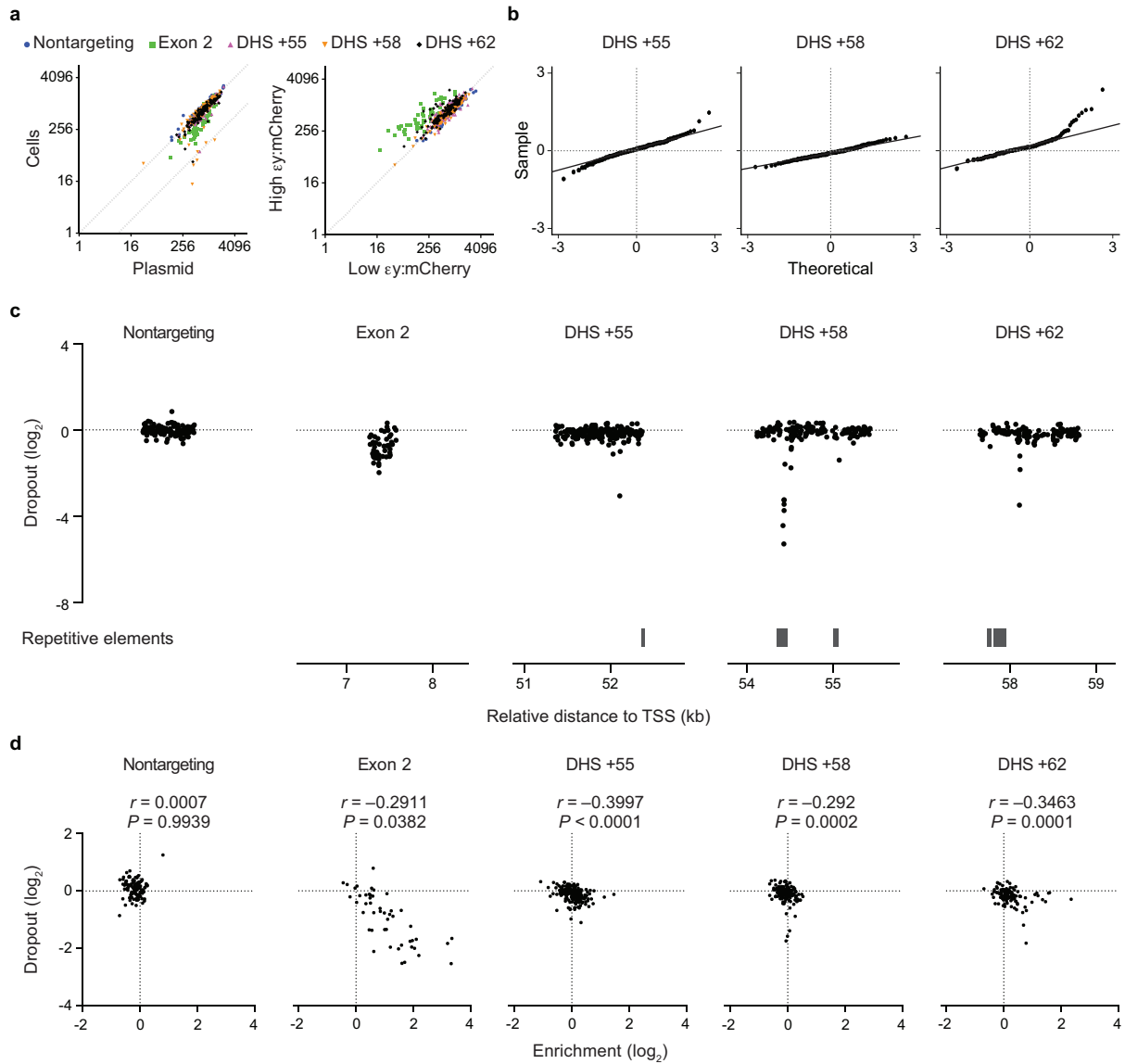
a–c, 200 bp at the functional cores of DHSs h+55, h+58 and h+62 defined by HMM states (active, red; repressive, green). HbF enrichment scores are shown by grey lines and circles. HbF indel enrichment per nucleotide based on amplicon genomic sequencing of sorted cells exposed to either sgRNA-1617 (top) or -1621 (bottom) is shown. Common SNPs (MAF > 1%) are shown with dotted lines with HbF-low allele in blue and HbF-high allele in red; no common SNPs are present at the h+58 region. JASPAR motifs ($P < 10^{-4}$) are depicted in black except for those with allele-specific significance depicted by allelic colour. Selected motifs annotated by transcription factor on the basis

of known erythroid-specific function or genomic position. Motif LOGOs at key positions with motif scores $P < 10^{-3}$ as described in text. Dotted boxes show regions of highest HbF enrichment score at each core with underlying predicted motifs. Orthologous sequences are listed from representative primates and nonprimates of distributed phylogeny. PhyloP (scale from -4.5 to 4.88) and PhastCons (from 0 to 1) estimates of evolutionary conservation among 100 vertebrates are shown. An arrow indicates a 144 bp insertion in the mouse genome relative to the human reference adjacent to the orthologous GATA1 motif at h+58.



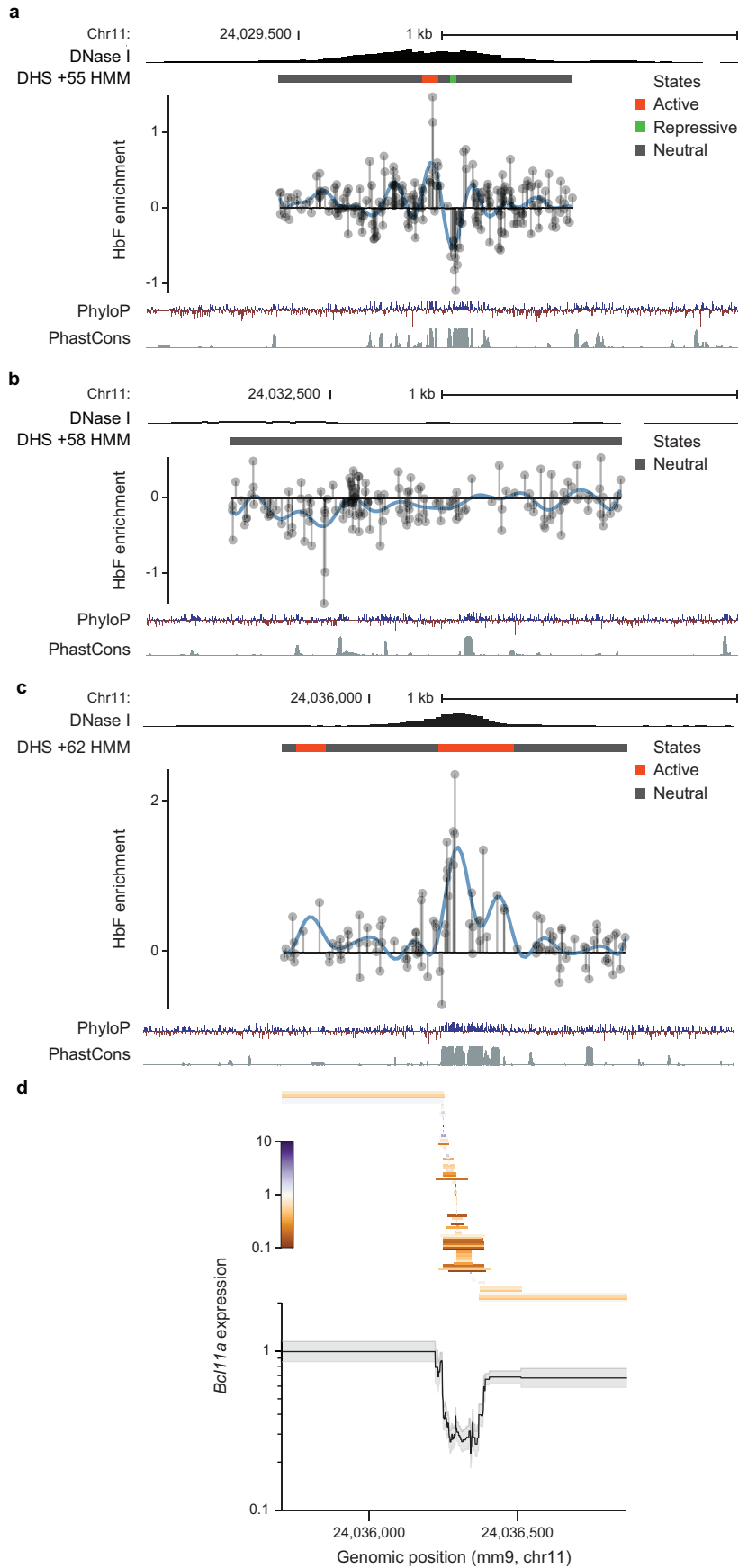
Extended Data Figure 6 | Tiled pooled *in situ* CRISPR-Cas9 *Bcl11a* enhancer screen. **a**, Schematic of the mouse *Bcl11a* locus (mm9, transcription from left to right) with erythroid chromatin marks (top, dark blue H3K27ac from ref. 55; middle, light blue H3K27ac from ref. 56; and bottom, black DNase I from ref. 28) and regions of primary sequence homology to the human DHSs displayed. *y* axes for H3K27ac tracks are both scaled to maximum 3.5 reads per million. Composite enhancer as previously defined²⁸. **b**, Ranked enhancers in mouse erythroid precursors by H3K27ac signal intensity^{55,56}, with super-enhancers shaded. Super-enhancer associated genes are indicated by Venn diagram. **c**, Strategy to knock-in by homology-directed repair the fluorescent protein mCherry into the mouse embryonic globin *Hbb-y* locus (encoding the $\epsilon\gamma$ embryonic globin chain). **d**, Distribution of NGG and NAG PAM sgRNAs mapped to genomic cleavage position with vertical lines

representing cleavage sites for sgRNAs mapped to plus and minus strands. **e**, Distance to adjacent genomic cleavage position for NGG (left) and NAG (right) PAM sgRNAs. **f**, Representation of the 1,271 NGG and NAG sgRNAs within the plasmid pool by deep sequencing. The median was 735 normalized reads and the 10th and 90th percentiles (indicated by the vertical dotted lines) ranged from 393 to 1,240 normalized reads. **g**, Library composition by target sequence and PAM restriction. **h**, mCherry expression upon exposure to Cas9 and an individual NGG sgRNA targeting *Bcl11a* exon 2 in MEL $\epsilon\gamma$:mCherry reporter cells. **i**, $\epsilon\gamma$:mCherry sort of library transduced cells. **j**, Control sgRNA enrichment. Boxes demonstrate 25th, median and 75th percentiles and whiskers minimum and maximum values. **** $P < 0.0001$. **k**, Enrichment scores of NGG sgRNAs between four biological replicates.



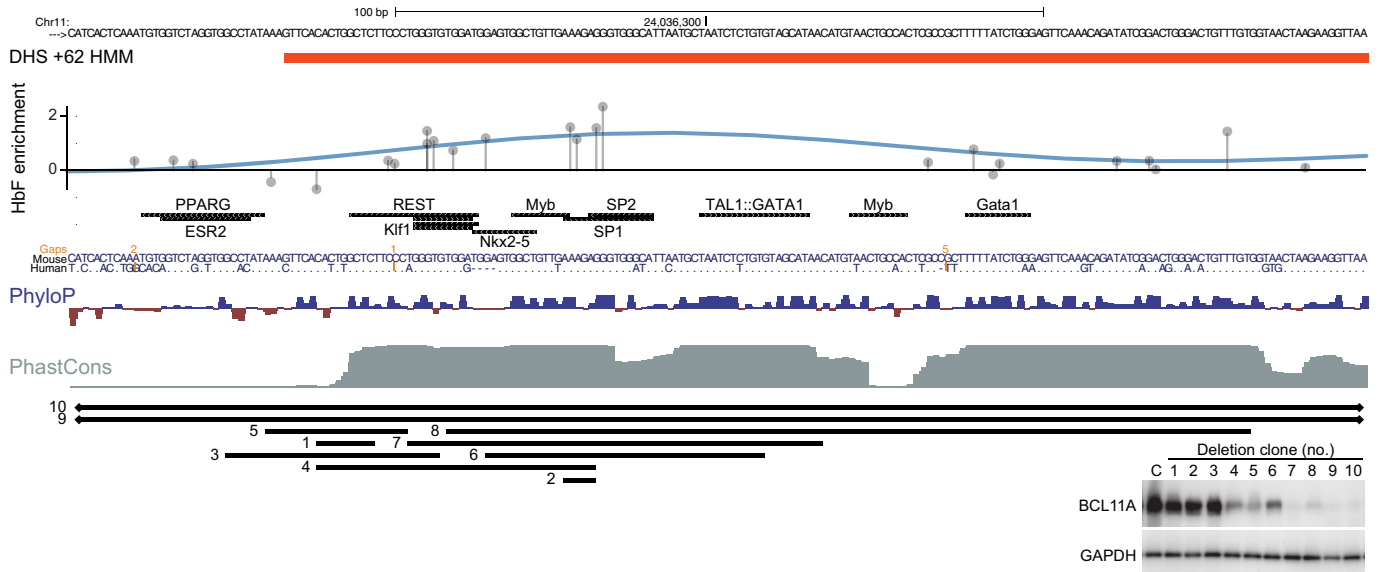
Extended Data Figure 7 | *Bcl11a* enhancer screen analyses. **a**, NGG sgRNA representation in plasmid pool and cells at conclusion of experiment (left), and in $\epsilon y:mCherry$ -high and $\epsilon y:mCherry$ -low pools (right), with dotted lines at $x = y$ and $x = 8y$. **b**, Quantile–quantile plots of NGG sgRNA ϵy enrichment

scores. **c**, Cellular dropout scores of NGG sgRNAs relative to genomic cleavage position and repetitive elements. Non-targeting sgRNAs are pseudo-mapped with 5 bp spacing. **d**, Correlation between cellular dropout and ϵy enrichment scores.



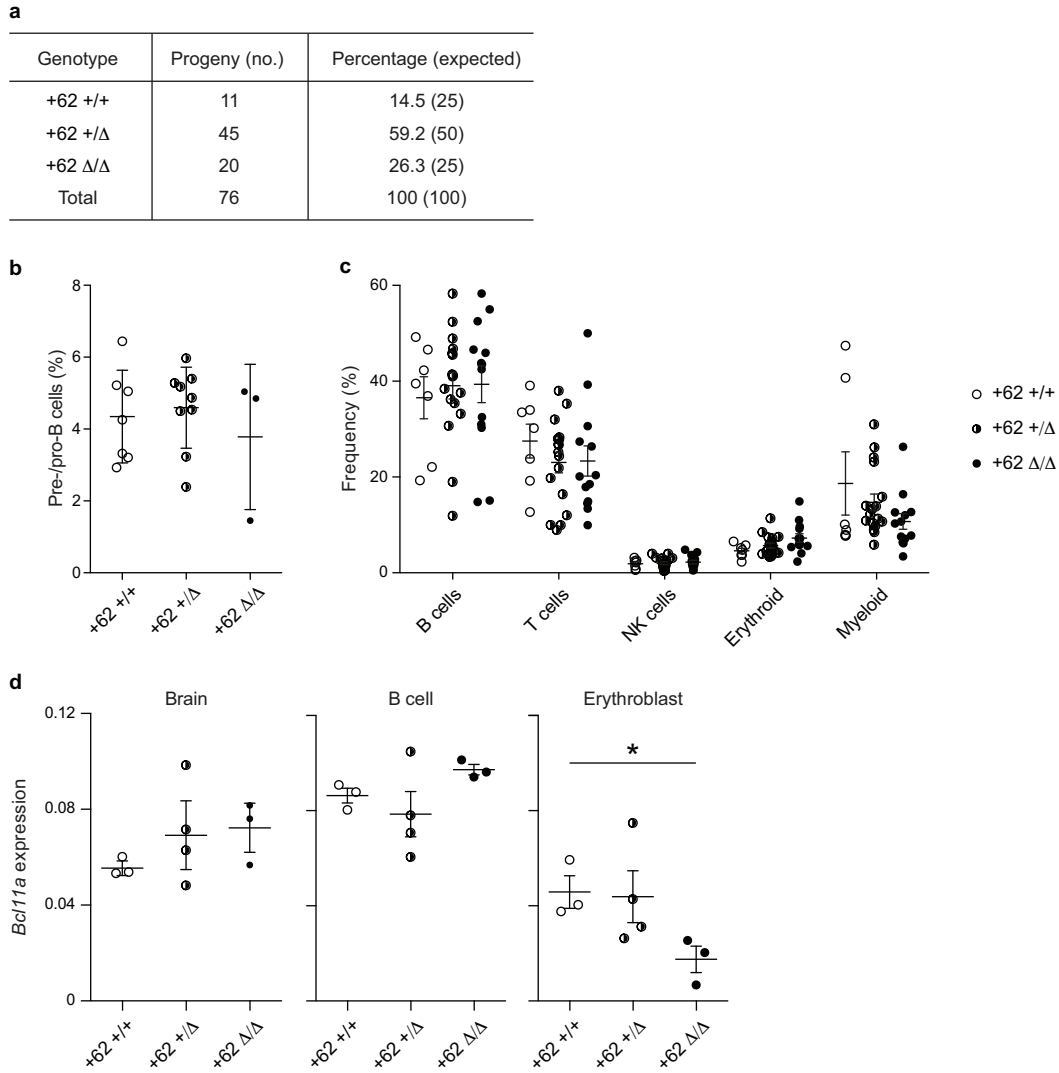
Extended Data Figure 8 | Functional sequences at the *Bcl11a* erythroid enhancer. **a–c**, HMM segmentation of active functional states at m+55, m+58 and m+62 orthologues. HbF enrichment scores are shown as grey lines and circles with the blue line representing smoothed enrichment score. DNase I sequencing from mouse fetal liver erythroid precursors²⁸. PhyloP (scale from -3.3 to 2.1) and PhastCons (from 0 to 1) estimates of evolutionary conservation among 30 vertebrates are shown. **d**, Top: *Bcl11a* expression determined by RT-qPCR displayed as a heat map in 108 hemizygous m+62

orthologue deletion clones ordered by genomic position of deletion midpoint. Each bar demonstrates the genomic position of the deletion breakpoints and the associated colour demonstrates the level of *Bcl11a* expression. Bottom: *Bcl11a* expression determined by RT-qPCR in 108 hemizygous m+62 orthologue deletion clones. Per nucleotide mean effect size was calculated as the mean fold change in *Bcl11a* expression from all clones in which that nucleotide was deleted. Grey shading represents one s.d. The *Bcl11a* expression data are shown with same x axis as in panel c immediately above.



Extended Data Figure 9 | Evaluation of the m+62 functional core. 200 bp at the functional core of the m+62 orthologue defined by HMM state. Enrichment scores are shown as grey lines and circles with the blue line representing smoothed enrichment score. JASPAR motifs ($P < 10^{-4}$) are depicted with selected motifs annotated by transcription factor name on the basis of known erythroid-specific function or genomic position. Orthologous

human sequences are listed. PhyloP (scale from -3.3 to 2.1) and PhastCons (from 0 to 1) estimates of evolutionary conservation among 30 vertebrates are shown. Individual numbered hemizygous deletion clones with indicated breakpoints were evaluated by BCL11A immunoblot (C, control). Clones 9 and 10 encompass the entire m+62 orthologue.



Extended Data Figure 10 | Requirement of *Bcl11a* erythroid enhancer during murine ontogeny. **a**, Progeny of heterozygous *Bcl11a* m+62 orthologue deletion intercrosses as compared to expected Mendelian ratio. **b**, Fraction of fetal liver comprised of B-cell progenitors at E16.5 from various genotypes. **c**, Peripheral blood analysis from 4-week-old mice to examine the

frequency of various circulating haematopoietic lineages in *Bcl11a* m+62 orthologue deletion wild-type, heterozygous, and homozygous mice. **d**, *Bcl11a* expression in β-YAC/+62 deletion mice (each symbol represents the mean expression from technical replicates from an individual mouse). **P* < 0.05; error bars represent s.e.m.