Supplementary Materials
Materials and Methods
Supplementary Text
Figs. S1-S9
Tables S1-S6
References (39-55)

Materials and Methods

Cell culture

Human $CD34^+$ cells from mobilized peripheral blood of healthy donors were obtained from Centers of Excellence in Molecular Hematology at Yale University, New Haven, Connecticut and Fred Hutchinson Cancer Research Center, Seattle, Washington. The cells were subject to *ex vivo* erythroid maturation with a two-phase serum-free liquid culture protocol as previously described *(39)*. Peripheral blood mononuclear cells (PBMCs) were obtained from healthy donors from Boston Children's Hospital. Erythroid differentiation from PBMCs was performed as previously described *(40)*. Mouse erythroleukemia (MEL) cells and 293T cells were cultured as previously described *(39)*. Stably v-Abl transformed pre-B lymphocyte murine cells (derived as described *(41)*) were cultured in RPMI plus 2% penicillin-streptomycin, 15% FCS, 2% HEPES, 1% non-essential amino acids, 1% sodium pyruvate, 1% L-glutamine and 100 μM β-mercaptoethanol.

ChIP and DNase I sensitivity

Chromatin immunoprecipitation and massively parallel sequencing were performed as described *(39)*. The following antibodies were used: H3K27me3 (Millipore, 07-449), H3K4me3 (Millipore, 04-745), H3K4me1 (Abcam, ab8895), H3K27ac (Abcam, ab4729), RNA Polymerase II (PolII, Santa Cruz, sc-899), GATA1 (Abcam, ab11852) and TAL1 (Santa Cruz, sc-12984). DNase I cleavage density performed and analyzed as previously described *(42)*. For ChIP-qPCR, relative enrichment was determined by comparing amplification of ChIP material to 1% input chromatin by the ΔCt method. Loci previously reported to be occupied and non-occupied by GATA1 and TAL1 were used as positive and negative controls respectively *(39)*.

Chromosome conformation capture (3C)

3C assay was performed as previously described *(39)* except as below. Nuclei from formaldehyde cross-linked primary human erythroid precursors were digested with HindIII prior to ligation and reversal of cross-links. Quantitative real-time PCR was performed using iQ SYBR Green Supermix (Bio-Rad, 170-8880). A fragment containing the *BCL11A* promoter was used as the anchor region. To correct for amplification efficiency of different primers, a control template was prepared by digesting and ligating an equimolar mixture of two bacterial artificial chromosomes (BACs) comprising the complete human *BCL11A* locus (RP11-606L8 and RP11-139C22) and one the human β-globin cluster (CTD-3055E11). An interaction between fragments in HS1/HS2 and HS3 of the human β-globin locus control region (LCR) served as a positive control. Interaction frequency was expressed as amplification relative to the known LCR interaction, normalized to the BAC control template.

Fine-mapping BCL11A *locus*

Markers (all coordinates hg19) were selected from within the three *BCL11A* intron-2 DHSs +62 (chr2:60,717,492-60,718,860), +58 (chr2:60,721,411-60,722,674) and +55 (chr2:60,724,802-60,726,084). 21 markers were identified from the 1000 Genomes Project database using the North European (CEU), Nigerian (YRI) and African-American (ASW) reference populations (Table S1). 38 additional variants were present in

dbSNP135 (Table S1). We sequenced by Sanger chemistry the three DHS intervals in the DNA of 52 and 36 sickle cell disease (SCD) patients from the CSSCD cohort with high (> 8%) and low (< 2%) HbF levels, respectively. From this sequencing effort, seven novel sequence variants were identified (Table S2). Because most markers cluster in small genomic intervals, it was not possible to design genotyping assays for some of them. Of 66 non-redundant variants identified in the three DHSs, genotyping assays for 40 markers were performed in 1,263 participants from the CSSCD, an African-American SCD cohort for which genomic DNA (gDNA) is available and HbF levels are known *(21)*. Markers were genotyped using the Sequenom iPLEX platform. Individuals and DNA sequence variants with a genotyping success rate < 90% were excluded. Overall genotype concordance estimated from triplicates was 100%. SNPs passing quality control (QC; $n = 38$) are listed in Tables S1 and S2, and shown schematically in Figs. 2A and S2A below the three DHSs. A substantial fraction of the genotyped SNPs are rare in the reference populations so not surprisingly monomorphic in the CSSCD ($n = 18$). After QC, 1,178 individuals and 20 polymorphic SNPs remained for the analysis. HbF levels were modeled as previously described *(9, 43)*. Association and conditional analyses of single variants (MAF > 1%) were performed with PLINK *(44)* using linear regression under an additive genetic model. Analysis of common variants (MAF > 1%) revealed that rs1427407 in DHS +62 had the strongest association to HbF level ($P = 7.23 \times 10^{-50}$; Figs. 2A and S2B). Conditional analysis demonstrated that after conditioning on rs1427407 and rs7606173, no more SNPs were significant (Fig. S2B). Adjusting for principal components (PCs) on 855 individuals for whom genome-wide genotyping data was available to account for admixture and other confounders yielded similar results.

For rare and low-frequency variants (MAF < 5%), we performed set-based analyses using each of the three DHSs +62, +58 and +55 as the testing unit. For these analyses, we used the sequence kernel association test (SKAT-O) program *(45)* with default parameters. We selected the 5% threshold for MAF in order to maximize statistical power given our limited sample size, but note that markers with a MAF between 1% and 5% were also analyzed in the single variant analyses presented above. This variant overlap is accounted for using conditional analyses with the common variants independently associated with HbF levels. Two sets were found to be statistically significant, namely DHS +62 and DHS +55, but after conditioning on rs1427407 and rs7606173, results were no longer statistically significant, suggesting weak LD between the rare/low-frequency variants and the common SNPs (Table S3). We did not find evidence that rare and low-frequency sequence variants within the *BCL11A* DHSs influence HbF levels in SCD subjects, despite Sanger re-sequencing these DHSs in 88 subjects with extreme HbF phenotype.

The rs1427407–rs7606173 haplotype frequencies in CSSCD are: T–G 24.5%, T–C 0.085%, G–C 42.3%, G–G 33.1%. The mean HbF level is 4.05% (SD 3.10) in 213 rs1427407–rs7606173 G–C individuals, 7.08% (SD 4.50) in 254 rs1427407–rs7606173 T–G/G–T heterozygotes and 11.21% (SD 4.37) in 60 rs1427407–rs7606173 T–G individuals (Fig. S3). For comparisons of HbF levels between genotypes, the *P*-values were determined by one-tailed student t-tests.

Molecular haplotyping

For two heterozygous SNPs on the same chromosome, there are two possible phases: A–B/a–b (model 1) and A–b/a–B (model 2). For SNPs within the 12-kb *BCL11A* intron-2 fragment +52.0-64.4 kb, phase was determined by cloning PCR products and determining co-distribution of SNP alleles. To determine phase of rs7569946 and rs1427407 alleles (separated by 30.1 kb on chromosome 2), emulsion fusion PCR was performed as previously described *(24, 25)* with minor modification. Fusion PCR brings two regions of interest, from separate parts of the same chromosome, together into a single product. By carrying out the reaction in emulsion with aqueous microdroplets surrounded by oil, the preponderance of amplicons are derived from a single template molecule. Genomic DNA from individuals known to be doubly heterozygous for rs7569946 and rs1427407 served as template in the following 100 μl reaction (with final concentrations listed): KOD Hot Start DNA Polymerase (14 U, Novagen, 71086), KOD buffer (1X), MgSO4 (1.5 mM), dNTPs (0.2 mM each), rs7569946-F and rs1427407-R primers (1 μM each), rs7569946-R primer (30 nM), rs7569946-R-revcomp-rs1427407–F bridging inner primer (30 nM), gDNA (200 ng). The 100 μl aqueous reaction was added dropwise with stirring to 200 μl oil phase to create an emulsion. Two 125 μl aliquots of emulsion were amplified under the following conditions: 95 degrees 2 minutes; 45 cycles of 95 degrees 20 seconds, 60 degrees 10 seconds, 70 degrees 30 seconds; 70 degrees 2 minutes. Hexane extracted fusion PCR product was subject to nested PCR in 25 μl as follows: KOD Hot Start DNA Polymerase (0.5 U), KOD buffer (1X), MgSO4 (1.5 mM), dNTPs (0.2 mM each), rs7569946-nested-F and rs1427407-nested-R primers (300 nM each), extracted fusion PCR product (75 nl); 95 degrees 2 minutes; 35 cycles of 95 degrees 20 seconds, 60 degrees 10 seconds, 70 degrees 30 seconds; 70 degrees 2 minutes. The nested product was confirmed by agarose gel electrophoresis to constitute a single band of expected size. The purified product was cloned with the Zero Blunt PCR Cloning kit (Life Technologies, K2700-20). The Sanger sequencing of fusion amplicons enumerated clones of 4 possible sequences: A–B, a–b, A–b and a–B. The likelihood of each phase was calculated based on a multinomial distribution assumption (Table S4). The likelihood ratio for the two configurations was calculated as a measure for the statistical significance of the data fitting haplotype model 1 (as compared to model 2). A ratio approaching infinity suggests model 1, a ratio of 1 suggests equipoise and a ratio approaching zero suggests model 2.

Pyrosequencing

Healthy $CD34^+$ cell donors were screened to identify five donors heterozygous for rs1427407. These $CD34^+$ cells were subject to *ex vivo* erythroid differentiation. Chromatin was isolated and ChIP performed with GATA1 and TAL1 antibodies. Input chromatin as compared to GATA1 or TAL1 precipitated material was subject to pyrosequencing to determine allelic balance of rs1427407. Healthy $CD34^+$ donors were screened to identify three donors heterozygous for the rs1427407–rs7606173 G–C/T–G haplotype. These $CD34^+$ cells were subject to *ex vivo* erythroid differentiation. Complementary DNA (cDNA) and gDNA were subject to pyrosequencing to determine allelic balance of rs7569946.

PCR conditions as follows: 2X HotStarTaq master mix (Qiagen, 203443), $MgCl_2$ (final concentration 3 mM), template DNA (0.1-1 ng) and SNP-specific forward and reverse-biotinylated primers (200 nM each). PCR cycling conditions were: 94ºC 15 min;

45 cycles of 94ºC 30 s; 60ºC 30 s; 72ºC 30 s; 72ºC 5 min. One primer of each pair was biotinylated. The PCR product strand containing the biotinylated primer was bound to streptavidin beads and combined with a specific sequencing primer. The primed single stranded DNA was sequenced and genotype analyzed using the Pyrosequencing PSQ96 HS System (Qiagen Pyrosequencing) following the manufacturer's instructions.

Transgenic mice

The enhancer reporter construct pWHERE-Dest was obtained from Dr. William Pu. Modified from pWHERE (Invivogen, pwhere) as previously described *(46)*, the construct has murine *H19* insulators flanking a CpG-free *lacZ* variant driven by a minimal *Hsp68* minimal promoter with a Gateway destination cassette at the upstream MCS. Enhancer fragments were amplified from mouse gDNA, recombined into pDONR221 vector (Invitrogen, 12536-017) by BP clonase (Invitrogen, 11789020) and recombined into pWHERE-Dest vector with LR clonase (Invitrogen, 11791020). Plasmids were digested with PacI to remove vector backbone. The *lacZ* enhancer reporter fragments were purified by gel electroelution and then concentrated using Wizard DNA Clean-Up System (Promega, A7280). Transgenic mice were generated by pronuclear injection to FVB fertilized eggs. Approximately 10 ng/μl of DNA solution was used for series of injections. CD-1 females were used as recipients for injected embryos. 10.5 to 14.5 dpc embryos were dissected from surrogate mothers with whole-mount and tissue X-gal staining performed as previously described *(47)*. X-gal stained cytospins were counterstained with Nuclear Fast Red (Vector Laboratories, H-3403). Tails used for PCR genotyping. Animal procedures were approved by the Children's Hospital Instititutional Animal Care and Use Committee.

Human erythroid precursor enhancer assay

Genomic DNA fragments containing putative enhancer elements were cloned into pLVX-Puro (Clontech, 632164) upstream of a minimal *TK* promoter and *GFP* reporter gene as described *(39)*. 293T cells were transfected with FuGene 6 reagent (Promega, E2691) according to manufacturer's protocol. The media was changed after 24 hours to SFEM medium supplied with 2% penicillin-streptomycin, and after 36 hours, supernatant was collected and filtered. $CD34^+$ cell-derived erythroid cultures were transduced with lentivirus on expansion days 4 and 5 by spin-infection as previously described *(39)*. Cells were resuspended in erythroid differentiation media 24 hours after the second infection. Selection with puromycin 1 μg/ml commenced 48 hours after infection. Transduced cells were analyzed after five days in differentiation media by flow cytometry for GFP mean fluorescence intensity.

Flow cytometry

Live cells were gated by exclusion of 7-aminoactinomycin D (7-AAD, BD Pharmingen, 559925). Bone marrow (for erythroblast) and spleen (for lymphocyte) suspensions were isolated from young adult transgenic mice. Following hypotonic lysis of mature red blood cells, live cells ($7\text{-}AAD^-$) sorted based on staining with CD71-biotin (BD, 557416), streptavidin-APC (BD, 554067), Ter-119-PE (BD, 553673), CD19-APC (BD, 550992) or CD3-PE (BD, 100308). $CD71^+Ter119^+$, $CD19^+$ and $CD3^+$ sorted populations used for cytospin and RNA isolation.

TALEN-mediated chromosomal deletion

Transcription activator-like effector nucleases (TALENs) were designed to generate cleavages at mouse *Bcl11a* intron-2 at sites +50.4 kb (termed 5' site) and +60.4 kb (3' site) relative to the TSS. The TALENs recognize the following sequences: CTTAAGGCAAGAATCACT (5' left), CCATGCCTTTCCCCCCCT (5' right), GAGTTAAAATCAGAAATCT (3' left), CTGACTAATTGATCAT (3' right). TALENs were synthesized with Golden Gate cloning *(48)* using the NN RVD to recognize G. The synthesized DNA binding domains were cloned into pcDNA3.1 (Invitrogen, V790-20) with the FokI nuclease domain, Δ152 N-terminal domain and +63 C-terminal domain previously described *(49)*. 2.5 μg of each of the four TALEN plasmids with 0.5 μg pmaxGFP (Lonza) were delivered to 2 x 10^6 MEL or pre-B cells by electroporation per manufacturer's protocol (Lonza, VCA-1005). GFP-positive cells were sorted by flow cytometry after 48 hours. Cells seeded by limiting dilution in 96-well plates to isolate individual clones. Clones screened by PCR of gDNA to detect the amplification of a short product from upstream of the 5' site and downstream of the 3' site indicating deletion of the intervening segment. Monoallelic deleted clones were subject to a second round of TALEN-mediated deletion to obtain biallelic deleted clones. Clones with biallelic deletion were identified by detecting absence of amplification from within the deleted fragment. Deletion frequency was approximately one in 50 alleles. Deletion was validated with Southern blotting. Genomic DNA was digested with BmtI; a 561-bp probe (amplified from gDNA upstream of the 5' site) hybridizes to a 3.6 kb fragment from the wild-type allele and a 8.9 kb fragment from the Δ50.4-60.4 deleted allele.

RT-qPCR and immunoblotting

RNA isolation with RNeasy columns (Qiagen, 74106), reverse transcription with iScript cDNA synthesis kit (Bio-Rad, 170-8890), qPCR with iQ SYBR Green Supermix (Bio-Rad, 170-8880) and immunoblotting performed as described *(39)*. For the mouse β-globin cluster genes, a common primer pair recognizes the adult β-globins β2 and β1 while independent primers recognize the embryonic β-globins εy and βH1. The following antibodies were used for immunoblotting: BCL11A (Abcam, ab19487), GAPDH (Santa Cruz, sc-25778).

Supplementary Text

HbF-associated variation at BCL11A

Six GWAS of HbF level (or the highly correlated trait F-cell number) have been conducted in individuals of European, African and Asian descent, each identifying trait-associated variants within *BCL11A (7-12)*. The same variants are associated with the clinical severity of SCD and β-thalassemia *(9, 10, 50)*, consistent with HbF as a major modifier of these disorders. Variation at *BCL11A* is estimated to explain ~15% of the trait variance in HbF level *(7, 12, 43)*. Four different SNPs have been identified as most highly associated with the trait (rs1427407 *(7)*, rs11886868 *(8)*, rs4671393 *(9)* and rs766432 *(10-12)*); these sentinel SNPs cluster within 3 kb of each other in *BCL11A* intron-2 (Figs. 1A and S2A). Haplotypes including the sentinel SNPs appear to better explain the HbF association than any individual SNP *(12, 43)*. Fifty SNPs at the *BCL11A* locus and twenty-seven SNPs within intron-2 have been associated with HbF level with genome-wide significance ($P < 5 \times 10^{-8}$). Despite large-scale resequencing efforts, coding variants of *BCL11A* have not been described *(43)*.

Previously, we used the CSSCD to fine-map the association signal with HbF at the *BCL11A* locus and reported a strong association with rs4671393 *(43)*; in that study, rs1427407 was imputed. Two additional SNPs, rs766432 and rs11886868 have also been identified in prior studies as sentinel SNPs most highly trait-associated *(8, 10, 11, 51)*. In a subset of individuals ($n = 728$) for which genotypes at all four sentinel SNPs were available, the association result was not significant at rs4671393, rs766432 or rs11886868 following conditioning on genotypes at rs1427407; conversely, the association remained highly significant for rs1427407 upon conditioning on rs4671393, rs766432 or rs11886868 (Fig. S2C). Therefore, rs1427407 is the SNP most strongly associated with HbF level within the erythroid DHSs and better accounts for the trait association than other previously described sentinel SNPs.

Conditional analysis demonstrated associations that remained significant after conditioning on rs1427407. The most significant residual association was for rs7606173 in DHS +55 ($P = 9.66 \times 10^{-11}$); rs7599488 in DHS +62, which we had previously reported *(43)*, was only slightly less significant ($P = 2.43 \times 10^{-10}$) (Fig. S2B). Analysis of rare DNA sequence variants within the three DHSs did not yield additional independent HbF-associated signals (Table S3).

Allele-specific TF binding and BCL11A expression

Allele-specific biochemical studies were performed using informative heterozygotes to control for trans-acting differences between samples and to ensure equal abundance of both alleles, substantiated by equal representation of alleles in paired gDNA (Figs. 2B and 2C). rs1427407 is found directly at the center of a GATA1 and TAL1 binding peak at DHS +62 (Fig. 1B). In our ChIP assays, chromatin was sonicated to approximately 500-bp fragments. The five primary human erythroid precursor samples heterozygous for rs1427407 used for ChIP-qPCR were Sanger sequenced at the erythroid DHSs. The only other heterozygous SNP within 500-bp of rs1427407 in any of these samples was rs7599488 (304-bp 3' of rs1427407) which was heterozygous in just two of the five samples. This SNP does not fall within GATA1 or TAL1 binding motifs. It therefore appears unlikely that another SNP within DHS +62 could account for the observed allele-specific TF binding.

Association between BCL11A expression and HbF level

Our studies provide an estimate of the change in BCL11A expression that may result in a clinically meaningful increase in HbF level. Among a limited set of human lymphoblastoid cell lines we previously reported correlation of the high HbF-associated A-allele of rs4671393 with reduced BCL11A expression *(13)*. Extension of these experiments to a larger collection of genotyped lines failed to confirm this observation. Hence, we hypothesized that the HbF-associated rs1427407–rs7606173 haplotype might influence BCL11A expression in an erythroid-specific context, a possibility consistent with the DNase I sensitivity findings. BCL11A mRNA expression in primary erythroid precursors differed by 1.7-fold between the high-HbF rs1427407–rs7606173 T–G and low-HbF G–C haplotypes (Fig. 2C); correspondingly, median HbF levels were 10.6% and 3.1% in T–G and G–C homozygotes, respectively (Fig. S3). Of note, the results demonstrating allele-specific expression of BCL11A in primary human erythroid cells were observed in cells heterozygous for the rs1427407-rs7606173 haplotype, and thus the modest effects on BCL11A expression reflect the combined effects of all functional SNPs within the haplotype. While inheritance of a protective *BCL11A* haplotype is clinically beneficial on a population basis *(9, 10, 50)*, the average level of HbF in T–G homozygotes remains below that required to prevent morbidity from SCD. The sensitivity of HbF level to BCL11A expression, however, predicts that relief of disease severity might require only a modest further reduction in BCL11A expression.

Developmental regulation of globin genes and BCL11A

During human development, yolk sac-derived ε-globin is superseded in the first trimester by fetal liver-derived γ-globin. Following birth, as erythropoiesis shifts from the liver to the bone marrow, γ-globin is gradually silenced and β-globin predominates. Only a single switch in globin gene expression occurs in mouse ontogeny. During this transition, which occurs at mid-gestation, the circulating yolk sac-derived primitive erythrocytes express embryonic-stage globins εy and βH1, whereas the fetal liver definitive erythroblasts express adult-stage globins β1 and β2. Concordant with this developmental switch, BCL11A is expressed in the definitive but not primitive-stage erythroid lineage and required for the change in globin gene expression *(16, 52)*.

In the stable transgenic *BCL11A* +52.0-64.4 reporter lines at 10.5 dpc, lacZ expression was observed only in the fetal liver primordium and not in the circulating blood within the embryo, placenta or yolk sac (Fig. S4A). These results, coupled with the finding of lacZ expression in the 12.5 dpc definitive fetal liver erythroblasts but not yolk sac-derived primitive circulating erythrocytes (Fig. 3B), demonstrate that the BCL11A composite enhancer sequences drive expression in a developmentally-specific pattern concordant with endogenous globin gene switching.

A series of deletion mutants was generated to refine the minimal elements required for erythroid enhancer activity. Sequences containing the central +58 DHS were sufficient for erythroid enhancer activity. Those sequences containing only the flanking +62 or +55 elements were unable to direct erythroid gene expression (Fig. S4B). To test the ability of the DHSs to enhance gene expression in primary human erythroid precursors, we used lentiviral delivery of a *GFP* reporter system as previously described *(39)*. Similarly, the +58 DHS enhanced gene expression in this reporter assay (Fig. S5).

Cell lines with Bcl11a enhancer deletion

To investigate the requirement of the enhancer for BCL11A expression, we generated stable erythroid cells with disruption of the enhancer. Since there are no suitable adult-stage human erythroid cell lines, we turned to the murine system. Mouse erythroleukemia (MEL) cells depend on BCL11A for an adult-stage pattern of globin gene expression *(14)*. We identified an orthologous erythroid composite enhancer at mouse *Bcl11a* intron-2. Like the human GWAS-marked intron-2 *BCL11A* enhancer, these sequences possessed a series of erythroid-specific DHSs. In addition, these sequences were decorated by H3K4me1 and H3K27ac, lacked H3K4me3 and H3K27me3, and occupied by both GATA1 and TAL1 in mouse erythroid chromatin (Fig. S6). Composite regulatory elements including a series of adjacent DHSs have been shown to be critical for gene expression at numerous loci, including among others the β-globin locus control region, α-globin multispecies conserved sequences, and IgH regulatory region *(53-55)*. We observed species-specific unique features of the composite enhancer. For example, we identified the conserved mouse sequences to each of the three human DHSs +62, +58 and +55, and found erythroid DNase I hypersensitivity at the +62 and +55 conserved sequences, however the +58 conserved sequences lacked DNase I hypersensitivity. Of note, there are species-level differences in the timing of BCL11A expression *(13, 16)*, and it is tempting to speculate that differences in enhancer functional components could contribute.

PCR and Southern blotting verified excision of the +50.4-60.4 kb intronic segment of *Bcl11a* in three unique MEL clones and two unique pre-B lymphocyte clones (Fig. S7). Sanger-sequenced breakpoints were characteristic of TALEN-mediated cleavage with subsequent NHEJ repair (Fig. S8). Upon deletion of the intronic segment, we observed dramatic reduction in BCL11A transcript in the MEL cell clones by RT-qPCR, using primer pairs detecting exon junctions upstream, spanning or downstream of the deletion (Fig. 4A).

Supplementary Figures

Fig. S1

Chromatin state and TF occupancy at *BCL11A*.

ChIP-seq from human erythroblasts with indicated antibodies. DNase I cleavage density from indicated human tissues. *BCL11A* transcription from right to left.

Fig. S1

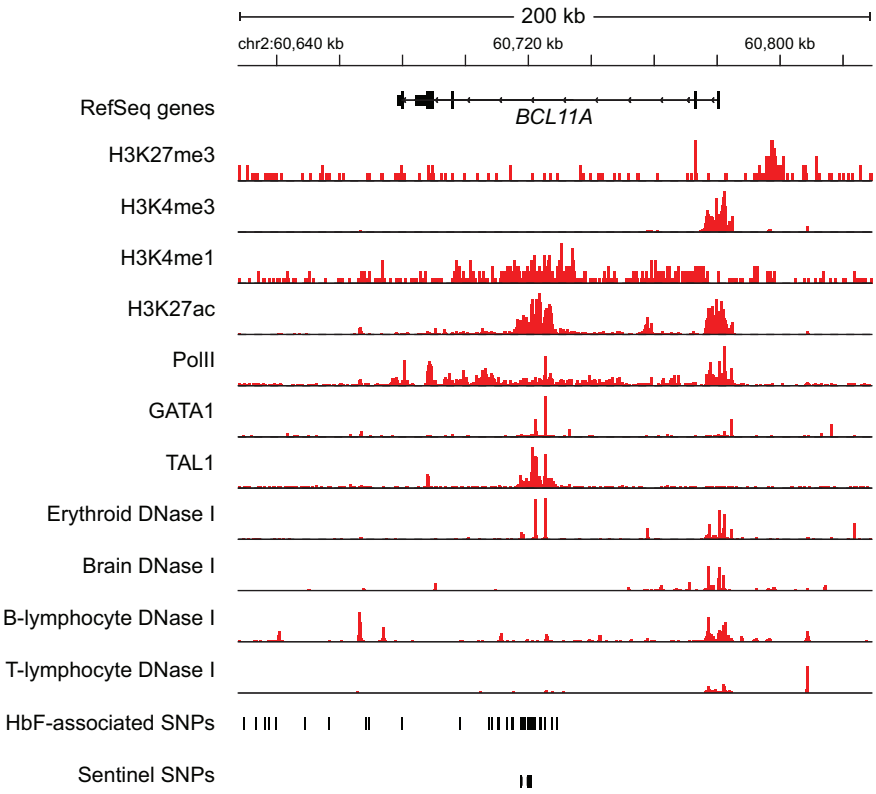


Fig. S2

HbF Association Analyses at *BCL11A*.

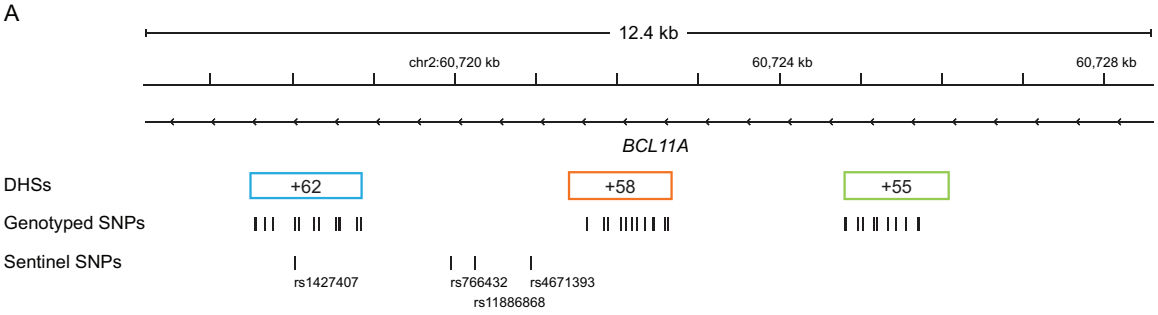
(A) Genotype data obtained in 1,178 individuals from CSSCD for 38 variants within *BCL11A* +62, +58 or +55 DHSs. Sentinel SNPs are those with the highest association to HbF level or F-cell number in prior GWAS *(7-12)*. These SNPs are shown with respect to *BCL11A* intron-2 with the 3 DHSs +62, +58 and +55 indicated.

(B) Association analysis of common (MAF > 1%) SNPs (n = 10) in DHSs +62, +58 or +55 from 1,178 individuals from CSSCD. All SNP coordinates chromosome 2, build hg19. Most highly significant association for each analysis in bold with associated SNP shaded.

(C) Conditional analyses of sentinel SNPs. All four sentinel SNPs were genotyped in 728 individuals from the CSSCD. Most highly significant association for each analysis in bold. It was not possible to calculate P for rs766432 when conditioning on rs4671393 (and vice versa) because these two markers are so strongly correlated (r^2 = 0.997). r^2 = 0.848 between rs1427407 and rs766432; r^2 = 0.709 between rs1427407 and rs11886868; r^2 = 0.850 between rs1427407 and rs4671393; r^2 = 0.761 between rs766432 and rs11886868; r^2 = 0.758 between rs11886868 and rs4671393.

Fig. S2

A



B

DHS	SNP	coordinate	MAF	r^2 with rs1427407	β	*P*	Conditional on rs1427407 β	Conditional on rs1427407 *P*	Conditional on rs1427407 and rs7606173 β	Conditional on rs1427407 and rs7606173 *P*
+62	rs111575474	60,717,559	0.0153	0.006	-0.2624	0.0976	-0.0851	0.5584	0.0486	0.7368
+62	rs74958177	60,717,776	0.0646	0.028	-0.3614	2.79×10^{-6}	-0.1838	0.0104	-0.0832	0.2518
+62	rs1427407	60,718,043	0.2460	-	0.6634	**7.23×10^{-50}**	-	-	-	-
+62	rs112105713	60,718,278	0.0115	0.002	-0.3285	0.0776	-0.2137	0.2097	-0.0859	0.6107
+62	rs7599488	60,718,347	0.3148	0.145	-0.0047	0.9116	0.2622	2.43×10^{-10}	0.0915	0.3547
+62	rs1896293	60,718,848	0.1089	0.025	-0.2623	2.52×10^{-5}	-0.1248	0.0310	0.0241	0.6952
+58	rs6738440	60,722,241	0.2734	0.114	-0.3820	1.25×10^{-18}	-0.1935	5.64×10^{-6}	-0.0223	0.6887
+55	rs147910897	60,724,818	0.0132	0.002	-0.3656	0.0329	-0.2586	0.0995	-0.1575	0.3101
+55	rs148529953	60,724,967	0.0140	0.007	-0.3521	0.0403	-0.1423	0.3668	-0.0098	0.9501
+55	rs7606173	60,725,451	0.4238	0.221	-0.4691	2.86×10^{-34}	-0.2632	**9.66×10^{-11}**	-	-

C

SNP	MAF	β	*P*	Conditional on rs1427407 β	Conditional on rs1427407 *P*	Conditional on rs766432 β	Conditional on rs766432 *P*	Conditional on rs11886868 β	Conditional on rs11886868 *P*	Conditional on rs4671393 β	Conditional on rs4671393 *P*
rs1427407	0.245	0.659	**4.56×10^{-29}**	-	-	0.656	**3.4×10^{-6}**	0.651	**2.8×10^{-10}**	0.666	**1.9×10^{-6}**
rs766432	0.275	0.579	2.48×10^{-24}	0.0036	0.979	-	-	0.479	1.1×10^{-5}	NA	NA
rs11886868	0.302	0.509	2.51×10^{-21}	0.0097	0.918	0.112	0.284	-	-	0.123	0.234
rs4671393	0.274	0.576	4.41×10^{-24}	-0.0065	0.961	NA	NA	0.466	1.7×10^{-5}	-	-

Fig. S3

HbF Level by rs1427407–rs7606173 Haplotype.

HbF level in CSSCD cohort by rs1427407–rs7606173 haplotype. Line indicates median, box interquartile range and whiskers 1st and 99th percentiles.

Fig. S3

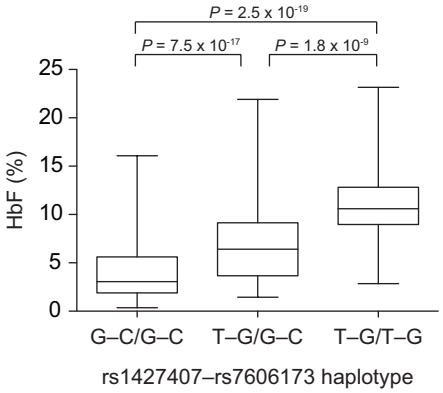


Fig. S4

The GWAS-Marked *BCL11A* Enhancer is Sufficient for Adult-Stage Erythroid Expression.

(A) A 12.4-kb fragment of *BCL11A* intron-2 (+52.0-64.4 kb from TSS) was cloned to a *lacZ* reporter construct. Stable transgenic mouse embryo X-gal stained at 10.5 dpc.

(B) Deletional mapping of *BCL11A* erythroid enhancer. Schematics demonstrate putative enhancers cloned to *lacZ* reporter constructs with included DHSs +62, +58 and/or +55 indicated. Transient transgenic embryos X-gal stained at 12.5 dpc. Embryos genotyped by *lacZ* PCR. The fraction of transgenic embryos with fetal liver X-gal staining is indicated.

Fig. S4

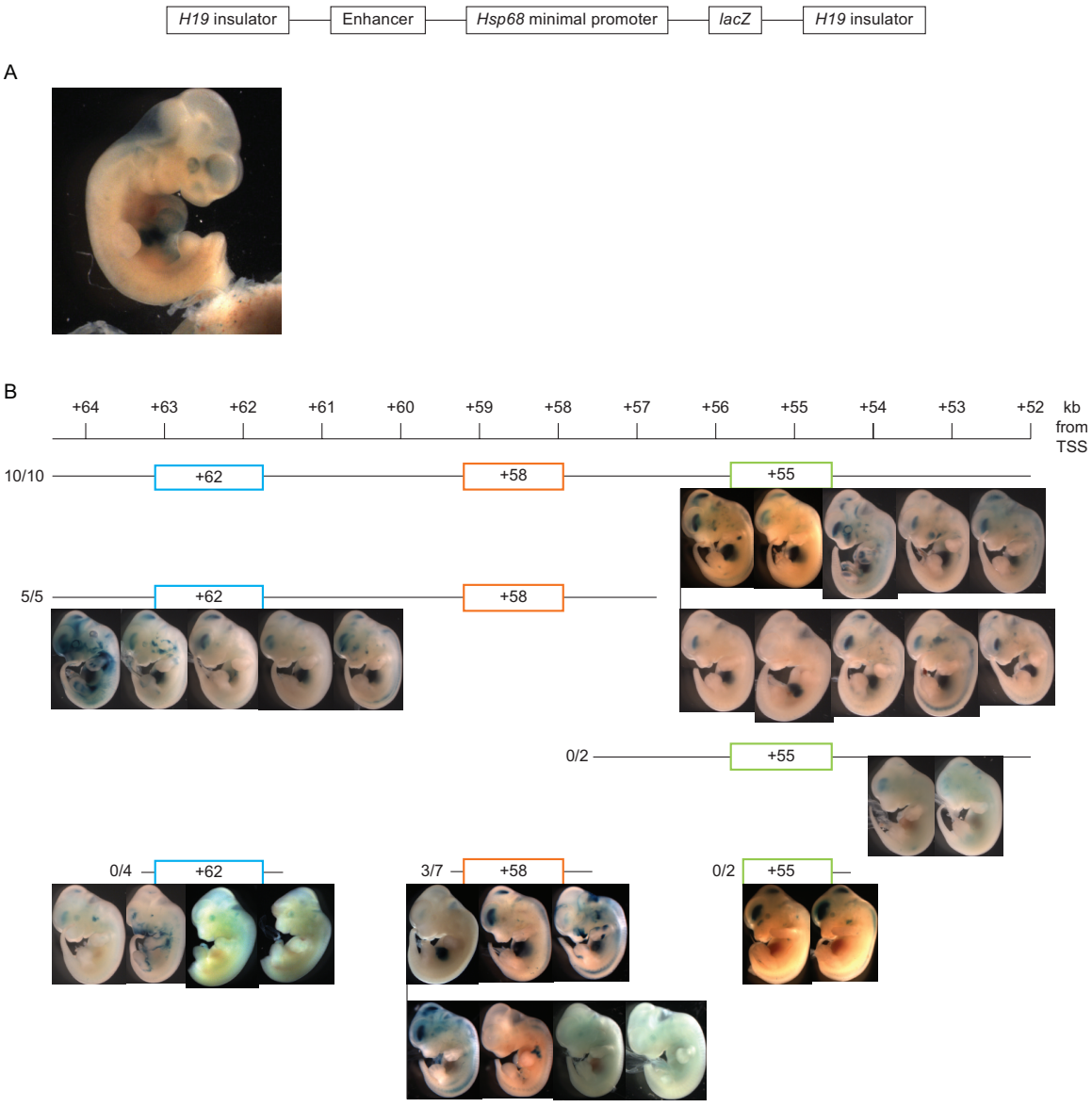


Fig. S5
Enhancer Activity in Primary Human Erythroid Precursors.
Putative *BCL11A* enhancer fragments (1-2 kb) were cloned to *GFP* reporter construct. Fragments cloned from gDNA possessing (+62, +58 and +55) or lacking (+164, +156, +153, +41, +32, -46 and -52) an enhancer chromatin signature. Enhancer reporter constructs delivered by lentiviral vectors to primary human erythroid precursors. Mean GFP fluorescence intensity measured in transduced cells. Error bars indicate s.d.

Fig. S5

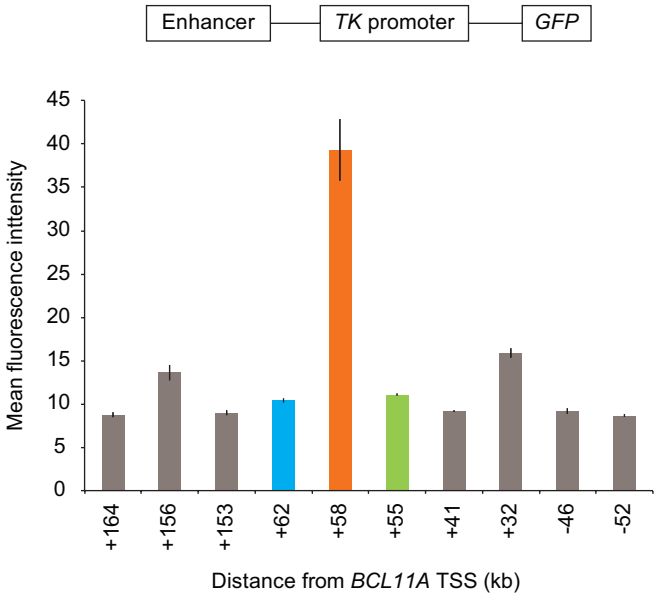


Fig. S6
An Orthologous Erythroid Enhancer Signature at Mouse *Bcl11a*.
DNase I cleavage density from indicated mouse tissues. Histone mark *(27)* and GATA1 and TAL1 *(22)* ChIP-seq tracks obtained from previously published mouse erythroid global chromatin profiling. Conserved sequences to the human *BCL11A* erythroid DHSs +62, +58 and +55 were determined using the liftOver tool of UCSC Genome Browser. Dotted rectangle bounds orthologous erythroid enhancer signature defining Δ50.4-60.4 element targeted for TALEN-mediated deletion.

Fig. S6

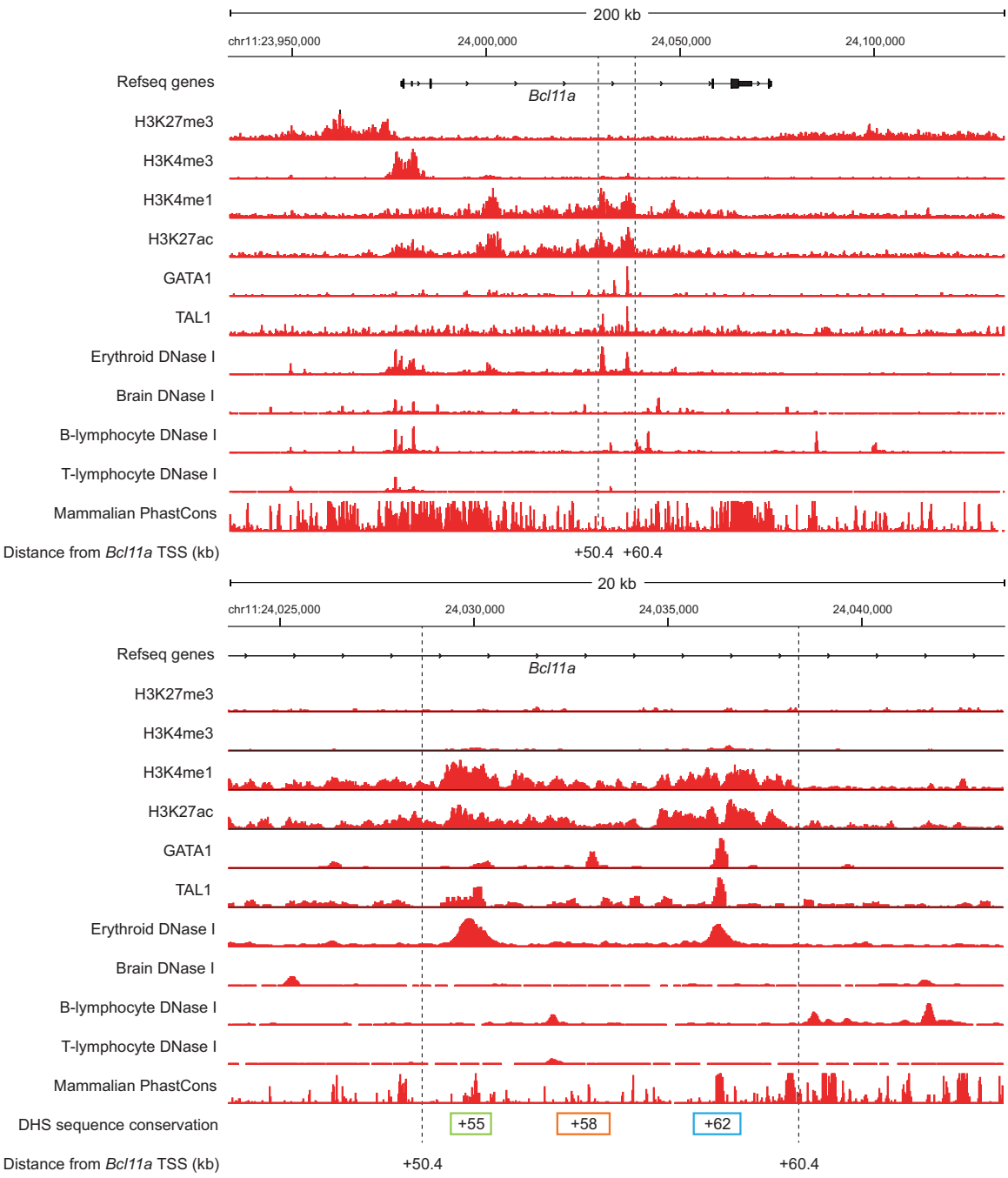

200 kb
chr11:23,950,000
24,000,000
24,050,000
24,100,000
Refseq genes
Bcl11a
H3K27me3
H3K4me3
H3K4me1
H3K27ac
GATA1
TAL1
Erythroid DNase I
Brain DNase I
B-lymphocyte DNase I
T-lymphocyte DNase I
Mammalian PhastCons
Distance from Bcl11a TSS (kb)
+50.4
+60.4
20 kb
chr11:24,025,000
24,030,000
24,035,000
24,040,000
Refseq genes
Bcl11a
H3K27me3
H3K4me3
H3K4me1
H3K27ac
GATA1
TAL1
Erythroid DNase I
Brain DNase I
B-lymphocyte DNase I
T-lymphocyte DNase I
Mammalian PhastCons
DHS sequence conservation
+55
+58
+62
Distance from Bcl11a TSS (kb)
+50.4
+60.4

Fig. S7
TALEN-Mediated Deletion.
(A) Schema of TALEN-mediated genome engineering strategy. Two pairs of TALEN sequence-specific nucleases designed to generate double strand breaks, one at *Bcl11a* +50.4 and the other at +60.4. Alleles were identified that had repaired the two DSBs by NHEJ with excision of the intervening 10 kb segment. Clones were screened by PCR with primers 5', 3' and internal to (del-1 and del-2) the 10 kb deletion as well as by Southern blotting.
(B) Southern blotting of BmtI-digested gDNA from Δ50.4-60.4 clones.
(C) Quantitative PCR from MEL and pre-B lymphocyte Δ50.4-60.4 clones, expressed relative to parental cells. Error bars indicate s.d.

Fig. S7

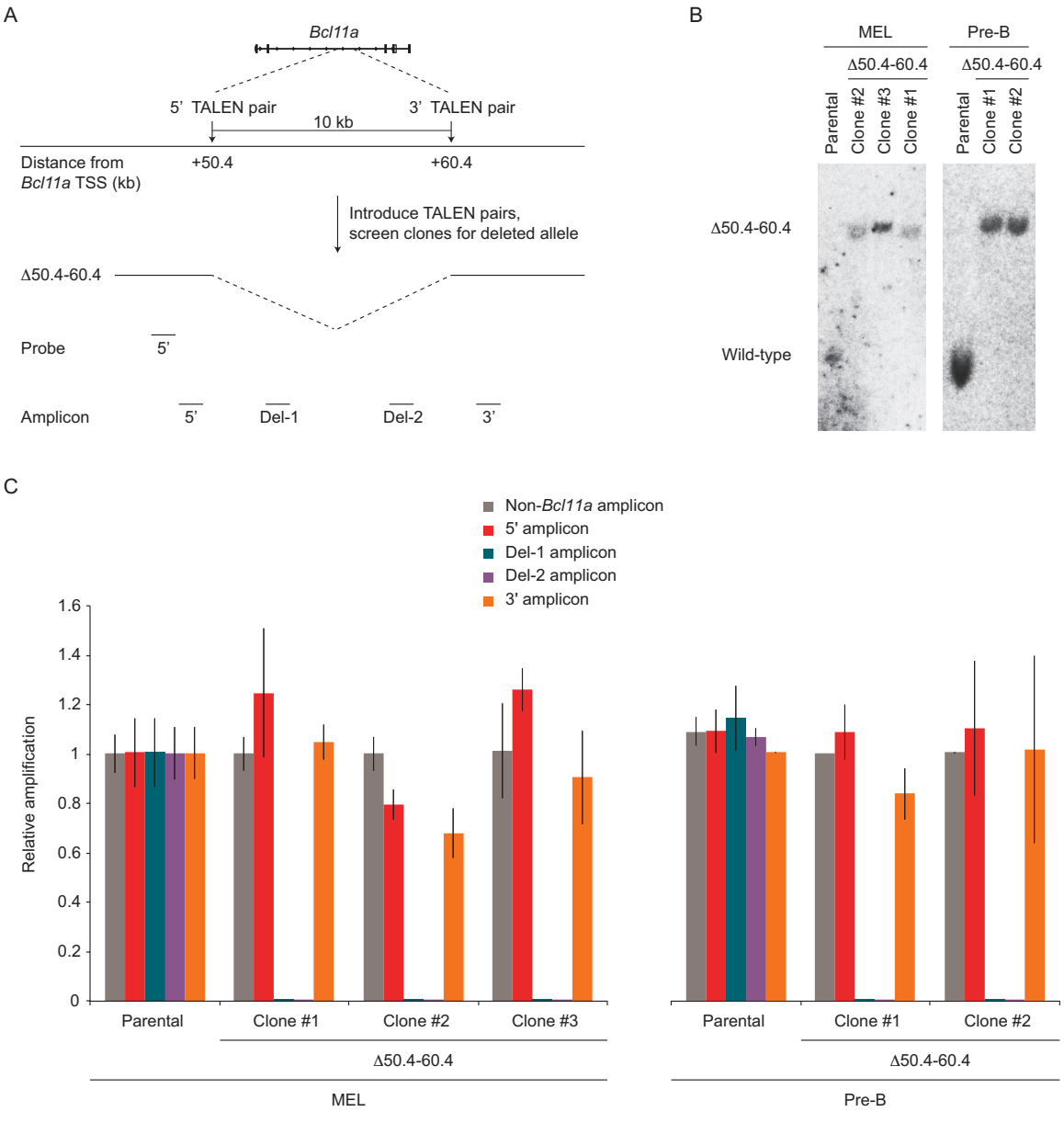

A
Bcl11a
5' TALEN pair
3' TALEN pair
10 kb
Distance from Bcl11a TSS (kb)
+50.4
+60.4
Introduce TALEN pairs, screen clones for deleted allele
Δ50.4-60.4
Probe
5'
Amplicon
5'
Del-1
Del-2
3'
B
MEL
Pre-B
Δ50.4-60.4
Δ50.4-60.4
Parental
Clone #2
Clone #3
Clone #1
Parental
Clone #1
Clone #2
Δ50.4-60.4
Wild-type
C
Non-Bcl11a amplicon
5' amplicon
Del-1 amplicon
Del-2 amplicon
3' amplicon
Relative amplification
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Parental
Clone #1
Clone #2
Clone #3
Δ50.4-60.4
MEL
Parental
Clone #1
Clone #2
Δ50.4-60.4
Pre-B

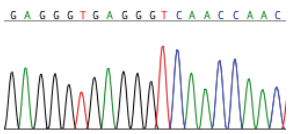
Fig. S8
Bcl11a Enhancer Deletion Breakpoints.
5'-site (+50.4) and 3'-site (+60.4) left and right TALEN recognition sequences with intervening spacers are shown on top. Each breakpoint was PCR amplified and cloned with a primer pair spanning the deletion. Sequences retrieved from each clone displayed on left with chromatograms on right. Some alleles showed evidence of end-joining directly from each cleaved spacer sequence whereas other alleles showed loss of hundreds of additional nucleotides from one or both TALEN recognition sites. One unique allele each was isolated from MEL clone #1 and pre-B lymphocyte clone #2.

Fig. S8

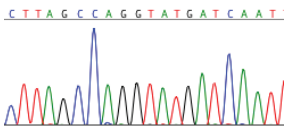
5′ left TALEN 5′ right TALEN
T**CTTAAGGCAAGAATCACT**GCTTAGCCAGGGCCCCA**AGGGGGGGAAAGGCATGG**A

3′ left TALEN 3′ right TALEN
T**GAGTTAAAATCAGAAATCT**CATCTTTCACAGGTT**ATGATCAATTAGTCAG**A

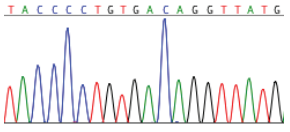
MEL clone #1a
-197 5′TALEN ---------- +314 3′ TALEN



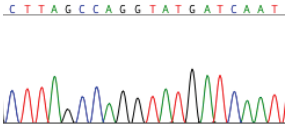
MEL clone #2a
T**CTTAAGGCAAGAATCACT**GCTTAGCCAGG--------T**ATGATCAATTAGTCAG**A



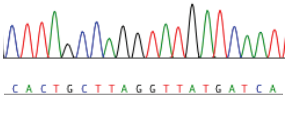
MEL clone #2b
-681 5′TALEN ----------ACAGGTT**ATGATCAATTAGTCAG**A



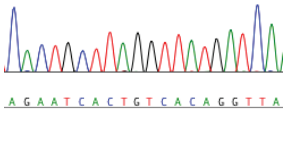
MEL clone #3a
T**CTTAAGGCAAGAATCACT**GCTTAGCCAGG--------T**ATGATCAATTAGTCAG**A



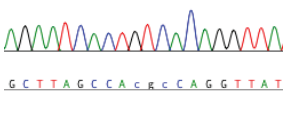
MEL clone #3b
T**CTTAAGGCAAGAATCACT**GCTTAG----------GTT**ATGATCAATTAGTCAG**A



Pre-B clone #1a
T**CTTAAGGCAAGAATCACT**G----------TCACAGGTT**ATGATCAATTAGTCAG**A



Pre-B clone #1b
T**CTTAAGGCAAGAATCACT**GCTTAGCCA-cgc-CAGGTT**ATGATCAATTAGTCAG**A



Pre-B clone #2a
T**CTTAAGGCAAGAATCACT**GCTT---------ACAGGTT**ATGATCAATTAGTCAG**A

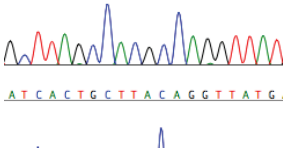
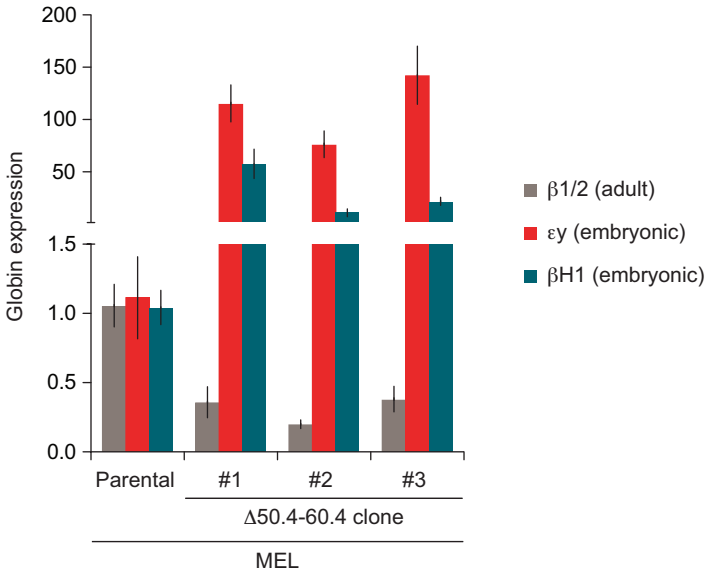


Fig. S9

Globin Gene Expression Upon *Bcl11a* Enhancer Deletion.

Globin gene expression in Δ50.4-60.4 MEL clones by RT-qPCR. Error bars indicate s.e.m. of at least 4 experiments.

Fig. S9



Supplemental Tables

Table S1. SNPs within *BCL11A* DHSs +62, +58 or +55.

Marker	Chr	Pos	Major allele	Minor allele	DHS	Genotyping status	MAF
rs149113684	2	60,717,544	C	A	+62	Monomorphic	0.0000
rs111575474	2	60,717,559	C	T	+62	Polymorphic	0.0157
rs148272134	2	60,717,643	C	A	+62	Failed Assay Design	-
rs182773253	2	60,717,676	A	G	+62	Monomorphic	0.0000
rs188706265	2	60,717,769	C	T	+62	Monomorphic	0.0000
rs74958177	2	60,717,776	A	G	+62	Polymorphic	0.0645
rs1427407	2	60,718,043	G	T	+62	Polymorphic	0.2460
rs35262352	2	60,718,076	A	-	+62	Failed Assay Design	-
rs79781583	2	60,718,077	A	T	+62	Failed Assay Design	-
rs201428515	2	60,718,088	G	A	+62	Monomorphic	0.0000
rs112105713	2	60,718,278	G	A	+62	Polymorphic	0.1145
rs7599488	2	60,718,347	C	T	+62	Polymorphic	0.3149
rs113636744	2	60,718,540	C	T	+62	Polymorphic	0.0042
rs35259900	2	60,718,555	C	T	+62	Failed Assay Design	-
rs111911554	2	60,718,569	A	G	+62	Failed Assay Design	-
rs137943695	2	60,718,574	G	A	+62	Monomorphic	0.0000
rs45579333	2	60,718,599	G	A	+62	Monomorphic	0.0000
rs77876582	2	60,718,639	C	T	+62	Failed Assay Design	-
rs112634025	2	60,718,708	G	A	+62	Failed Assay Design	-
rs45439602	2	60,718,721	G	A	+62	Failed Assay Design	-
rs112387548	2	60,718,762	C	T	+62	Failed Assay Design	-
rs191369155	2	60,718,781	G	A	+62	Failed Assay Design	-
rs6723022	2	60,718,807	A	C	+62	Monomorphic	0.0000
rs11422901	2	60,718,819	G	A	+62	Failed Assay Design	-
rs200632291	2	60,718,824	A	G	+62	Failed Assay Design	-
rs11387709	2	60,718,826	A	-	+62	Failed Assay Design	-
rs1896293	2	60,718,848	G	T	+62	Polymorphic	0.1088
rs71526487	2	60,721,587	T	C	+58	Failed Assay Design	-
rs185151573	2	60,721,639	G	C	+58	Monomorphic	0.0000
rs6721788	2	60,721,846	T	C	+58	Polymorphic	0.0025
rs76033449	2	60,721,900	G	A	+58	Polymorphic	0.0004
rs6706648	2	60,722,040	T	C	+58	Failed Genotyping	-
rs62142615	2	60,722,120	T	C	+58	Polymorphic	0.0081
rs35923541	2	60,722,197	T	-	+58	Monomorphic	0.0000
rs35815093	2	60,722,208	G	-	+58	Failed Assay Design	-
rs147659683	2	60,722,219	G	A	+58	Failed Assay Design	-
rs6738440	2	60,722,241	A	G	+58	Polymorphic	0.2732
rs189178945	2	60,722,449	G	A	+58	Monomorphic	0.0000
rs140819321	2	60,722,465	G	A	+58	Polymorphic	0.0064
rs181895125	2	60,722,609	A	G	+58	Monomorphic	0.0000
rs144676401	2	60,722,634	C	T	+58	Monomorphic	0.0000
rs147910897	2	60,724,818	T	C	+55	Polymorphic	0.0132
rs34322220	2	60,724,831	T	-	+55	Monomorphic	0.0000
rs148529953	2	60,724,967	A	G	+55	Polymorphic	0.0140
rs188426060	2	60,724,989	T	G	+55	Failed Assay Design	-
rs191734859	2	60,724,994	A	G	+55	Failed Assay Design	-
rs45442493	2	60,725,043	G	C	+55	Monomorphic	0.0000
rs59444712	2	60,725,047	T	C	+55	Failed Assay Design	-
rs35173197	2	60,725,052	G	-	+55	Failed Assay Design	-
rs188151753	2	60,725,071	G	A	+55	Failed Assay Design	-

rs181041409	2	60,725,143	C	A	+55	Failed Assay Design	-
rs142174420	2	60,725,169	C	A	+55	Monomorphic	0.0000
rs187333125	2	60,725,342	C	G	+55	Monomorphic	0.0000
rs45566439	2	60,725,384	C	T	+55	Failed Assay Design	-
rs7606173	2	60,725,451	G	C	+55	Polymorphic	0.4235
rs190502487	2	60,725,499	C	T	+55	Failed Assay Design	-
rs151187913	2	60,725,714	G	T	+55	Monomorphic	0.0000
rs113798461	2	60,725,727	T	C	+55	Monomorphic	0.0000
rs181699714	2	60,726,054	G	A	+55	Failed Genotyping	-

SNPs falling within *BCL11A* DHSs +62, +58 or +55. SNPs identified from either dbSNP135 or the 1000 Genomes Project database for YRI, CEU and ASW reference populations. Genotyping status and MAF within the CSSCD listed. Chr, chromosome. Pos, position.

Table S2. Additional Markers Identified by Sanger Resequencing.

Marker	Chr	Pos	Major allele	Minor allele	DHS	Genotyping status	MAF
ss711589103	2	60,717,561	T	A	+62	Polymorphic	0.00085
ss711589106	2	60,718,048	C	G	+62	Failed Assay Design	-
ss711589108	2	60,722,056	G	A	+58	Polymorphic	0.00424
ss711589109	2	60,722,355	C	T	+58	Polymorphic	0.00085
ss711589110	2	60,722,358	C	T	+58	Failed Assay Design	-
ss711589111	2	60,725,211	G	T	+55	Polymorphic	0.00509
ss711589113	2	60,725,564	C	A	+55	Polymorphic	0.00127

88 individuals from CSSCD with extreme HbF phenotype underwent Sanger re-sequencing of the three DHSs within *BCL11A*. Identified novel markers listed. Genotyping status and MAF within the CSSCD listed.

Table S3. Rare and Low-Frequency Variant Analysis.

DHS	Markers (*n*)	*P*	Conditional on rs1427407 *P*	Conditional on rs1427407 and rs7606173 *P*
+62	4	0.00149	0.06092	0.59353
+58	6	0.06506	0.03668	0.07146
+55	4	0.00681	0.35022	0.75880
all	14	0.00063	0.15035	0.69089

Rare and low-frequency variant analysis results (MAF < 5%). The analysis was performed using the set-based SKAT-O algorithm using the individual DHSs +62, +58 and +55 as three different sets. The bottom row "all" shows results with the three regions collapsed together.

Table S4. Emulsion Fusion Haplotyping PCR.

Donor no.	G–G	A–T	G–T	A–G	Likelihood ratio
1	19	22	4	2	1.63×10^{29}
2	22	14	2	3	3.78×10^{26}
3	25	23	9	10	4.69×10^{11}

Emulsion fusion PCR analysis of rs7569946–rs1427407 haplotype. Fusion PCR conducted in emulsion from three individual donors doubly heterozygous for rs7569946 and rs1427407, generating a fusion amplicon encompassing both SNPs. The fusion amplicon was cloned and individual clones Sanger sequenced. The number of clones of each genotype is listed. The likelihood ratio for the G–G/A–T as compared to G–T/A–G phase was calculated.

Table S5. Reporter Assay Fragment Coordinates.

		hg19, chr2		Distance from *BCL11A* TSS (bp)		
Reporter	Name	Start	End	Start	End	Length (bp)
lacZ	52.0-64.4	60,716,189	60,728,612	64,444	52,021	12,423
	56.8-64.4	60,716,189	60,723,870	64,444	56,763	7,681
	52.0-57.6	60,722,992	60,728,612	57,641	52,021	5,620
	+62	60,717,236	60,719,036	63,397	61,597	1,800
	+58	60,722,006	60,723,058	58,627	57,575	1,052
	+55	60,724,917	60,726,282	55,716	54,351	1,365
GFP	+164	60,616,396	60,618,032	164,237	162,601	1,636
	+156	60,623,536	60,624,989	157,097	155,644	1,453
	+153	60,626,565	60,628,177	154,068	152,456	1,612
	+62	60,717,236	60,719,036	63,397	61,597	1,800
	+58	60,721,212	60,722,958	59,421	57,675	1,746
	+55	60,724,780	60,726,471	55,853	54,162	1,691
	+41	60,739,075	60,740,154	41,558	40,479	1,079
	+32	60,748,003	60,749,009	32,630	31,624	1,006
	-46	60,826,438	60,827,601	-45,805	-46,968	1,163
	-52	60,831,589	60,833,556	-50,956	-52,923	1,967

Coordinates of the putative enhancer fragments cloned in the enhancer reporter assays. Chromosome 2 coordinates listed in hg19 as well as in reference to the *BCL11A* TSS.

Table S6. Oligonucleotide sequences.

Name	Sequence	Assay
mBcl11a-5'-F	AAAGAGCTGTCCGAAGTCCA	TALEN deletion PCR
mBcl11a-5'-R	GGGCACTTCCTAGTCCCTCT	TALEN deletion PCR
mBcl11a-del1-F	TTTGAGCAGGAGGGAATTTG	TALEN deletion PCR
mBcl11a-del1-R	ATGTTGTGGTCCCTGTGGTT	TALEN deletion PCR
mBcl11a-del2-F	GCAAGGCAGGTACCAAACAT	TALEN deletion PCR
mBcl11a-del2-R	TAGAGATTCCAGGCCCCTTT	TALEN deletion PCR
mBcl11a-3'-F	AGCAAGGAAAGGTGAAGCAG	TALEN deletion PCR
mBcl11a-3'-R	CCCAATGTCTTCCGAACTGT	TALEN deletion PCR
mBcl11a-upstreamTALEN-F	AGGCTGGTCTTGGGATTTTT	TALEN deletion PCR
mBcl11a-downstreamTALEN-R	GCCTTTAACAAGGGTGTCCA	TALEN deletion PCR
mBcl11a-5'probe-F	CATAGACCTGGGTCCTGGAA	5'-probe for Southern blot
mBcl11a-5'probe-R	TTGCAGAGTGACTCCTGTGG	5'-probe for Southern blot
hBCL11A-52.0-F	CCAGCCATACCCAAAACAAA	lacZ reporter cloning
hBCL11A-64.4-R	CTTTCCCTCTTGCCACTCAG	lacZ reporter cloning
hBCL11A-56.8-F	GGCAGAGAAGGCACAGTGA	lacZ reporter cloning
hBCL11A-57.6-R	GGCTGTCCTGGCATGTAAGT	lacZ reporter cloning
hBCL11A-63.4-F	AACAGACCCATGTGCTAGGC	lacZ/GFP reporter cloning
hBCL11A-61.6-R	TGTGTGGACTGCCTTTTCTG	lacZ/GFP reporter cloning
hBCL11A-58.6-F	GGGAAAAGGGAGAGGAAAAA	lacZ reporter cloning
hBCL11A-57.6-R	CTCAGAAAAATGACAGCACCA	lacZ reporter cloning
hBCL11A-55.7-F	GGACTCAGTGGCCTCTTTTG	lacZ reporter cloning
hBCL11A-54.4-R	GAAGATAATGGCAGCCCAGA	lacZ reporter cloning
hBCL11A-164.2-F	TGTGTGGCCAACCTGTAAAA	GFP reporter cloning
hBCL11A-162.6-R	CTCGCTCTGTTTCCCAGTTC	GFP reporter cloning
hBCL11A-157.1-F	CTCTCCGACGACCTCTTTTG	GFP reporter cloning
hBCL11A-155.6-R	GTAGGGAAGGGGCTACTTGG	GFP reporter cloning
hBCL11A-154.1-F	AGAGCCAAACTCCGTCTCAA	GFP reporter cloning
hBCL11A-152.5-R	AAATACCACAGCCCAACAGC	GFP reporter cloning
hBCL11A-59.4-F	GAACAGAGACCACTACTGGCAAT	GFP reporter cloning
hBCL11A-57.7-R	GGGGAAGGGGTATTGAATTG	GFP reporter cloning
hBCL11A-55.9-F	CTTCCACTGGATGGCACTTT	GFP reporter cloning
hBCL11A-54.2-R	ACTTCAGCCTCCAGCACTGT	GFP reporter cloning
hBCL11A-41.6-F	CCTCCCAGCAATGTAGGTGT	GFP reporter cloning
hBCL11A-40.5-R	TGGTGTGGTCCACTGTGACT	GFP reporter cloning
hBCL11A-32.6-F	GCAAGCTTAGCCCCTTCTTT	GFP reporter cloning
hBCL11A-31.6-R	TGAGGCAGAGTCAGATGTGG	GFP reporter cloning
hBCL11A-n45.8-F	CCCCGCTCAGAGTAAGTGAG	GFP reporter cloning
hBCL11A-n47.0-R	GGAAACTGCCTATCCCATGA	GFP reporter cloning
hBCL11A-n51.0-F	CAACACCCCGATTTCAGACT	GFP reporter cloning
hBCL11A-n52.9-R	GAATGGTCCCGATCTCTTGA	GFP reporter cloning
mGapdh-RT-F	TGGTGAAGGTCGGTGTGAAC	RT-qPCR (Gapdh)
mGapdh-RT-R	CCATGTAGTTGAGGTCAATGAAGG	RT-qPCR (Gapdh)
mBcl11a-RT-e1e2-F	AACCCCAGCACTTAAGCAAA	RT-qPCR (Bcl11a exon-1/2)
mBcl11a-RT-e1e2-R	ACAGGTGAGAAGGTCGTGGT	RT-qPCR (Bcl11a exon-1/2)
mBcl11a-RT-e2e3-F	GCCCCAAACAGGAACACATA	RT-qPCR (Bcl11a exon-2/3)
mBcl11a-RT-e2e3-R	GGGGCATATTCTGCACTCAT	RT-qPCR (Bcl11a exon-2/3)
mBcl11a-RT-e4e4-F	ATGCGAGCTGTGCAACTATG	RT-qPCR (Bcl11a exon-4/4, XL

		isoform)
mBcl11a-RT-e4e4-R	GTAAACGTCCTTCCCCACCT	RT-qPCR (Bcl11a exon-4/4, XL isoform)
mBcl11a-RT-e4e5-F	CAGCTCAAAAGAGGGCAGAC	RT-qPCR (Bcl11a exon-4/5, L isoform)
mBcl11a-RT-e4e5-R	GAGCTTCCATCCGAAAACTG	RT-qPCR (Bcl11a exon-4/5, L isoform)
mHbby-RT-F	TGGCCTGTGGAGTAAGGTCAA	RT-qPCR (eY)
mHbby-RT-R	GAAGCAGAGGACAAGTTCCCA	RT-qPCR (eY)
mHbb-bh1-RT-F	TGGACAACCTCAAGGAGACC	RT-qPCR (bH1)
mHbb-bh1-RT-R	ACCTCTGGGGTGAATTCCTT	RT-qPCR (bH1)
mHbb-b1-RT-F	TTTAACGATGGCCTGAATCACTT	RT-qPCR (b1/b2)
mHbb-b1-RT-R	CAGCACAATCACGATCATATTGC	RT-qPCR (b1/b2)
lacZ-RT-F	GCCAACATTGAGACACATGG	RT-qPCR (lacZ)
lacZ-RT-R	TGTCTCTCTGCACCATCCTG	RT-qPCR (lacZ)
lacZ-F	TTCAATGCTGTCAGGTGCTC	PCR genotyping (lacZ)
lacZ-R	GCCATGTGTCTCAATGTTGG	PCR genotyping (lacZ)
rs7569946-F	GTCTGCCCTCTTTTGAGCTG	haplotyping fusion PCR
rs7569946-R	GACTCCAGACAATCGCCTTT	haplotyping fusion PCR
rs7569946-R-rc-rs1427407-F	AAAGGCGATTGTCTGGAGTCAACCTT CTTAGCACCCACAAAC	bridging primer, haplotyping fusion PCR
rs1427407-R	CATGTTACTGCAACTTGCTTTTT	haplotyping fusion PCR
rs7569946-nested-F	AGATCCCTCCGTCCAGCTC	haplotyping fusion PCR
rs1427407-nested-R	TGAAAGTTCAAGTAGATATCAGAAGG	haplotyping fusion PCR
3C-hBCL11A-150.6-F	AGCAAACCACACAGACTGAAGA	3C
3C-hBCL11A-140.9-F	CCAGAGCCATTTACGTCACA	3C
3C-hBCL11A-114.1-F	CAGAAGGGAATAAGGTACTCTGGA	3C
3C-hBCL11A-111.5-F	GTTTGGGCCTCAAGGTCTTT	3C
3C-hBCL11A-109.1-F	GAGGTTGGGAGTAAGCATTCTG	3C
3C-hBCL11A-100.7-F	ACGCATCAGAATGCCCATAG	3C
3C-hBCL11A-92.3-F	TTTTGAAAGAAAACGCTGACA	3C
3C-hBCL11A-80.2-F	TTCCAGCTGGTTAAATTTAGGG	3C
3C-hBCL11A-77.2-F	AGAAGGGGCCAGAAGAACAG	3C
3C-hBCL11A-72.5-F	CCTTCTTTTTCTTTCTTGGTTGC	3C
3C-hBCL11A-66.8-F	CCCTGCGTGCCATTAAAATA	3C
3C-hBCL11A-61.2-F	AAAGGCCTTGGGAAGAAAGA	3C
3C-hBCL11A-59.1-F	GCAAGTCAGTTGGGAACACA	3C
3C-hBCL11A-57.1-F	GGACTCAGTGGCCTCTTTTG	3C
3C-hBCL11A-52.2-F	CTGTCTCTGTCTCCCCCAAG	3C
3C-hBCL11A-47-F	CCAATGCTCCTGTAACAAAGG	3C
3C-hBCL11A-43.5-F	AATGCAGTAGGCAAAGAAGCA	3C
3C-hBCL11A-38.6-F	GAAATTTGGAAGGCCACAGA	3C
3C-hBCL11A-29.3-F	GCTTGCAACAATTAAAAGATGG	3C
3C-hBCL11A-27.1-F	GGTGACAAGGGAGAACCACT	3C
3C-hBCL11A-20.9-F	TGATTTCCTTGCAGCCTTTT	3C
3C-hBCL11A-8.6-F	CACACCCACAGCAACAAATG	3C
3C-hBCL11Apromoter-R	TGCAGAGATCCCCCAAAGTA	3C
3C-hBCL11A-n8.3-F	CTCAGGGAGCAAGGGAAATA	3C
3C-hBCL11A-n12.6-F	CCCTCCCAACAGGGATTTAT	3C
3C-hBCL11A-n19.5-F	CAAAATTGAACACCTATGGTCTGA	3C
3C-hBCL11A-n29.8-F	AGGAAGACTTTGGCCTCCAT	3C
3C-hBCL11A-n34.6-F	TTCCAAACAATTATACACCAACAAA	3C
3C-hBCL11A-n54-F	TTTCATGGGGAATAGCCAAC	3C
3C-hBCL11A-n78.2-F	CCCTACTTGTTATTTGCTTCTGC	3C

3C-hBCL11A-n104.4-F	AGCTGAAGTTTCAGGGACCA	3C
3C-LCR-HS1-F	CCACACCTGCCTTCCTTAGA	3C
3C-LCR-HS3-F	TGCATATGATGGGGTAGCAG	3C
ChIP-hBCL11A-68.7-F	AAGAGAAGGGGGAATTTGGA	ChIP-qPCR
ChIP-hBCL11A-68.7-R	TGGTGATAAGGGCAGGAAAC	ChIP-qPCR
ChIP-hBCL11A-65.5-F	AGGAAGCTGCAGAAAGGTGA	ChIP-qPCR
ChIP-hBCL11A-65.5-R	TGCTTCCCCAGGTTTAGATG	ChIP-qPCR
ChIP-hBCL11A-64.7-F	CCACTGCTACCCAAAACGAT	ChIP-qPCR
ChIP-hBCL11A-64.7-R	CAAGAGCGAAACTCCACCTC	ChIP-qPCR
ChIP-hBCL11A-63.9-F	ACTGTGTGCCAAGTGACCAG	ChIP-qPCR
ChIP-hBCL11A-63.9-R	CAGCTTCCTTCAGGTGCTTC	ChIP-qPCR
ChIP-hBCL11A-63.1-F	CATGCTGCCTTTGTCTTCTG	ChIP-qPCR
ChIP-hBCL11A-63.1-R	TGTGGAGCTCTGGAATGATG	ChIP-qPCR
ChIP-hBCL11A-63.0-F	GAGCTCCACAATCCAACTCC	ChIP-qPCR
ChIP-hBCL11A-63.0-R	CCAGGAAGGAAATGAGAACG	ChIP-qPCR
ChIP-hBCL11A-62.5-F	ACCCACAAACATTTCCCTTCT	ChIP-qPCR
ChIP-hBCL11A-62.5-R	TTTGCTCTTCTCCAGGGTGT	ChIP-qPCR
ChIP-hBCL11A-62.4-F	TTTAAACAGCCACCCCACAC	ChIP-qPCR
ChIP-hBCL11A-62.4-R	ACCACGTAGTTGGGCTTCAC	ChIP-qPCR
ChIP-hBCL11A-62.2-F	TTTCAACCATGGTCATCTGC	ChIP-qPCR
ChIP-hBCL11A-62.2-R	CCCTCTGGCATCAAAATGAG	ChIP-qPCR
ChIP-hBCL11A-61.8-F	GAACCTGGGAGGCAGAAGAT	ChIP-qPCR
ChIP-hBCL11A-61.8-R	TTTTTGGTGAGACGGAGATTT	ChIP-qPCR
ChIP-hBCL11A-61.7-F	CCGGGCAACAAGAGTAAATC	ChIP-qPCR
ChIP-hBCL11A-61.7-R	ATGCCTAGGGTGTTTTGACG	ChIP-qPCR
ChIP-hBCL11A-61.5-F	CTCCGTGTTGAGAGCCAAGT	ChIP-qPCR
ChIP-hBCL11A-61.5-R	TGTGTGGACTGCCTTTTCTG	ChIP-qPCR
ChIP-hBCL11A-61.3-F	CAGAAAAGGCAGTCCACACA	ChIP-qPCR
ChIP-hBCL11A-61.3-R	CCTCTCCAGATTCCCTCTCA	ChIP-qPCR
ChIP-hBCL11A-61.0-F	AGCGAGACCCTGTCTCAAAA	ChIP-qPCR
ChIP-hBCL11A-61.0-R	TCCAGCAGGCTTCAAAAAGT	ChIP-qPCR
ChIP-hBCL11A-60.8-F	GGTGGATAACCCCATCTCAG	ChIP-qPCR
ChIP-hBCL11A-60.8-R	GGAAATGAGAATGCCCTTTG	ChIP-qPCR
ChIP-hBCL11A-60.5-F	CAGTCTAGAAAGCCCCCTCA	ChIP-qPCR
ChIP-hBCL11A-60.5-R	GTGGGGGTTCAGTGGTTAGA	ChIP-qPCR
ChIP-hBCL11A-60.3-F	TCCATGGTGTGGAGTGTGTT	ChIP-qPCR
ChIP-hBCL11A-60.3-R	ACCCACATGGCAACCAATAG	ChIP-qPCR
ChIP-hBCL11A-60.0-F	CCATTCCCTGGAGAGTTCAA	ChIP-qPCR
ChIP-hBCL11A-60.0-R	GGGGTCTCTTCCCATCATTT	ChIP-qPCR
ChIP-hBCL11A-59.9-F	ATGGGAAGAGACCCCAAAAC	ChIP-qPCR
ChIP-hBCL11A-59.9-R	GGACTCCGAACACCACACTT	ChIP-qPCR
ChIP-hBCL11A-59.5-F	GGGATCAGAGGTGAACAGGA	ChIP-qPCR
ChIP-hBCL11A-59.5-R	TTTAATCAGCTTCCGCCACT	ChIP-qPCR
ChIP-hBCL11A-59.0-F	TGGGGAGAGAAGAGTGGAAA	ChIP-qPCR
ChIP-hBCL11A-59.0-R	TTGCCAATTGGAGATTAGGG	ChIP-qPCR
ChIP-hBCL11A-58.7-F	TGCTCCGAGCTTGTGAACTA	ChIP-qPCR
ChIP-hBCL11A-58.7-R	GGGAAAGGGCCTGATAACTT	ChIP-qPCR
ChIP-hBCL11A-58.3-F	GAGAGTGCAGACAGGGGAAG	ChIP-qPCR
ChIP-hBCL11A-58.3-R	CCTCTTTCGGAAGGCTCTCT	ChIP-qPCR
ChIP-hBCL11A-58.0-F	TGGACTTTGCACTGGAATCA	ChIP-qPCR
ChIP-hBCL11A-58.0-R	GATGGCTGAAAAGCGATACA	ChIP-qPCR
ChIP-hBCL11A-57.3-F	GGGGAGATGATTGAAAGCAA	ChIP-qPCR
ChIP-hBCL11A-57.3-R	AGAACTTTCCCGGTTCTGGT	ChIP-qPCR
ChIP-hBCL11A-57.0-F	GCTCTGGACACACAGCAAAA	ChIP-qPCR

ChIP-hBCL11A-57.0-R	TCAAATCCTTGCCTTGAACC	ChIP-qPCR
ChIP-hBCL11A-56.6-F	CCTCAAATCTCCCTCACTGG	ChIP-qPCR
ChIP-hBCL11A-56.6-R	GGGAAATGGGTCCTGCTTTA	ChIP-qPCR
ChIP-hBCL11A-56.3-F	AGGGAGTACACCGCAGACAC	ChIP-qPCR
ChIP-hBCL11A-56.3-R	AAGGAAGGCTGCAAGGAAAT	ChIP-qPCR
ChIP-hBCL11A-55.9-F	GACTTAAACTGCCGCTCCTG	ChIP-qPCR
ChIP-hBCL11A-55.9-R	TGACTGGTAAGAGCCGATTG	ChIP-qPCR
ChIP-hBCL11A-55.3-F	GCTGGGGTGAGTCAAAAGTC	ChIP-qPCR
ChIP-hBCL11A-55.3-R	GGTCACCTTAAGGAGCCACA	ChIP-qPCR
ChIP-hBCL11A-54.8-F	GCACCTGCATTTGTTTTTCA	ChIP-qPCR
ChIP-hBCL11A-54.8-R	GGGTCAGATCACCTCTGCTC	ChIP-qPCR
ChIP-hBCL11A-54.4-F	AGGCATCCAAAGGGAAGAAT	ChIP-qPCR
ChIP-hBCL11A-54.4-R	GAAGATAATGGCAGCCCAGA	ChIP-qPCR
ChIP-hBCL11A-54.0-F	TGGGAAAGGTTGCACATTCT	ChIP-qPCR
ChIP-hBCL11A-54.0-R	GGGCCTCAGGCTCTTTATCT	ChIP-qPCR
ChIP-hBCL11A-53.4-F	CCACTGCCAGGCTGTTTACT	ChIP-qPCR
ChIP-hBCL11A-53.4-R	GACCGAAAGGAGGAGAGGAG	ChIP-qPCR
ChIP-hBCL11A-53.1-F	CAGTTCCCCCATTATGCACT	ChIP-qPCR
ChIP-hBCL11A-53.1-R	CCCTTCTCTGAAGGCACATC	ChIP-qPCR
ChIP-hBCL11A-52.7-F	TTCAAGCCTTGGTGGATAGG	ChIP-qPCR
ChIP-hBCL11A-52.7-R	GCCAGGAAATTGGTGGTAGA	ChIP-qPCR
ChIP-hBCL11A-52.3-F	TGCCCACATGAGACATCTTT	ChIP-qPCR
ChIP-hBCL11A-52.3-R	AAATTGGCTGCCATTGAATC	ChIP-qPCR
ChIP-hBCL11A-51.3-F	CCACCAGAAGTCCTGGAAAA	ChIP-qPCR
ChIP-hBCL11A-51.3-R	TTGGAGGGACCTGATCTCTG	ChIP-qPCR
ChIP-hBCL11A-50.2-F	CCAAGATGGAGAAGCCACAT	ChIP-qPCR
ChIP-hBCL11A-50.2-R	TCTGTCTTGGGTCTCCTGGT	ChIP-qPCR
ChIP-hBCL11A-49.8-F	GAGAAGCCCTCAGCAAACAC	ChIP-qPCR
ChIP-hBCL11A-49.8-R	GGTTGCATCTTGGCTCCTAA	ChIP-qPCR
ChIP-hBCL11A-49.5-F	GAAATGCAGGAAAGGAACGA	ChIP-qPCR
ChIP-hBCL11A-49.5-R	TCTAGCAGATGGGGTTTTGG	ChIP-qPCR
ChIP-hOct4-prom-F	AGTCTGGGCAACAAAGTGAGA	ChIP-qPCR
ChIP-hOct4-prom-R	AGAAACTGAGGAGAAGGATG	ChIP-qPCR
ChIP-hHS3-F	ATAGACCATGAGTAGAGGGCAGAC	ChIP-qPCR
ChIP-hHS3-R	TGATCCTGAAAACATAGGAGTCAA	ChIP-qPCR
ChIP-hHS-40-F	CAGATAACTGGGCCAACCAT	ChIP-qPCR
ChIP-hHS-40-R	ATTCACCCCTTTCCCTTGTC	ChIP-qPCR
ChIP-hGAPDH-F	CGTAGCTCAGGCCTCAAGAC	ChIP-qPCR
ChIP-hGAPDH-R	CGAACAGGAGGAGCAGAGAG	ChIP-qPCR

Oligonucleotides used in indicated experiments.

Supplementary References

39. J. Xu *et al*., *Dev*. *Cell*. **23**, 796-811 (2012).
40. E. van den Akker, T. J. Satchwell, S. Pellegrin, G. Daniels, A. M. Toye, *Haematologica*. **95**, 1594-1598 (2010).
41. A. L. Bredemeyer *et al*., *Nature*. **442**, 466-470 (2006).
42. R. E. Thurman *et al*., *Nature*. **489**, 75-82 (2012).
43. G. Galarneau *et al*., *Nat*. *Genet*. **42**, 1049-1051 (2010).
44. S. Purcell *et al*., *Am*. *J*. *Hum*. *Genet*. **81**, 559-575 (2007).
45. M. C. Wu *et al*., *Am*. *J*. *Hum*. *Genet*. **89**, 82-93 (2011).
46. A. He, S. W. Kong, Q. Ma, W. T. Pu, *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. **108**, 5632-5637 (2011).
47. M. A. McDevitt, Y. Fujiwara, R. A. Shivdasani, S. H. Orkin, *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. **94**, 7976-7981 (1997).
48. T. Cermak *et al*., *Nucleic Acids Res*. **39**, e82 (2011).
49. J. C. Miller *et al*., *Nat*. *Biotechnol*. **29**, 143-148 (2011).
50. R. Galanello *et al*., *Blood*. **114**, 3935-3937 (2009).
51. H. T. Bae *et al*., *Blood*. **120**, 1961-1962 (2012).
52. K. E. McGrath *et al*., *Blood*. **117**, 4600-4608 (2011).
53. D. Noordermeer, W. de Laat, *IUBMB Life*. **60**, 824-833 (2008).
54. D. R. Higgs, D. Vernimmen, B. Wood, *Adv*. *Genet*. **61**, 143-173 (2008).
55. E. Pinaud *et al*., *Adv*. *Immunol*. **110**, 27-70 (2011).