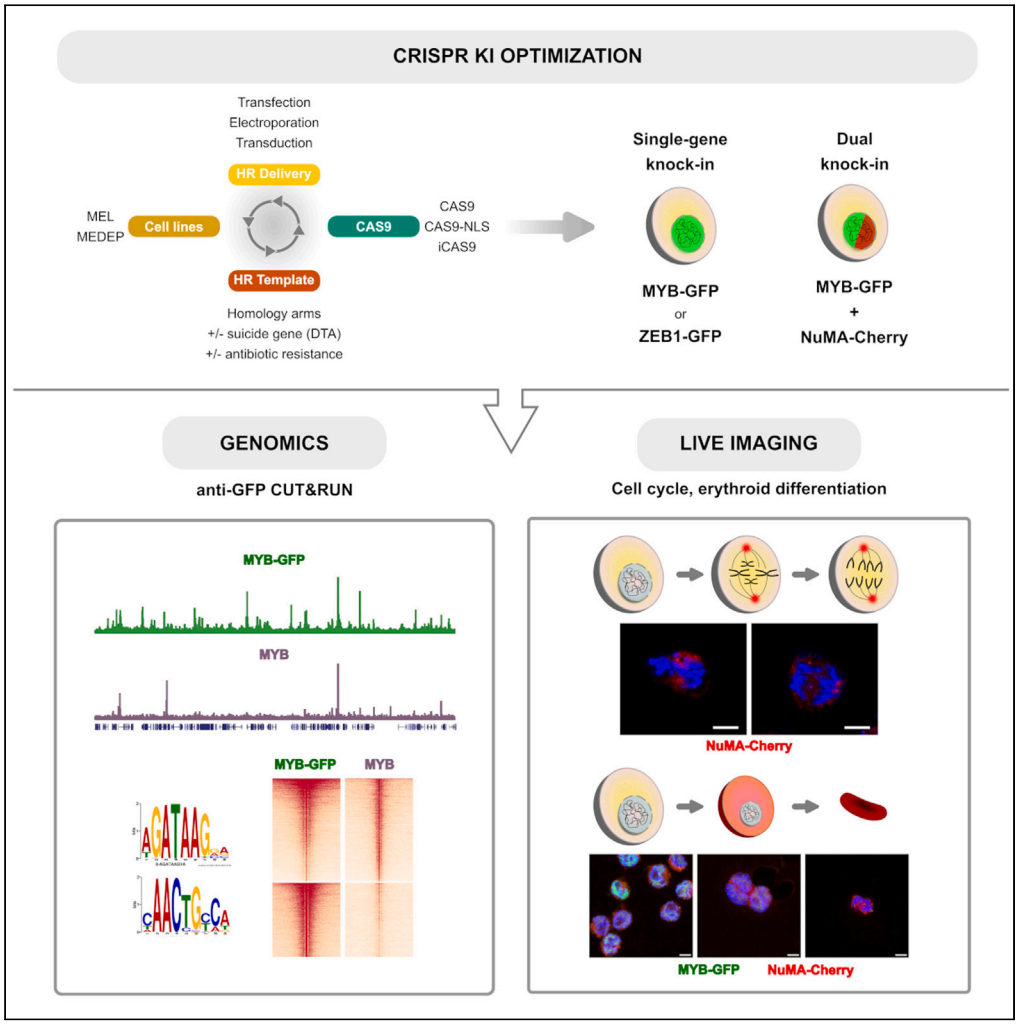
iScience

CellPress
OPEN ACCESS

Article

Efficient genome editing in erythroid cells unveils novel MYB target genes and regulatory functions



Virginie Deleuze, Leonor Garcia, Betty Rouaisnel, Mohammad Salma, Alexia Kinoo, Charlotte Andrieu-Soler, Eric Soler

eric.soler@igmm.cnrs.fr

Highlights

Efficient genome editing in hard-to-transfect erythroid cell lines

High-efficiency in-frame GFP insertion in low expressed gene loci *Myb* and *Zeb1*

Multi-gene targeting in individual cells: *Myb-Gfp/Numa1-Cherry* double KI

MYB-GFP CUT&RUN analysis uncovers previously unidentified MYB target genes

Deleuze et al., iScience *26*, 107641
September 15, 2023
https://doi.org/10.1016/j.isci.2023.107641

iScience

CellPress
OPEN ACCESS

Article

Efficient genome editing in erythroid cells unveils novel MYB target genes and regulatory functions

Virginie Deleuze,[1,4] Leonor Garcia,[1,4,5] Betty Rouaisnel,[1] Mohammad Salma,[1,2] Alexia Kinoo,[1] Charlotte Andrieu-Soler,[1,2,3] and Eric Soler[1,2,3,6,*]

SUMMARY

Targeted genome editing holds great promise in biology. However, efficient genome modification, including gene knock-in (KI), remains an unattained goal in multiple cell types and loci due to poor transfection efficiencies and low target genes expression, impeding the positive selection of recombined cells. Here, we describe a genome editing approach to achieve efficient gene targeting using hard to transfect erythroid cell lines. We demonstrate robust fluorescent protein KI efficiency in low expressed transcription factor (TF) genes (e.g., *Myb* or *Zeb1*). We further show the ability to target two independent loci in individual cells, exemplified by MYB-GFP and NuMA-Cherry double KI, allowing multicolor labeling of regulatory factors at physiological endogenous levels. Our KI tagging approach allowed us to perform genome-wide TF analysis at increased signal-to-noise ratios, and highlighted previously unidentified MYB target genes and pathways. Overall, we establish a versatile CRISPR-Cas9-based platform, offering attractive opportunities for the dissection of the erythroid differentiation process.

INTRODUCTION

Genome engineering is a powerful approach for numerous biological applications. It allows functional testing of genes and their regulatory elements, and allows creating recombinant alleles for protein tagging or fluorescent reporter systems at endogenous gene loci. CRISPR-CAS9 technology has revolutionized the way genome engineering can be handled.[1–7] It represents a relatively simple molecular approach as it only requires the introduction of the CAS9 endonuclease (or CAS9 functional variants) together with single guide RNAs (gRNAs) to target specific genomic sites with sequence complementary to the gRNAs. Cleavage of the target sequence by the CAS9 enzyme creates a local double-strand break, which can be repaired either by the error-prone non-homologous end joining (NHEJ), creating out-of-frame mutations, or by homology-directed repair (HDR) upon addition of a recombination template.[7] The latter allows inserting sequences of interest such as short sequence tags or fluorescent markers into endogenous loci to facilitate endogenous protein detection, purification or live imaging, without suffering from overexpression artifacts.

Despite its apparent simplicity, CRISPR-CAS9-mediated gene engineering requires introducing exogenous DNA and/or RNA material into the cells of interest, and selection of single clones bearing the desired modifications. These steps may prove extremely inefficient in hard-to-transfect cells, or cells that poorly recombine, resulting in very low genome engineering yields. Some erythroid cell lines have been notoriously known for being hard to transfect.[8–11] The mouse erythroleukemic cell line (MEL), for instance, is one such example. It is, however, a widely used model for erythroid research.[12–20] It is also part of the mouse ENCODE cell lines for which numerous epigenomic datasets have been generated,[21] hence being able to manipulate gene function in such model cell lines is of high interest. Another layer of difficulty relates to low expressed genes. Whereas abundantly expressed genes offer facilitated detection potential (when either fused to a selection marker or to a fluorescent protein), low expressed genes do not offer such an easy detection capacity, and selecting recombinant cell lines based on antibiotic resistance or fluorescence from low abundance transcripts is often problematic.

We tested here several different approaches to modify gene loci in the MEL erythroid cell line and systematically compared efficiencies, leading to the identification of robust conditions allowing knock-in (KI) insertion of in-frame exogenous sequences into endogenous gene loci, including low abundance protein genes. We further devised strategy to carry out similarly successful KI in the MEDEP erythroid cell line.[22] We show successful GFP KI generation in the low expressed transcription factor (TF) genes *Myb* and *Zeb1*, and we also show the ability

[1]IGMM, University Montpellier, CNRS, Montpellier, France
[2]Laboratory of Excellence GR-Ex, Université de Paris, Paris, France
[3]Senior author
[4]These authors contributed equally
[5]Present address: Institut de Génomique Fonctionnelle, INSERM U 1191, CNRS UMR 5203, Montpellier, France
[6]Lead contact
*Correspondence: eric.soler@igmm.cnrs.fr
https://doi.org/10.1016/j.isci.2023.107641

to generate double KI cell lines combining MYB-GFP and NuMA-Cherry simultaneous labeling allowing genome-wide profiling of chromatin occupancy by CUT&RUN and cellular imaging using the GFP and Cherry moieties. We demonstrate that we are able to detect MYB chromatin occupancy sites with increased sensitivity and provide a characterization of its regulatory network and target pathways in erythroid progenitors. Taken altogether these data provide a framework for successful genome modification of erythroid cell lines and provide novel insights into the role of key erythroid transcription factors.

RESULTS

Stable or inducible expression of CAS9 does not impair MEL cell behavior

We initially assessed the possibility to generate KI clones using the MEL cell line. We first generated stable MEL cell clones constitutively expressing the *Streptococcus pyogenes* Cas9 nuclease (MEL-CAS9), or an improved version containing additional nuclear localization sequences (NLS) for increased nuclear accumulation[23] (MEL-CAS9-NLS). We also generated doxycycline-inducible (Dox) Cas9 MEL cell clones (MEL-iCAS9). All clonal populations were tested for proper growth and differentiation to ascertain that stable or inducible Cas9 expression does not perturb MEL cell behavior. As shown in Figure S1A, CAS9 expression did not alter MEL cells growth, and did not alter their capacity to undergo terminal erythroid differentiation as judged by their hemoglobinization (Figure S1B) and activation of specific differentiation markers (Figure S1C). These data indicate that MEL cells can tolerate CAS9 expression.

Efficient CRISPR-mediated knock-in in MEL cells

We initially chose to target the *Myb* proto-oncogene locus for the following reasons: first, it is a critical regulator of the erythroid lineage[12,24–26] for which the molecular mechanisms of action remain obscure. Second, the *Myb* gene encodes a TF (c-MYB) that is expressed at very low levels (within the range of $\sim 10^2$–10^3 protein copies per cell),[14,27,28] despite its importance for erythropoiesis. It therefore represents a difficult target for KI and for selection of recombinant alleles. Third, its genome-wide and intracellular dynamics are unknown, and endogenous fluorescent alleles would be useful to shed light on *Myb* functions.

The general KI strategy is depicted in Figure 1A. In short, we designed HDR template vectors containing ~700 bp homology arms corresponding to the genomic sequences surrounding a gene's STOP codon (in our case, the *Myb* gene).

The vectors contain a ~25 nm flexible protein linker sequence[29] followed by an eGFP sequence. We designed gRNA in the vicinity of the STOP codon to introduce a DNA double-strand break upon CAS9 expression to favor HDR cassette integration. This should result in the replacement of the STOP codon by the in-frame insertion of the linker-fused GFP sequence, while retaining all endogenous sequences at the target locus, including the 3'UTR. We designed the targeting vectors in several different flavors, as shown in Figure 1B. All vectors contain a U6 promoter driving expression of the gRNAs of interest, and the homologous recombination cassettes. We tested various cassettes composed of (i) a single linker-GFP sequence flanked by homology arms (HDR template); or (ii) a similar construct containing a suicide gene (diphteria toxin gene, DTA) downstream of the 3' homology arm in order to favor homologous recombination and eliminate random integration events (HDR-DTA template); or (iii) a linker-GFP sequence fused to a self-cleavable peptide P2A fused to a puromycin resistance gene, flanked by homology arms, with the aim to enrich for proper recombination events using puromycin resistance (HDR-PuroR template).

We combined these different targeting vectors with two delivery methods in our CAS9-expressing MEL cell lines: first using an integration-deficient lentivirus (LV) vector system, or by transfection. Surprisingly, whereas MEL cells can be very efficiently transduced with LV vectors, none of the HDR constructs yielded *Myb*-GFP recombination events, even when using increasing doses of concentrated integrase-deficient LV preparations (Figure 1B and data not shown). We therefore sought to test and optimize the delivery of HDR templates by transfection methods. We first tested different transfection regimen using a plasmid encoding an eGFP. Among the different transfection protocols tested, the Neon transfection system gave the best yield in MEL cells (Figure S2). We therefore used Neon transfection as the HDR template delivery method. Surprisingly, the targeting constructs containing either the DTA suicide gene or the puromycin resistance gene gave no evidence of any positive KI event, as judged by PCR detection of correct integration of the construct (see STAR Methods for details) in any of the parental cell lines (i.e., MEL-CAS9, MEL-CAS9-NLS, or MEL-iCAS9) (Figure 1B and data not shown).

However, we could reproducibly detect correct integration events (*Myb*-GFP KI) in the pool of transfected MEL-CAS9, MEL-CAS9-NLS, or MEL-iCAS9 cells using the HDR template devoid of selection markers (Figures 2A and S3). We next wanted to evaluate and compare the KI efficiency in these different parental cell lines after transfection of the HDR template. The various MEL cell clones were transfected with the HDR template, the presence of proper recombination events was first confirmed in the transfected pool by PCR, and individual GFP^+ cells were then sorted out in order to isolate independent single clones. After clonal expansion of each single cell, PCR was used to determine if the HDR template was correctly integrated at the target (*Myb*) locus (as opposed to random genomic insertion, i.e., false positive clones). Results depicted in Figures 2B and S4 show that the different MEL cell lines were all able to yield proper integration events, albeit at varying yields. Whereas MEL-CAS9 (4 positive clones out of 62; 6.5%) and MEL-CAS9-NLS (4 positive clones out of 41; 10.8%) showed similar but moderate capacity to KI the HDR construct, the MEL-iCAS9 showed the best yield with more than 70% of the single GFP+ clones bearing at least one *Myb*-GFP KI allele (11 clones out of 15; 73.3%) (Figure 2B). Among the KI clones 45–50% were found to be homozygous for GFP insertion, irrespective of the parental cell line (MEL-CAS9, MEL-CAS9-NLS, or MEL-iCAS9), indicating that both alleles were efficiently targeted at elevated frequency (see Figure S5 for a complete summary of the KI yields and genetic status of the clones). As expected, heterozygous cells show intermediate fluorescence levels as compared to homozygous cells showing higher mean fluorescence intensities (Figure 3A). Presence of a protein product at a size corresponding to MYB-GFP fusion (104 kDa, Figure 3B) was apparent by Western blotting indicating successful integration of the GFP sequence in the endogenous *Myb* locus.

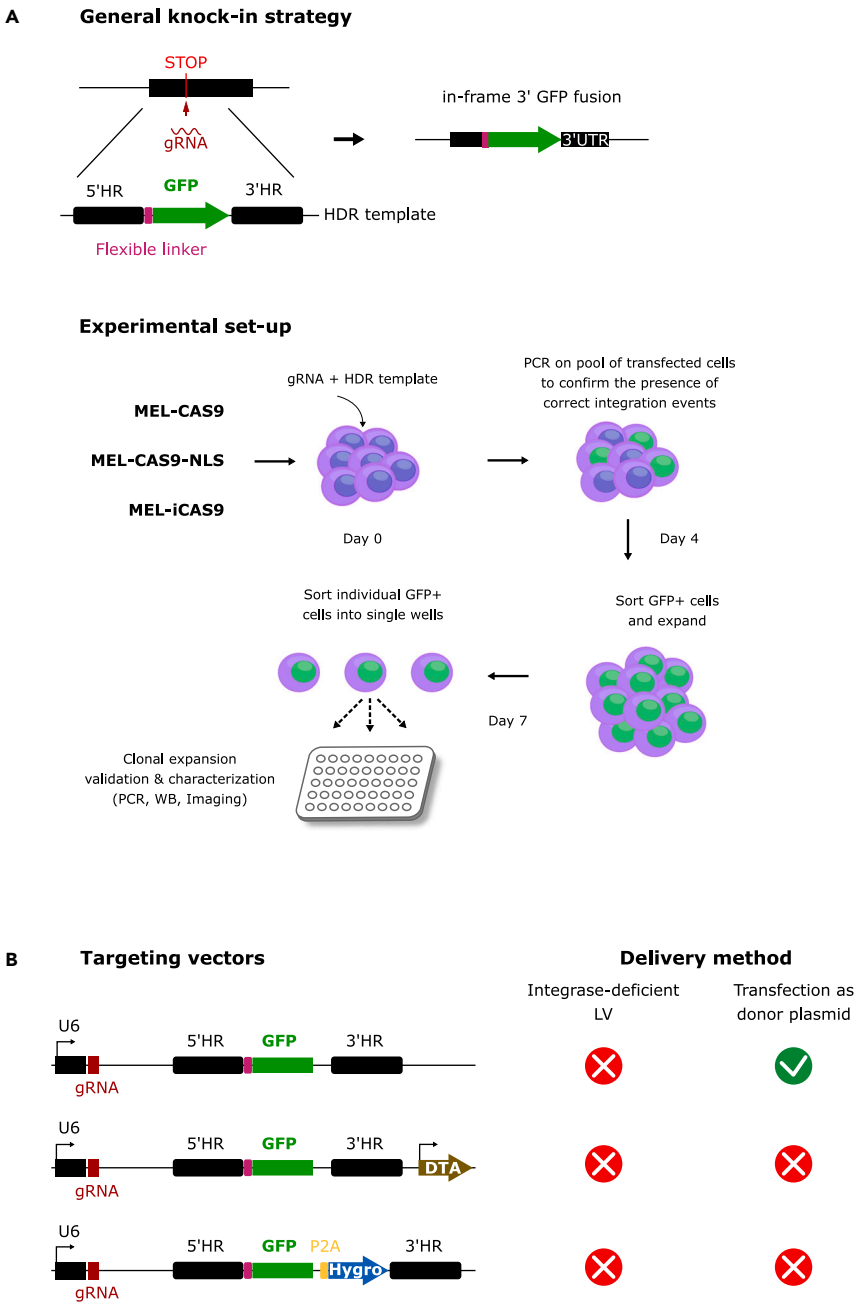


Figure 1. General knock-in strategy and targeting vectors

(A) *Top*: The general KI strategy is depicted. A gRNA with sequence complementarity in the vicinity of the STOP codon is designed, and used in combination with an HDR template containing a flexible linker and a GFP sequences flanked by 5′ and 3′ homology arms. *Bottom*: Experimental set-up consisting in testing the pool of transfected cells for proper recombination events by PCR, then sorting and expanding the GFP+ fraction, prior to single cells dispensing for clonal expansion. (B) The different targeting vectors tested and associated delivery methods. Crosses (red) and ticks (green) indicate absence or presence of recombination events, respectively. U6: U6 promoter, DTA: diphteria toxin gene, P2A: self-cleavage peptide P2A, Hygro: hygromycin resistance gene.
See also Figures S1 and S11.

In order to show the general applicability of our optimized conditions, we sought to target another weakly expressed gene locus in MEL cells. We selected the *Zeb1* TF gene, which is among the lowest expressed TF in MEL cells.[14] Again, using a similar strategy we were able to derive *Zeb1*-GFP KI clones (with a success rate of 2 out of 14 clones, 14%) (Figure S6A). Correct integration was confirmed at the genomic level by PCR, at the protein level by Western blotting, and by FACS (Figures 3C–3E). As expected, homozygous *Zeb1*-GFP KI cells show increased fluorescence intensity as compared to heterozygous cells (Figure 3D). They also show the unique presence of the ZEB1-GFP fusion protein in

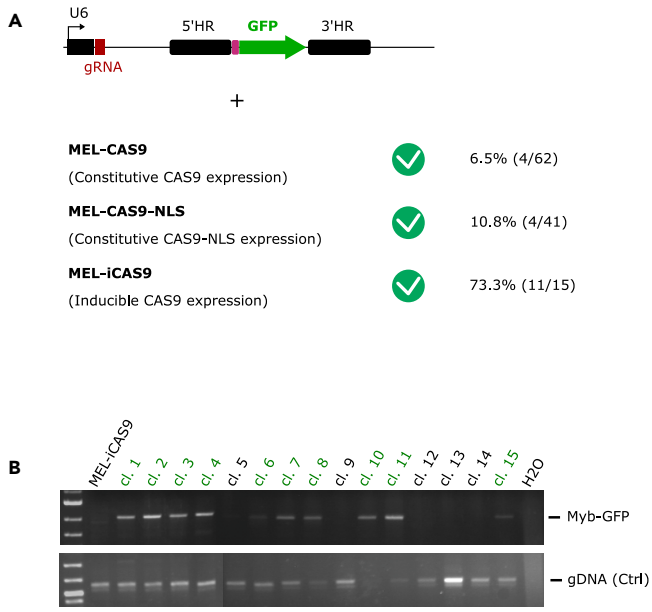


Figure 2. Knock-in efficiency

(A) The KI efficiencies using the indicated targeting vector in the various CAS9-expressing MEL cells are shown as percentage. The number of KI clones over the total number of screened clones is indicated between brackets.

(B) PCR screening of sorted GFP+ MEL cell clones. Top lane: KI-specific PCR product, bottom lane: genomic DNA control PCR. The *Myb*-GFP KI clones are labeled in green on top of the gel.

See also Figures S2–S5.

the absence endogenous non-modified ZEB1 protein, whereas heterozygous cells show both endogenous ZEB1 and ZEB1-GFP proteins (Figure 3E). *Zeb1*-GFP cells display similar growth kinetics compared to the parental MEL cells, and were able to properly differentiate as judged by the strong suppression of *Myb* gene expression and induction of the *Epb4.2* and *Gypa* differentiation genes, although *Gypa* was somewhat induced at slightly lower levels in the KI clones (Figures S6B and S6C). These results show that the strategy led to proper insertion of the GFP sequence by homologous recombination in MEL cells, leading to the expression of a ZEB1-GFP fusion protein at endogenous levels.

Taken altogether these results delineate the best conditions, among those tested, for knocking-in a fluorescent reporter in weakly expressed TF loci.

Characterization of *Myb*-GFP KI MEL cells

Myb-GFP MEL cells appear indistinguishable from their WT counterparts in terms of growth and differentiation (Figures 4A and 4B). As the MYB protein is critically required for proper MEL cell growth, this result indicates that the MYB-GFP fusion protein is functional and is able to fulfill its essential functions even in a homozygous state. At the transcriptional level, the *Myb* gene is rapidly suppressed when cells enter terminal erythroid differentiation[12,30] (Figures 4B and S1C). We monitored fluorescence intensity of the *Myb*-GFP clones during differentiation and observed a rapid decrease of fluorescence during the first 24 h, confirming proper transcriptional dynamics of the modified alleles (Figure 4C). Fluorescence levels dropped to background levels by 24–30 h of differentiation, reflecting the acute drop of *Myb* transcript abundance upon entry into terminal erythroid differentiation. Taken altogether, these data indicate that the *Myb*-GFP alleles behave similarly as the non-modified wild-type alleles, are subjected to proper regulatory inputs and that the MYB-GFP protein is functional.

MYB-GFP allows genome-wide analysis with increased sensitivity

The regulatory role(s) of MYB remain obscure. It is considered to be a potent cell cycle regulator, sustaining the proliferation potential of immature erythroid and hematopoietic progenitors, and blocking their differentiation.[24,30–32] It is also a critical regulator of the globin network, acting on the balance between fetal and adult β-globin genes expression through the regulation of the critical fetal globin regulators BCL11A and KLF1, and the globin genes themselves.[25,26,33] Despite its importance, genome-wide occupancy maps of MYB in hematopoietic and erythroid progenitors remain scarce. We therefore sought to determine whether the MYB-GFP cells could be used to efficiently profile MYB chromatin occupancy by using the GFP moiety as an affinity tag. We performed parallel CUT&RUN (Cleavage Under Target & Release Using Nuclease)[34,35] experiments using an anti-MYB or an anti-GFP antibody in MEL cells and in homozygous *Myb*-GFP MEL cells, respectively. We used an anti-MYB antibody that has been successfully used in landmark studies to detect MYB binding sites in leukemic cells[36,37] (see STAR Methods). Both CUT&RUN experiments yielded numerous high quality binding sites at the genome-wide scale. Global comparative analysis of the two datasets showed extensive overlap (i.e., 78% of MYB binding sites are detected in the MYB-GFP dataset,

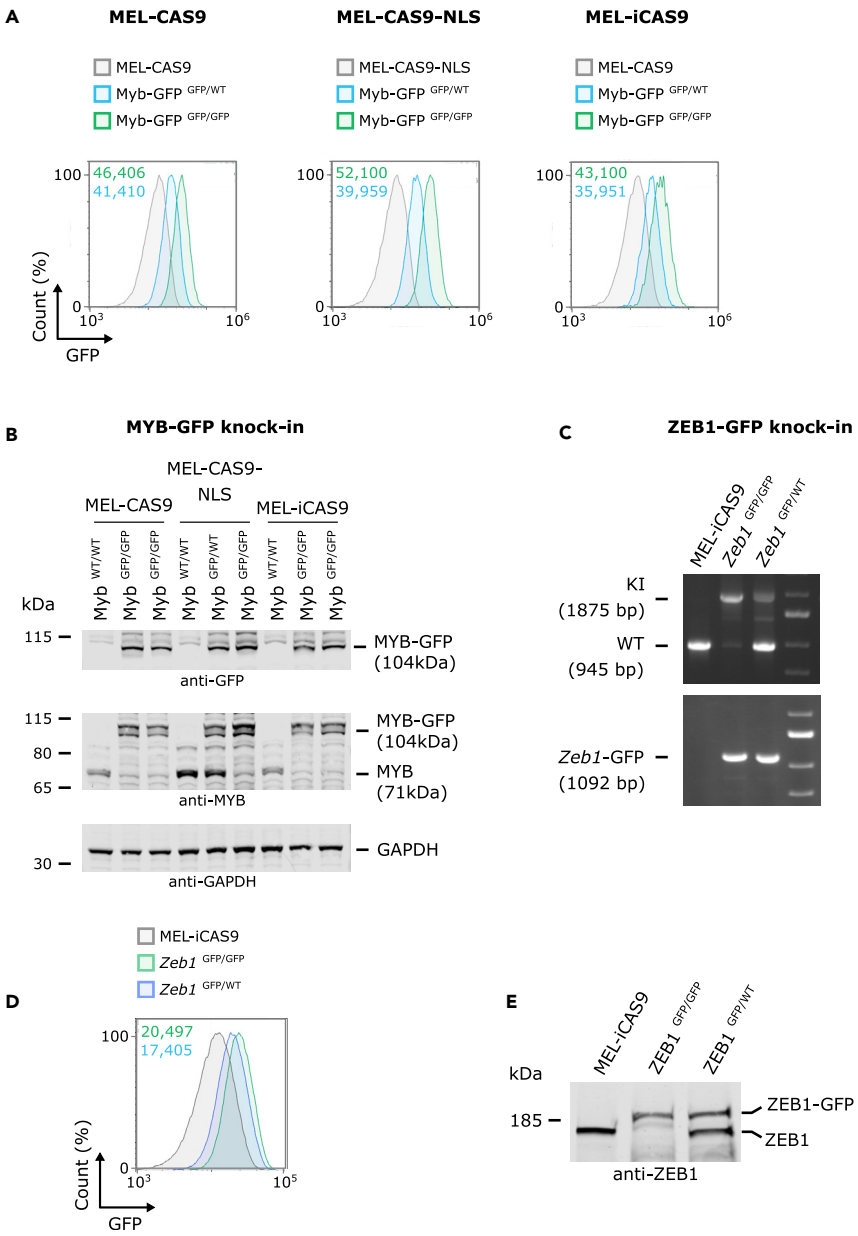


Figure 3. Cellular and molecular characterization of *Myb*-GFP and *Zeb1*-GFP KI MEL cells

(A) FACS analysis of heterozygous (*Myb*-GFP $^{GFP/WT}$) and homozygous (*Myb*-GFP $^{GFP/GFP}$) KI cells obtained from the different CAS9-expressing parental MEL cells. The MFI are indicated in each plot.

(B) Western blotting analysis of *Myb*-GFP KI MEL cells using anti-GFP, and anti-MYB antibodies. Anti-GAPDH was used as loading control.

(C) Two different PCR genotyping of Zeb1-GFP KI cells are shown, highlighting the presence of the endogenous Zeb1 (945 bp) signal in the control and heterozygous clones, but not in the homozygous KI cells. Specific amplification of the KI allele is also shown, which is uniquely found in the KI cells (bottom panel, 1,092 bp).

(D) FACS analysis of Zeb1-GFP KI cells. The heterozygous (blue) and homozygous (green) MFI values are indicated.

(E) Western blotting analysis of the Zeb1-GFP KI cells using an anti-ZEB1 antibody, highlighting endogenous ZEB1 and ZEB1-GFP proteins.

See also Figure S6.

Figure 5A); however, it appeared clearly that the MYB-GFP CUT&RUN detected many more sites (8071 vs. 3109, Figures 5A and 5B). The peak intensities (RPKM) obtained in the MYB-GFP dataset were generally much higher compared to the endogenous MYB CUT&RUN, with higher signal-to-noise ratios (Figures 5C, 5D, and S7). The increased sensitivity and low background obtained through the use of the GFP tag allowed us to detect multiple novel binding sites throughout the genome, as exemplified by the target gene *Cdc20* (Figures 5C and 5D). We next

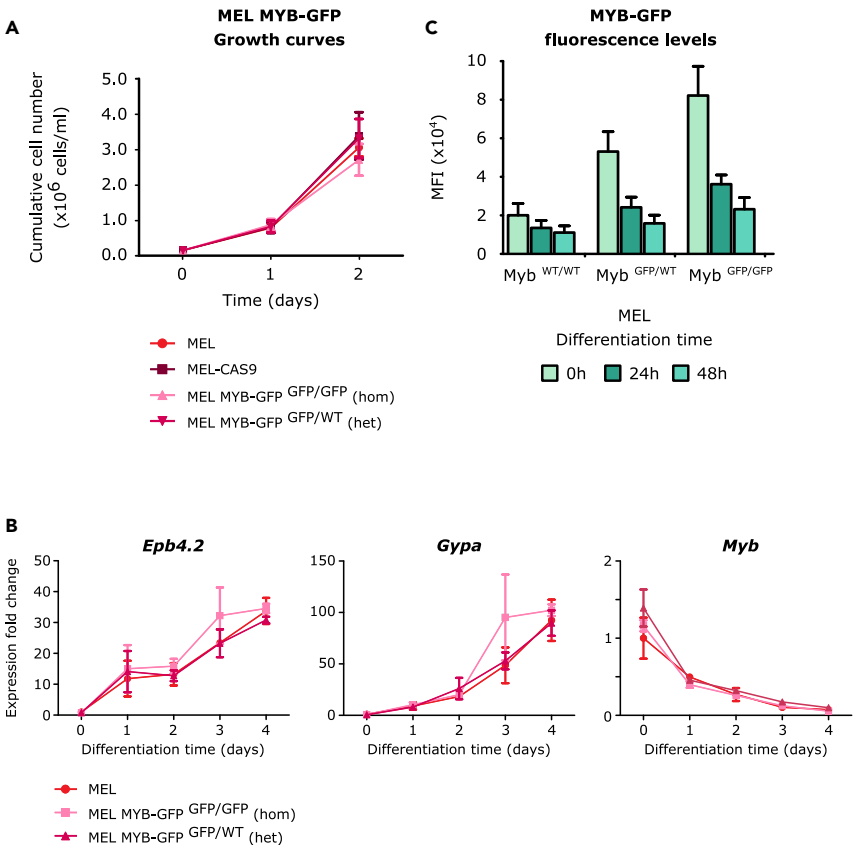


Figure 4. Functional characterization of *Myb*-GFP KI MEL cells

(A) Growth curves of MEL cells, MEL-CAS9, and two MEL *Myb*-GFP clones: 1 homozygous (hom), and 1 heterozygous (het).

(B) Expression of the *Epb4.2*, *Gypa*, and *Myb* genes upon induction of differentiation.

(C) Mean Fluorescence intensity (MFI) of heterozygous and homozygous MEL *Myb*-GFP cells showing the rapid drop of GFP as a function of differentiation time. Note that homozygous KI cells show increased MFI compared to heterozygous cells, as expected. Data are represented as mean ± SEM.

performed motif enrichment analysis to confirm that these novel sites are *bona fide* true MYB binding sites. We find that the MYB DNA binding motif is one of the most significantly enriched in the MYB-GFP CUT&RUN, strongly suggesting that the sites detected are likely direct MYB binding sites (Figure 6A). Surprisingly, we found that the MYB CUT&RUN binding sites are dominated by GATA1 motifs, and show a lower enrichment of MYB motifs compared to the MYB-GFP dataset (Figure 6B). This observation prompted us to compare MYB genome-wide occupancy with the location of regulatory elements and the master regulators GATA1 and LDB1.[13,38] We found a substantial overlap between MYB and GATA1/LDB1 at putative enhancer regions demarcated by H3K4me1, H3K27Ac, and low levels of H3K4me3 (Figure 6C). Interestingly, whereas anti-MYB CUT&RUN signals preferentially associate with the position of putative enhancers, anti-GFP CUT&RUN signals distribute more evenly between enhancer and promoter elements (defined by high H3K4me3 and low H3K4me1) (Figures 6C and 6D). As the GATA1/LDB1 complex is known to preferentially target erythroid enhancers[13,25,38–42] but not promoters, these MYB-bound promoters are mostly devoid of GATA1 or LDB1 occupancy (Figure 6C). Given the fact that the anti-MYB antibody preferentially picks-up distally bound putative enhancer regions, and that the anti-GFP signals are more evenly shared between enhancers and promoters, these results are in accordance with the motif enrichment analyses (i.e., strong GATA motif enrichment in the anti-MYB dataset versus preferential MYB motif enrichment in the MYB-GFP dataset) (see Discussion).

We next interrogated whether such additional MYB binding sites (highlighted in the MYB-GFP dataset) and associated target genes could better inform MYB functions in erythroid cells. Pathway enrichment analysis carried-out on the common and unique target sites revealed a strong cell cycle signature in the promoter-bound MYB-GFP targets, matching the known role of MYB as a proliferation-associated regulator and revealing new targets (Figure 5E), whereas MYB/MYB-GFP overlapping sites associate with various signaling pathways (e.g., FoxO and AMPK) and autophagy. Interestingly, mRNA transport and splicing were also specifically enriched in the genes targeted by MYB-GFP, suggesting that MYB controls transcripts processing in erythroid cells (Figures 5E and S8). We next monitored changes in expression of these target genes upon MEL cells differentiation. A majority (66%) of MYB targets are downregulated during differentiation (Figure S9), including the cell cycle promoting genes *E2f1*, *Myc*, or *Cdk6*, in accordance with the cell-cycle arrest associated with terminal erythroid differentiation. In addition, we also observed that a number of MYB targets become upregulated in differentiating MEL cells (34%), as exemplified by the *Hmbs*

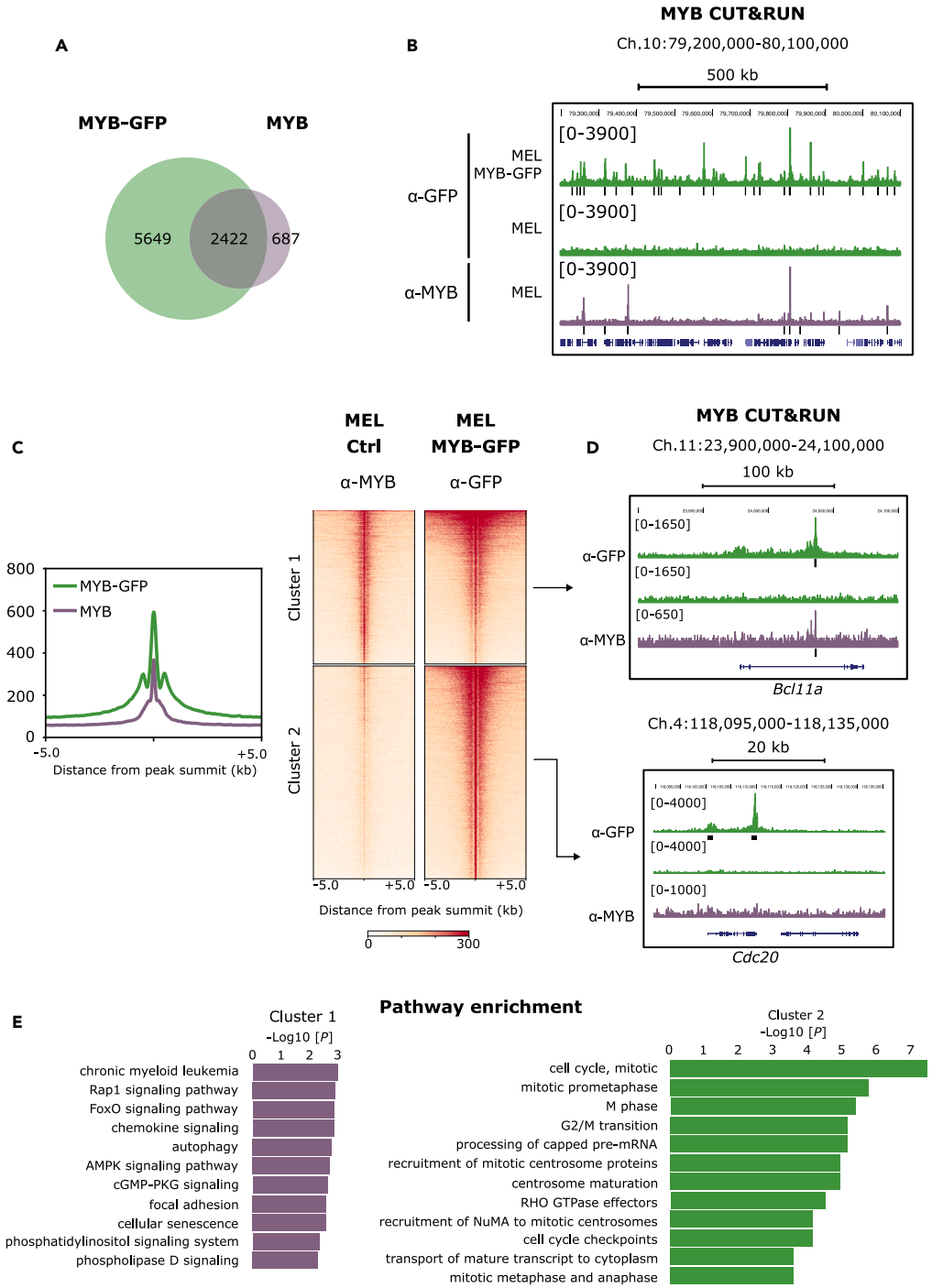


Figure 5. MYB-GFP genome-wide analysis

(A) Venn diagram showing the overlap between MYB and MYB-GFP CUT&RUN peaks.

(B) MYB-GFP and MYB CUT&RUN signals using anti-GFP or anti-MYB antibodies in MEL *Myb*-GFP KI and MEL cells, respectively. The use of anti-GFP in *Myb*-GFP KI MEL cells allows detecting many more sites at the genome-wide scale. The peak calling tracks are shown underneath the signal tracks.

(C) (left) Metaprofile analysis of CUT&RUN signals obtained using anti-MYB or anti-GFP antibodies; (right) Heatmap showing comparative analysis of MYB and MYB-GFP signals. Two clusters were observed: MYB/MYB-GFP overlapping sites (Cluster 1), and MYB-GFP unique sites (Cluster 2).

(D) Genome browser examples of CUT&RUN signal intensities at overlapping (*Bcl11a* locus) or unique (*Cdc20* locus) sites.

(E) Pathway enrichment analysis performed on Cluster 1 (MYB/MYB-GFP overlapping sites) and Cluster 2 (MYB-GFP unique sites).

See also Figures S7 and S8.

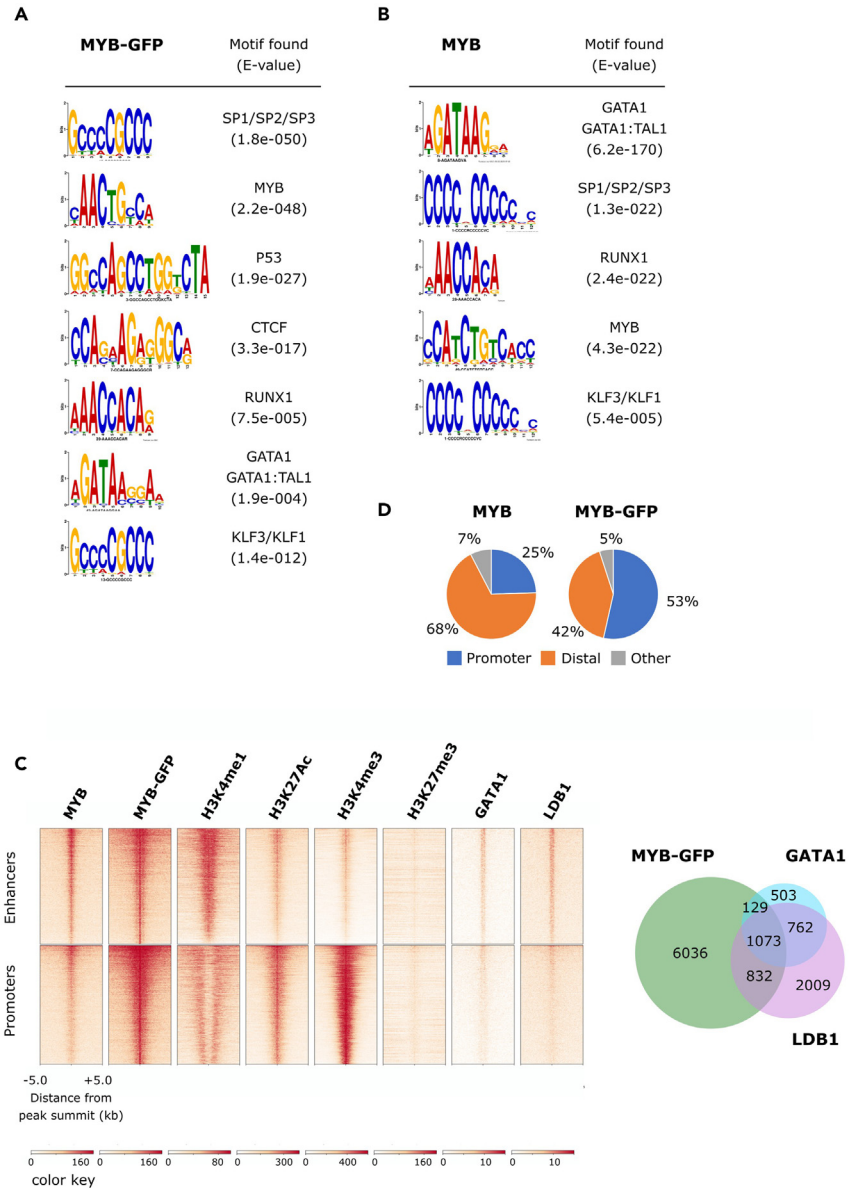


Figure 6. Genome-wide features of MYB-GFP binding sites

(A and B) DNA motif enrichment analyses were performed using (A) MYB-GFP and (B) MYB CUT&RUN peaks. The identified motifs are shown together with the corresponding factor and significance (E-value).

(C) Heatmaps (centered on the MYB and MYB-GFP signals) showing the overlap between MYB, MYB-GFP, epigenetic marks (H3K4me1, H3K27Ac, H3K4me3, and H3K27me3), and the master erythroid regulators GATA1 and LDB1. A Venn diagram representation is shown on the right.

(D) Genome-wide localization of MYB and MYB-GFP binding sites (Promoters were defined as -3kb to +1kb around transcription start sites).

See also Figures S9, S10, and S12.

(hydroxymethylbilane synthase), *Fech* (ferrochelatase), *Urod* (uroporphyrinogen decarboxylase) genes, involved in heme biosynthesis,[43] that are co-targeted by the GATA1/LDB1 complex (Figure S10). These data suggest that MYB is involved in both activating and repressing activities to control the proliferation-to-differentiation balance of erythroid progenitors.

Our data therefore provide improved characterization of MYB genome-wide occupancy map, and refine its associated gene regulatory networks. In particular we find that MYB specifically targets genes involved in the control of cell cycle checkpoints, especially at the G2/M transition and mitotic stages. Interestingly, as MYB shows extensive binding independently of the master regulators GATA1 and LDB1, our data indicate that these factors control distinct as well as overlapping regulatory networks in erythroid cells. The sharp decrease in MYB expression at the entry of terminal differentiation may, therefore, simultaneously suppress the MYB-activated mitotic/cell cycle

regulatory program, and release the differentiation block exerted at canonical erythroid maturation genes, allowing the GATA1/LDB1 complex to fully activate its targets (e.g., heme synthesis enzymes or globin genes).

Efficient CRISPR-mediated knock-in in MEDEP cells

MEDEP cells[22] emerged as an interesting murine model, being one of the closest to primary bone marrow erythroid progenitors in terms of whole proteome dynamics.[14] They respond to hormonal and growth factor signaling (e.g., dexamethasone and SCF), and are able to spontaneously differentiate up to the enucleation state upon stimulation by erythropoietin. MEDEP cells therefore emerged as a unique model system to study murine erythropoiesis. Having demonstrated the capacity to efficiently KI a fluorescent marker into endogenous loci in MEL cells, we wanted to validate its applicability to the MEDEP cell line. We employed a similar strategy as with MEL cells, being generating MEDEP clones expressing CAS9 as a stable or inducible form. To our surprise, we were unable to obtain any MEDEP lines with stable CAS9 expression, suggesting that prolonged CAS9 expression in this cell line is toxic and/or counter-selected. We therefore proceeded using transient expression of CAS9 in co-transfection experiments with the HDR templates. We first tested different transfection regimen and selected the Neon transfection system as being the most efficient (Figure S2). We transfected the *Myb*-GFP HDR construct into MEDEP cells and tested the presence of proper recombination events in the pool of transfected cells prior to single cell sorting of GFP^+ cells and clonal expansion. Out of 10 GFP^+ clones, 5 were validated by PCR as being *Myb*-GFP KI cells, indicating a success rate of 50% in the GFP^+ population. Within these 5 *Myb*-GFP MEDEP clones, 4 were heterozygous and 1 showed homozygous insertion of the HDR construct (Figures 7A–7C). This ratio is somewhat lower than the one observed in MEL cells (~50%), but indicates that both MEL and MEDEP cells are able to target both alleles without the need to employ multiple rounds of transfection/selection.

MEDEP-*Myb*-GFP cells were further characterized to ascertain their proper modification. FACS analyses showed that the heterozygous clones display intermediate GFP fluorescence intensities, and that the homozygous clone consistently displayed higher intensities, as expected (Figures 7B and 7F). Protein analysis using an anti-GFP antibody highlighted a high molecular weight protein corresponding to the MYB-GFP protein product at the expected size of 104 kDa in all clones. When using an anti-MYB antibody, both the endogenous protein and the MYB-GFP protein were detected in all heterozygous clones, and only the MYB-GFP protein in the homozygous clone (Figure 7C). This result further confirms that both alleles were properly targeted in the homozygous clone.

MEDEP-*Myb*-GFP KI cells showed comparable growth capacities compared to control cells (Figure 7D). Live cell imaging confirmed the presence of a nuclear fluorescent signal (Figure 7E). As expected, MYB-GFP levels showed a dramatic drop upon induction of erythroid differentiation reaching background levels at 30 h post induction, paralleling the decreased abundance of *Myb* mRNA (Figure 7F). Taken altogether, these results show that the KI alleles are properly regulated and that the MYB-GFP protein is functional in the MEDEP-*Myb*-GFP cells.

Multiplexed knock-in insertion of two fluorescent tags in erythroid cells

We next wondered if our approach could be suitable for multiple KI generations in the same cells. We chose to insert a Cherry fluorescent tag into the cell cycle marker and mitotic regulator *Numa1* (nuclear mitotic apparatus protein 1), a nuclear factor involved in spindle pole orientation during mitosis.[44] We could successfully generate double KI MEL and MEDEP cells harboring both MYB-GFP and NUMA-Cherry proteins expressed from their endogenous loci (Figure 8A). The success rate was higher in MEDEP cells (16 positive clones out of 48 tested; 33%) than in MEL cells (3 positive clones out of 65; ~5%), but we could successfully obtain double KI cells in each cell line, showing the efficiency of our approach for multiple rounds of KI generation. Confocal imaging of the MYB-GFP::NUMA-Cherry cells showed the presence of MYB-GFP in the nuclear compartment, as expected (Figure 8B). MYB-GFP intensities dropped rapidly during differentiation and were undetectable by 48 h. NUMA-Cherry signals were relatively homogeneous before differentiation, and gradually accumulated into foci as differentiation proceeded (Figure 8B, 48 h timepoints). To our surprise, NUMA1-Cherry which is generally described as a nuclear protein, consistently displayed both cytoplasmic and nuclear localization in MEDEP and MEL cells. This result was confirmed at the endogenous level by imaging untagged NUMA in wild-type (unmodified) MEL and MEDEP cells, showing that all cells examined displayed both nuclear and cytoplasmic NUMA localization (Figure 8C). These observations indicate that this peculiar NUMA1 localization is not caused by perturbation of NUMA due to its tagging with mCherry but rather represents endogenous intracellular distribution in erythroid cells. In addition, NUMA-Cherry displayed its characteristic accumulation at spindle poles in mitotic cells, showing that the mCherry tag did not alter its intracellular dynamics (Figure 8D). Interestingly, several reports have highlighted functional cytoplasmic localization of NUMA1 in neurons.[45,46] Our study therefore adds erythroid cells to the biological systems associated with cytoplasmic NUMA localization. Thus, we demonstrate here the potential to perform multiple modifications in erythroid cell lines, to generate fluorescent alleles for direct live protein visualization and localization without suffering from overexpression artifacts.

DISCUSSION

Genome engineering has become an indispensable tool to assess gene and genome function. Thanks to CRISPR technology, modifying genomic DNA has become at reach for many purposes, with relative ease. However, whereas modifying easy-to-transfect cells can be a relatively easy task, modifying and selecting recombinant cells in hard-to-transfect cell lines or primary cells remains cumbersome. It gets even harder when dealing with weakly expressed genes and proteins such as several transcription factors that are expressed at low protein

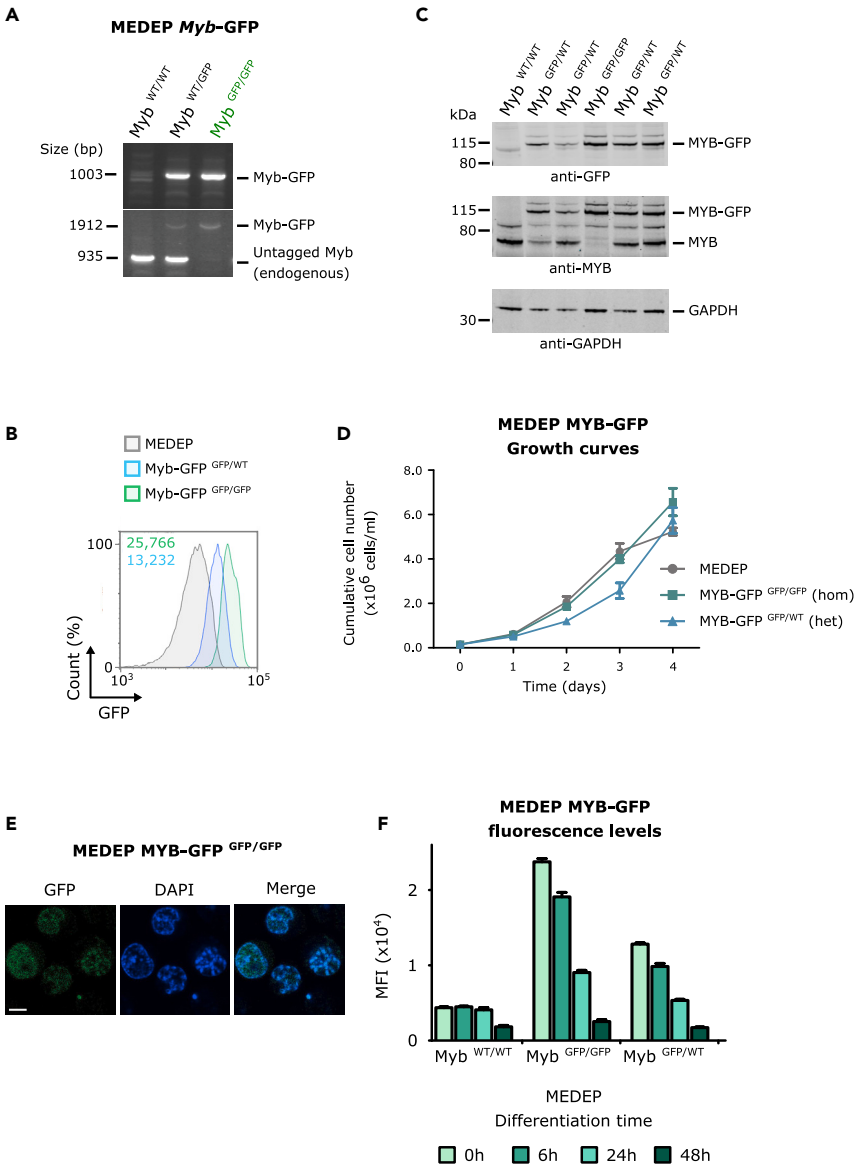


Figure 7. Characterization of MEDEP *Myb*-GFP KI cells

(A) PCR genotyping of heterozygous (*Myb* $^{WT/GFP}$) or homozygous (*Myb* $^{GFP/GFP}$) MEDEP KI cells.

(B) FACS analysis of GFP fluorescence in heterozygous and homozygous *Myb*-GFP KI. The MFI values are indicated.

(C) Western blot analysis using anti-GFP and anti-MYB antibodies. GAPDH was used as loading control.

(D) Growth curves of heterozygous (het) and homozygous (hom) Myb-GFP KI MEDEP cells. Data are represented as mean ± SEM.

(E) Confocal imaging of Myb-GFP KI cells. Scale bar: 5 μm.

(F) Mean Fluorescence intensity (MFI) of heterozygous and homozygous MEDEP *Myb*-GFP cells showing the rapid drop of GFP as a function of differentiation time (0h–48 h). Note that homozygous KI cells show increased MFI compared to heterozygous cells.

copy numbers per cell, despite their essential functions. The associated problems come from the fact that selecting for proper recombination events from a weakly expressed locus in a complex cell population remains a tremendous task.

In order to circumvent these issues and to provide a streamlined process for easy modification of hard-to-transfect cell lines, we compared different strategies and delineated the best conditions for easy KI generation. We tested whether tagging a gene with an antibiotic resistance cassette (in the form of a P2A-Puro^R protein moiety) could select for proper recombination events and eliminate unmodified cells from the population. We chose to target weakly expressed genes which represent difficult loci to target. To our surprise, this strategy did not yield any KI events. This may be explained by the fact that since *Myb* is a weakly expressed gene, its corresponding protein levels may not be sufficient to generate a sufficient pool of puromycin resistance proteins to efficiently survive the selection process. Decreasing the amount

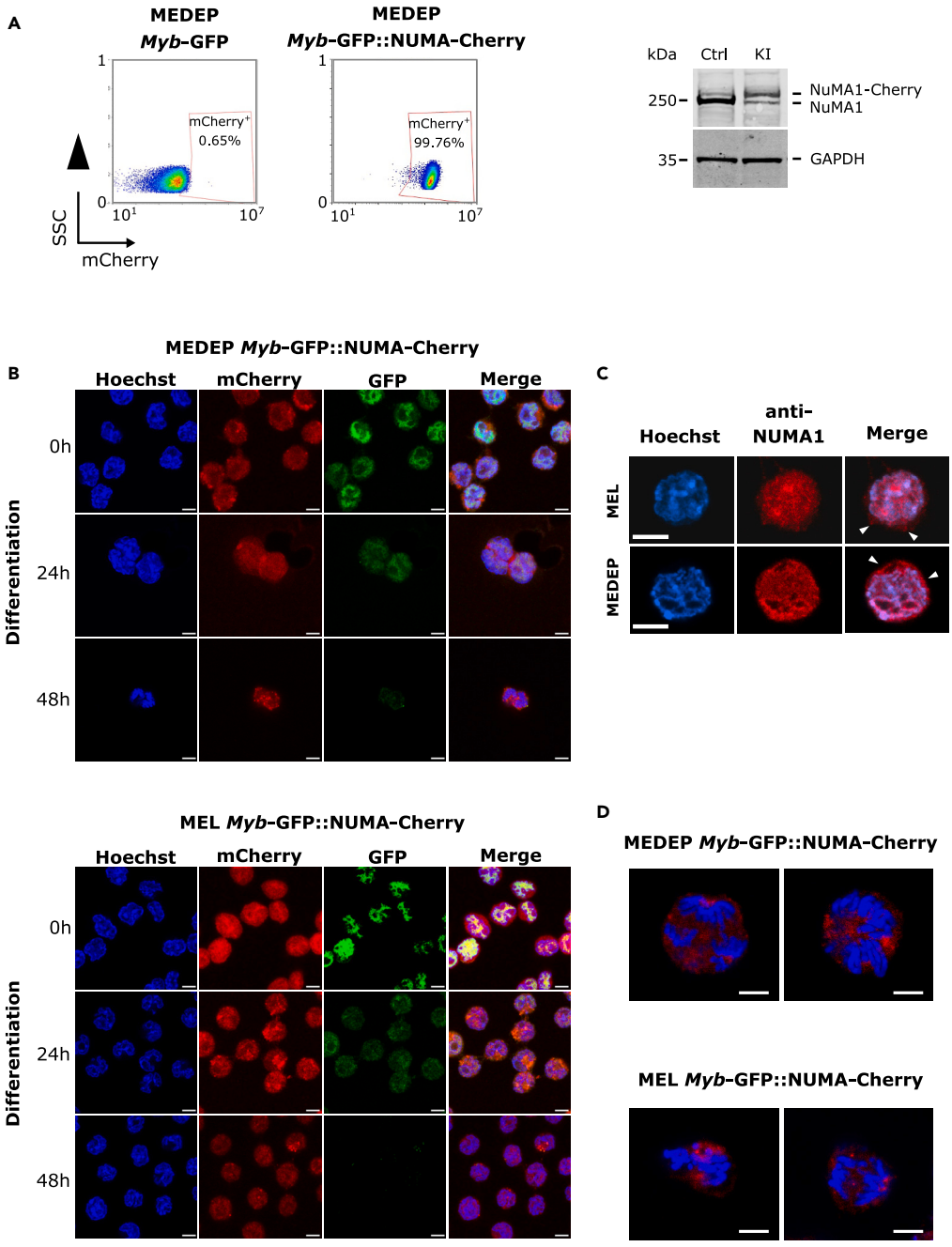


Figure 8. MYB-GFP::NUMA-Cherry double KI generation

(A and B) (A) (left) Representative FACS analysis of MEDEP *Myb*-GFP and *Myb*-GFP::*Numa*-Cherry, showing strong mCherry fluorescence in the double KI cells. (right) Anti-NUMA1 Western blotting performed on MEDEP *Myb*-GFP (Ctrl) or *Myb*-GFP::*Numa1*-Cherry (KI) (B) Confocal imaging of double KI MEDEP and MEL cells showing both MYB-GFP and NUMA-Cherry at steady state (0 h) and in differentiating conditions (24 h and 48 h).

(C) Immunofluorescence imaging showing cytoplasmic and nuclear NUMA1 localization in wild-type MEL and MEDEP cells. Arrow heads indicate regions of cytoplasmic NUMA1 localization.

(D) Relocalization of NUMA1-Cherry to the spindle poles during mitosis. Hoechst was used for nuclear staining. Scale bar: 5 μm.

of puromycin used during the selection period did not ameliorate the yield (not shown). Importantly, it is also worth mentioning that we observed incomplete P2A cleavage in fusion protein constructs in MEL cells (Figure S11), which may also decrease the availability of the PuroR enzyme if part of it is sequestered in the nuclear compartment if remaining partly fused to a TF. On the other hand, adding an independent counterselection cassette (DTA suicide gene) in the HDR constructs did not improve efficiency. In such case, the time needed to efficiently

integrate the HDR and get rid of the remaining DNA sequence (containing DTA) might be long enough to allow expression of the DTA and harm the cells. Interestingly, we observed that in MEL cells, a pulse of inducible CAS9 expression was more efficient to generate KI cells compared to stably expressed CAS9 parental cell lines (Figure 2). On the other hand, we find that MEDEP cells do not tolerate stable CAS9 expression, and that transient transfection of CAS9 expressing vectors was better suited. The reasons for these differences are unclear but these results underscore the need to rigorously assess various experimental strategies in each cell line of interest to define the best conditions for KI generation.

As antibiotic selection-containing constructs were inefficient, we reasoned that fusing a fluorescent tag to the protein of interest would represent a better strategy to enrich for KI cells by fluorescence-based cell sorting. This indeed allowed us to select for proper recombination events in *Myb*-GFP and *Zeb1*-GFP KI cells. Using this strategy, we could successfully isolate GFP KI cells, with varying success rates, reaching up to 73% from sorted single GFP^+ cells, in the best conditions. We also found an equal ratio of heterozygous to homozygous KI MEL cells indicating that recombination was efficient enough in this cell type to target both alleles without requiring several rounds of recombination/selection. Recombination rate in the GFP^+ MEDEP cell population was also satisfying (~50%), although the ratio of heterozygous to homozygous clones was somewhat lower than in MEL cells (~20%). Importantly, we could show that it is possible to perform successive recombination regimen to allow tagging of multiple genes in the same cells (MYB-GFP together with NUMA-Cherry). This opens-up attractive avenues for combinatorial analyses of regulatory factors controlling erythroid differentiation.

The advantages of the GFP tag are that it allows selecting for positive recombinant cells, to perform endogenous protein localization experiments and genome-wide profiling analyses. We showed that by using anti-GFP CUT&RUN in our MYB-GFP cells, we were able to detect a vast number of additional genomic binding sites (>2-fold more) and target genes that were not detected using anti-MYB CUT&RUN. Motif enrichment analysis identified known MYB DNA binding motifs associated with the GFP-MYB CUT&RUN peaks, with high significance, strongly suggesting that the newly identified sites are *bona fide* MYB binding regions and not unspecific sites. We noticed that the widely used anti-MYB antibody[36,37] shows a bias toward distal putative enhancer sites, whereas anti-GFP profiling produced a more even coverage between promoters and enhancers. The reasons for this bias are unclear, but it is worth mentioning that the anti-MYB antibody is actually directed against a phosphorylated form of MYB, the significance of which being obscure. Our results therefore strongly suggest that this commercially available anti-MYB antibody is biased toward subsets of MYB binding sites, especially at enhancer regions. The use of our MYB-GFP KI cells, which does not suffer from such bias, showed superior efficiency and highlighted many more promoters, which otherwise would have remained undetected. Our study therefore expands the MYB network and provides a frame for a better understanding of its essential functions.

We showed that MYB overlaps substantially with the GATA1/LDB1 complex. In addition, we detected a large fraction of MYB binding sites that do not overlap with GATA1/LDB1. This suggests that MYB and GATA1/LDB1 factors are involved in overlapping as well as distinct regulatory networks. Binding of MYB at a subset of GATA1/LDB1 sites may prevent premature activation of the terminal erythroid differentiation program controlled by GATA1/LDB1. This suggests that MYB likely assembles into distinct multiprotein complexes.[37,47] MYB-containing repressive complexes may target a subset of GATA1-bound enhancers to downmodulate terminal differentiation genes, whereas MYB-associated activating complexes may act at cell cycle genes to ensure proper erythroid progenitors proliferation. The sharp drop in MYB levels at the entry into terminal differentiation may release the transcriptional block exerted at canonical erythroid genes, as well as lower the expression of cell cycle-related genes. Indeed, we find that downregulated genes that are targeted by MYB mostly associate with cell cycle and DNA repair, whereas MYB target genes that become upregulated upon induction of differentiation are enriched for heme biosynthetic pathway, a critical process involved in hemoglobin formation (Figure S12). Taken altogether, our results provide information on the fine-tuning of the erythroid differentiation program, and expand characterization of the essential regulator MYB and its target genes. Our targeted genome modification approach therefore represents an attractive strategy to shed light on the regulators controlling erythroid cells, and should provide a valuable tool for further dissection of this complex and dynamic developmental process.

Limitations of the study

This study focuses on hard-to-transfect, non-adherent erythroid cell lines that are widely used models to study erythropoiesis. We currently do not know whether our optimized conditions may directly be applicable to other hematopoietic (e.g., lymphoid and myeloid) or non-hematopoietic cells. We demonstrate the ability to perform multiple rounds of KI generation (e.g., MYB-GFP::NUMA1-Cherry). We did not try to modify more than two loci in the same cells, and importantly we currently cannot predict how multiple (two or more) rounds of KI generation may impact the amount of putative off-target effects. We therefore systematically verify the growth and differentiation capacities of the modified cells to ensure that they retain their biological properties and remain valuable models for downstream analyses.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107641.

ACKNOWLEDGMENTS

This study was funded by grants from Laboratory of Excellence GR-Ex, reference ANR-11-LABX-0051 ("Investissements d'avenir" program of the French National Research Agency, reference ANR-11-IDEX-0005-02), Labex EpiGenMed ("Investissements d'avenir" program, reference ANR-10-LABX-12-01), the Fondation pour la Recherche Médicale (Equipe FRM DEQ20180339221), and the LNCC (Ligue Nationale Contre le Cancer - comité Hérault). Sequencing was performed by the GenomEast platform, a member of the "France Genomique" consortium (ANR-10-INBS-0009). The authors wish to thank Bernard Jost for precious advice on CUT&RUN sequencing library generation; Anne Coleno, Marie Kervella for initial input in construct generation, Stéphanie Viala and Myriam Boyer from MRI platform for the help with FACS and cell sorting, Peggy Reynaud, Céline Faux, and Dom Helmlinger for their help in setting-up CUT&RUN technology. We acknowledge the imaging facility MRI, member of the national infrastructure France-BioImaging infrastructure supported by the French National Research Agency (ANR-10-INBS-04, «Investments for the future»).

AUTHOR CONTRIBUTIONS

Conceptualization, C.A.S. and E.S.; Methodology, V.D, C.A.S., and E.S.; Investigation, V.D., L.G., B.R., and M.S.; Data Curation, M.S., V.D., L.G., and C.A.S.; Writing – Original Draft, E.S.; Funding Acquisition, E.S.; Supervision, C.A.S and E.S.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 27, 2023
Revised: July 9, 2023
Accepted: August 11, 2023
Published: August 15, 2023

REFERENCES

1. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816–821. https://doi.org/10.1126/science.1225829.
2. Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096. https://doi.org/10.1126/science.1258096.
3. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826. https://doi.org/10.1126/science.1232033.
4. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823. https://doi.org/10.1126/science.1231143.
5. Cho, S.W., Kim, S., Kim, J.M., and Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. *31*, 230–232. https://doi.org/10.1038/nbt.2507.
6. Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelson, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87. https://doi.org/10.1126/science.1247005.
7. Boti, M.A., Athanasopoulou, K., Adamopoulos, P.G., Sideris, D.C., and Scorilas, A. (2023). Recent Advances in Genome-Engineering Strategies. Genes *14*, 129. https://doi.org/10.3390/genes14010129.
8. Elnitski, L., and Hardison, R. (1999). Efficient and reliable transfection of mouse erythroleukemia cells using cationic lipids. Blood Cells. Mol. Dis. *25*, 299–304. https://doi.org/10.1006/bcmd.1999.0257.
9. Bouhassira, E.E., Westerman, K., and Leboulch, P. (1997). Transcriptional Behavior of LCR Enhancer Elements Integrated at the Same Chromosomal Locus by Recombinase-Mediated Cassette Exchange. Blood *90*, 3332–3344. https://doi.org/10.1182/blood.V90.9.3332.
10. Taxman, D.J., and Wojchowski, D.M. (1995). Erythropoietin-induced Transcription at the Murine βmaj-Globin Promoter. J. Biol. Chem. *270*, 6619–6627. https://doi.org/10.1074/jbc.270.12.6619.
11. Wang, H., Zhang, Y., Cheng, Y., Zhou, Y., King, D.C., Taylor, J., Chiaromonte, F.,

Kasturi, J., Petrykowska, H., Gibb, B., et al. (2006). Experimental validation of predicted mammalian erythroid *cis* -regulatory modules. Genome Res. *16*, 1480–1492. https://doi.org/10.1101/gr.5353806.

12. Stadhouders, R., Thongjuea, S., Andrieu-Soler, C., Palstra, R.-J., Bryne, J.C., van den Heuvel, A., Stevens, M., de Boer, E., Kockx, C., van der Sloot, A., et al. (2012). Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. EMBO J. *31*, 986–999. https://doi.org/10.1038/emboj.2011.450.
13. Stadhouders, R., Cico, A., Stephen, T., Thongjuea, S., Kolovos, P., Baymaz, H.I., Yu, X., Demmers, J., Bezstarosti, K., Maas, A., et al. (2015). Control of developmentally primed erythroid genes by combinatorial co-repressor actions. Nat. Commun. *6*, 8893. https://doi.org/10.1038/ncomms9893.
14. Gautier, E.-F., Leduc, M., Ladli, M., Schulz, V.P., Lefèvre, C., Boussaid, I., Fontenay, M., Lacombe, C., Verdier, F., Guillonneau, F., et al. (2020). Comprehensive proteomic analysis of murine terminal erythroid differentiation. Blood Adv. *4*, 1464–1477. https://doi.org/10.1182/bloodadvances.2020001652.
15. Yu, X., Martella, A., Kolovos, P., Stevens, M., Stadhouders, R., Grosveld, F.G., and Andrieu-Soler, C. (2020). The dynamic emergence of GATA1 complexes identified in *in vitro* embryonic stem cell differentiation and *in vivo* mouse fetal liver. Haematologica *105*, 1802–1812. https://doi.org/10.3324/haematol.2019.216010.
16. Cha, H.J., Uyan, Ö., Kai, Y., Liu, T., Zhu, Q., Tothova, Z., Botten, G.A., Xu, J., Yuan, G.-C., Dekker, J., and Orkin, S.H. (2021). Inner nuclear protein Matrin-3 coordinates cell differentiation by stabilizing chromatin architecture. Nat. Commun. *12*, 6241. https://doi.org/10.1038/s41467-021-26574-4.
17. Yan, B., Yang, J., Kim, M.Y., Luo, H., Cesari, N., Yang, T., Strouboulis, J., Zhang, J., Hardison, R., Huang, S., and Qiu, Y. (2021). HDAC1 is required for GATA-1 transcription activity, global chromatin occupancy and hematopoiesis. Nucleic Acids Res. *49*, 9783–9798. https://doi.org/10.1093/nar/gkab737.
18. Lee, J., Krivega, I., Dale, R.K., and Dean, A. (2017). The LDB1 Complex Co-opts CTCF for Erythroid Lineage-Specific Long-Range Enhancer Interactions. Cell Rep. *19*, 2490–2502. https://doi.org/10.1016/j.celrep.2017.05.072.
19. Giraud, G., Kolovos, P., Boltsis, I., van Staalduinen, J., Guyot, B., Weiss-Gayet, M., IJcken, W.v., Morlé, F., and Grosveld, F. (2021). Interplay between FLI-1 and the LDB1 complex in murine erythroleukemia cells and during megakaryopoiesis. iScience *24*, 102210. https://doi.org/10.1016/j.isci.2021.102210.
20. Krivega, I., and Dean, A. (2017). LDB1-mediated enhancer looping can be established independent of mediator and cohesin. Nucleic Acids Res. *45*, 8255–8268. https://doi.org/10.1093/nar/gkx433.
21. Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364. https://doi.org/10.1038/nature13992.
22. Hiroyama, T., Miharada, K., Sudo, K., Danjo, I., Aoki, N., and Nakamura, Y. (2008). Establishment of mouse embryonic stem cell-derived erythroid progenitor cell lines able to produce functional red blood cells. PLoS One *3*, e1544. https://doi.org/10.1371/journal.pone.0001544.
23. Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., and Root, D.E. (2014). Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nat. Biotechnol. *32*, 1262–1267. https://doi.org/10.1038/nbt.3026.
24. Mucenski, M.L., McLain, K., Kier, A.B., Swerdlow, S.H., Schreiner, C.M., Miller, T.A., Pietryga, D.W., Scott, W.J., and Potter, S.S. (1991). A functional c-myb gene is required for normal murine fetal hepatic hematopoiesis. Cell *65*, 677–689. https://doi.org/10.1016/0092-8674(91)90099-K.
25. Stadhouders, R., Aktuna, S., Thongjuea, S., Aghajanirefah, A., Pourfarzad, F., van Ijcken, W., Lenhard, B., Rooks, H., Best, S., Menzel, S., et al. (2014). HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers. J. Clin. Invest. *124*, 1699–1710. https://doi.org/10.1172/JCI71520.
26. Andrieu-Soler, C., and Soler, E. (2022). Erythroid Cell Research: 3D Chromatin, Transcription Factors and Beyond. Int. J. Mol. Sci. *23*, 6149. https://doi.org/10.3390/ijms23116149.
27. Gautier, E.-F., Ducamp, S., Leduc, M., Salnot, V., Guillonneau, F., Dussiot, M., Hale, J., Giarratana, M.-C., Raimbault, A., Douay, L., et al. (2016). Comprehensive Proteomic Analysis of Human Erythropoiesis. Cell Rep. *16*, 1470–1484. https://doi.org/10.1016/j.celrep.2016.06.085.
28. Karayel, Ö., Xu, P., Bludau, I., Velan Bhoopalan, S., Yao, Y., Ana Rita, F.C., Santos, A., Schulman, B.A., Alpi, A.F., Weiss, M.J., and Mann, M. (2020). Integrative proteomics reveals principles of dynamic phosphosignaling networks in human erythropoiesis. Mol. Syst. Biol. *16*, e9813. https://doi.org/10.15252/msb.20209813.
29. Kim, D.I., Jensen, S.C., Noble, K.A., Kc, B., Roux, K.H., Motamedchaboki, K., and Roux, K.J. (2016). An improved smaller biotin ligase for BioID proximity labeling. Mol. Biol. Cell *27*, 1188–1196. https://doi.org/10.1091/mbc.E15-12-0844.
30. Ramsay, R.G., and Gonda, T.J. (2008). MYB function in normal and cancer cells. Nat. Rev. Cancer *8*, 523–534. https://doi.org/10.1038/nrc2439.
31. Lieu, Y.K., and Reddy, E.P. (2009). Conditional c-myb knockout in adult hematopoietic stem cells leads to loss of self-renewal due to impaired proliferation and accelerated differentiation. Proc. Natl. Acad. Sci. USA *106*, 21689–21694. https://doi.org/10.1073/pnas.0907623106.
32. Lemma, R.B., Ledsaak, M., Fuglerud, B.M., Sandve, G.K., Eskeland, R., and Gabrielsen, O.S. (2021). Chromatin occupancy and target genes of the haematopoietic master transcription factor MYB. Sci. Rep. *11*, 9008. https://doi.org/10.1038/s41598-021-88516-w.
33. Andrieu-Soler, C., and Soler, E. (2020). When basic science reaches into rational therapeutic design: from historical to novel leads for the treatment of β-globinopathies. Curr. Opin. Hematol. *27*, 141–148. https://doi.org/10.1097/MOH.0000000000000577.
34. Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife *6*, e21856. https://doi.org/10.7554/eLife.21856.
35. Salma, M., Andrieu-Soler, C., Deleuze, V., and Soler, E. (2023). High-throughput methods for the analysis of transcription factors and chromatin modifications: Low input, single cell and spatial genomic technologies. Blood Cells Mol. Dis. *101*, 102745. https://doi.org/10.1016/j.bcmd.2023.102745.
36. Mansour, M.R., Abraham, B.J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A.D., Etchin, J., Lawton, L., Sallan, S.E., Silverman, L.B., et al. (2014). Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science *346*, 1373–1377. https://doi.org/10.1126/science.1259037.
37. Takao, S., Forbes, L., Uni, M., Cheng, S., Pineda, J.M.B., Tarumoto, Y., Cifani, P., Minuesa, G., Chen, C., Kharas, M.G., et al. (2021). Convergent organization of aberrant MYB complex controls oncogenic gene expression in acute myeloid leukemia. Elife *10*, e65905. https://doi.org/10.7554/eLife.65905.
38. Soler, E., Andrieu-Soler, C., de Boer, E., Bryne, J.C., Thongjuea, S., Stadhouders, R., Palstra, R.-J., Stevens, M., Kockx, C., van IJcken, W., et al. (2010). The genome-wide dynamics of the binding of Ldb1 complexes during erythroid differentiation. Genes Dev. *24*, 277–289. https://doi.org/10.1101/gad.551810.
39. Kassouf, M.T., Hughes, J.R., Taylor, S., McGowan, S.J., Soneji, S., Green, A.L., Vyas, P., and Porcher, C. (2010). Genome-wide identification of TAL1's functional targets: insights into its mechanisms of action in primary erythroid cells. Genome Res. *20*, 1064–1083. https://doi.org/10.1101/gr.104935.110.
40. Fujiwara, T., O'Geen, H., Keles, S., Blahnik, K., Linnemann, A.K., Kang, Y.-A., Choi, K., Farnham, P.J., and Bresnick, E.H. (2009). Discovering Hematopoietic Mechanisms through Genome-wide Analysis of GATA Factor Chromatin Occupancy. Mol. Cell *36*, 667–681. https://doi.org/10.1016/j.molcel.2009.11.001.
41. Yu, M., Riva, L., Xie, H., Schindler, Y., Moran, T.B., Cheng, Y., Yu, D., Hardison, R., Weiss, M.J., Orkin, S.H., et al. (2009). Insights into GATA-1-Mediated Gene Activation versus Repression via Genome-wide Chromatin Occupancy Analysis. Mol. Cell *36*, 682–695. https://doi.org/10.1016/j.molcel.2009.11.002.
42. Cheng, Y., Wu, W., Kumar, S.A., Yu, D., Deng, W., Tripic, T., King, D.C., Chen, K.-B., Zhang, Y., Drautz, D., et al. (2009). Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. Genome Res. *19*, 2172–2184. https://doi.org/10.1101/gr.098921.109.
43. Nilsson, R., Schultz, I.J., Pierce, E.L., Soltis, K.A., Naranuntarat, A., Ward, D.M., Baughman, J.M., Paradkar, P.N., Kingsley, P.D., Culotta, V.C., et al. (2009). Discovery of Genes Essential for Heme Biosynthesis through Large-Scale Gene Expression Analysis. Cell Metab. *10*, 119–130. https://doi.org/10.1016/j.cmet.2009.06.012.
44. Kiyomitsu, T., and Boerner, S. (2021). The Nuclear Mitotic Apparatus (NuMA) Protein: A Key Player for Nuclear Formation, Spindle Assembly, and Spindle Positioning. Front. Cell Dev. Biol. *9*, 653801. https://doi.org/10.3389/fcell.2021.653801.

45. Capizzi, M., Carpentier, R., Denarier, E., Adrait, A., Kassem, R., Mapelli, M., Couté, Y., and Humbert, S. (2022). Developmental defects in Huntington's disease show that axonal growth and microtubule reorganization require NUMA1. Neuron *110*, 36–50.e5. https://doi.org/10.1016/j.neuron.2021.10.033.
46. Torii, T., Ogawa, Y., Liu, C.-H., Ho, T.S.-Y., Hamdan, H., Wang, C.C., Oses-Prieto, J.A., Burlingame, A.L., and Rasband, M.N. (2020). NuMA1 promotes axon initial segment assembly through inhibition of endocytosis. J. Cell Biol. *219*, 201907048. https://doi.org/10.1083/jcb.201907048.
47. Xu, Y., Milazzo, J.P., Somerville, T.D.D., Tarumoto, Y., Huang, Y.-H., Ostrander, E.L., Wilkinson, J.E., Challen, G.A., and Vakoc, C.R. (2018). A TFIID-SAGA Perturbation that Targets MYB and Suppresses Acute Myeloid Leukemia. Cancer Cell *33*, 13–28.e8. https://doi.org/10.1016/j.ccell.2017.12.002.
48. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
49. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
50. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.
51. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.
52. Wang, Q., Li, M., Wu, T., Zhan, L., Li, L., Chen, M., Xie, W., Xie, Z., Hu, E., Xu, S., and Yu, G. (2022). Exploring Epigenomic Datasets by ChIPseeker. Curr. Protoc. *2*, e585. https://doi.org/10.1002/cpz1.585.
53. Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383. https://doi.org/10.1093/bioinformatics/btv145.
54. Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinf. *11*, 237. https://doi.org/10.1186/1471-2105-11-237.
55. Zhu, L.J. (2013). Integrative analysis of ChIP-chip and ChIP-seq dataset. Methods Mol. Biol. *1067*, 105–124. https://doi.org/10.1007/978-1-62703-607-8_8.
56. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS A J. Integr. Biol. *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.
57. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.
58. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.
59. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Res. *43*, W39–W49. https://doi.org/10.1093/nar/gkv416.
60. Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinforma. Oxf. Engl. *27*, 1696–1697. https://doi.org/10.1093/bioinformatics/btr189.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
mouse anti-Flag-M2	Sigma	Cat# F1804: RRID: AB_262044
rabbit anti-Gapdh	Cell Signaling	Cat# 3683S
rabbit anti-GFP	Sigma	Cat# SAB4301138; RRID: AB_2750576
mouse anti-Myb	Millipore	Cat# 05-175; RRID: AB_11213983
rabbit anti-Zeb1	Novus	Cat# NBP2-20942
goat anti-Rabbit (IR700)	Advansta	Cat# R-05054-250
goat anti-Rabbit (IR800)	Advansta	Cat# R-05060-250
goat anti-Mouse (IR700)	Advansta	Cat# R-05055-250
goat anti-Mouse (IR800)	Advansta	Cat# R-05061-250
mouse anti-Myb-p	Abcam	Cat# ab45150
Chemical, peptides, and recombinant proteins		
DMEM	Gibco	41965-039
IMDM	Gibco	319800-022
SVF	Eurobio	CVFSVF00-01
Antibiotique-antimycotique (100X)	Gibco	15240096
ITS 100x	Gibco	I 3146
Ascorbic Acid	Sigma	A4403
a-thioglycerol	Sigma	M6145
Dexamethasone	Sigma	D2915
mSCF	Home made	N/A
DMSO	Sigma	D8418
Erythropoietin	Peprotech	100-64
Puromycin	Gibco	A1113803
Neomycin (G418)	Gibco	11811031
Hygromycin	Invitrogen	1068/010
Hygromycin	Sigma	H0654-1G
Blasticidine	InvivoGen	30-414-BL
Nitrocellulose membrane Protran®	Amersham	10600003
Concanavalin-coated magnetic beads	Bangs Laboratories	BP531
Digitonin	Sigma	D141
Protein A/G–Micrococcal Nuclease (pA/G-MNase) fusion protein	Produced in Helmlinger's laboratory, Montpellier, France	N/A
Manganese Chloride ($MnCl_2$)	Sigma Aldrich	203734
Calcium Chloride ($CaCl_2$)	Fisher Scientific	BP510
Potassium Chloride (KCl)	Sigma Aldrich	P3911
Hydroxyethyl piperazineethanesulfonic acid	Sigma Aldrich	H3375
Sodium chloride (NaCl)	Sigma Aldrich	S5150-1L
Ethylenediaminetetraacetic acid (EDTA)	Research Organics	3002E
Ethylene glycol-bis(β-aminoethyl ether)-N,N,N,N-tetraacetic acid (EGTA)	Sigma Aldrich	E3889

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Spermidine	Sigma Aldrich	S2501
Roche Complete Protease Inhibitor EDTA-Free tablets	Sigma Aldrich	5056489001
Sodium dodecyl sulfate (SDS)	Sigma Aldrich	L4509
Agencourt SPRIselect beads	Beckman Coulter	B23317
GFP-Booster	ChromoTek	ATTO488
RFP-booster	ChromoTek	ATTO594
Hoechst 33342, trihydrochloride, trihydrate,	Invitrogen	H1399
Critical commercial assays		
Neon	Invitrogen	MPK5000
TransIT-X2	Mirus	MIR6000
GenCellinTM-HTC	BioCell Challenge	HTC750
TurboFect	Thermoscientific	R0531
QuickExtract DNA Extraction Solution	Lucigen	QE09050
HP RNA Isolation Kit	Roche	11828665001
SuperScriptTM II	Invitrogen	10297-018
PowerUpTM SYBRTM Green Master Mix	Applied Biosystems	A25742
Gel and PCR Clean-up kit	Macherey and Nagel	740609.250
MicroPlex Library Preparation kit v3	Diagenode	C05010001
Deposited data		
CUT&RUN (MYB and MYB-GFP)	This paper	GEO : GSE228372 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE228372
Experimental models: Cell lines		
MEDEP	Hiroyama et al.[22]	https://doi.org/10.1371/journal.pone.0001544
MEL	Kind Gift from J. Strouboulis	https://doi.org/10.1385/0-89603-178-0:421
Oligonucleotides		
sgRNA Myb	CACCGCGCGTTCTCAGCTCGAACTC	AAACGAGTTCGAGCTGAGAACGCGC
sgRNA Zeb1	CACCGAGCTTAGGAGTTCTTCTAAA	AAACTTTAGAAGAACTCCTAAGCTC
sgRNA Numa1	CACCGCAGTGGGGCCACGTAAGCAC	AAACGTGCTTACGTGGCCCCACTGCC
Myb-GFP	CATGGTCCTGCTGGAGTTCGTG (GFP)	AAAAATGTACAACAAATGCAAAATTCTC (Myb)
Zeb-GFP	CTTATAAATGTTCTGCCATGTGGTC (Zeb)	TCCTCGCCCTTGCTCAC (GFP)
Numa-cherry	CATCGTGGAACAGTACGAACG (Cherry)	ACATGAACCCCTGCAACAGT (Numa)
Myb	TCCGGCTCGGAAATACG	AAAAATGTACAACAAATGCAAAATTCTC
Zeb	GAGCGAGATGGAGAGCGAAA	ACAACAAACGAATCAACAACTGC
Numa	CAATAGAGGTTGGCAACACAGG	CCTGAGATCCAGAACCTATCTGTC
Myb	CGACGAAGACCCTGAGAAGG	GCTGCAAGTGTGGTTCTGTG
Ebp4.2	CTCGCTGCTGGTCTCTGG	GGAGAAGGACAGGTTGGTGG
Gypa	TCCTTACACCCTGTCTGCCT	CAACAAGCCAGCCCTCCTTA
Prkag1	GCACCGGATGTGGGGCGGGGT	GCTGGCGAGCTCTCTGCAGCAACCGAC
Software and algorithms		
NovoExpress	Agilent	https://www.agilent.com/en/product/research-flow-cytometry/flow-cytometry-software/novocyte-novoexpress-software-1320805
LightCycler 480	Roche	N/A
Image Studio Lite	LI-COR	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Trim Galore (version 0.6.6)	Babraham Bioinformatics	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
STAR (version 2.7.9a)	https://doi.org/10.1093/bioinformatics/bts635	https://github.com/alexdobin/STAR
bowtie2 (version 2.3.4.3)	https://doi.org/10.1038/nmeth.1923	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools (version 1.11)	https://doi.org/10.1093/gigascience/giab008	https://www.htslib.org
Picard (version 2.17.11)	Broad Institute	http://broadinstitute.github.io/picard/
HOMER (version 4.10.3)	https://doi.org/10.1016/j.molcel.2010.05.004	http://homer.ucsd.edu/homer/index.html
ChIPseeker (version 1.34.1)	https://doi.org/10.1093/bioinformatics/btv145	https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html
ChIPpeakAnno (version 3.32.0)	https://doi.org/10.1186/1471-2105-11-237	https://bioconductor.org/packages/release/bioc/html/ChIPpeakAnno.html
clusterProfiler (version 4.6.0)	https://doi.org/10.1089/omi.2011.0118	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
deepTools (version 3.4.2)	https://doi.org/10.1093/nar/gkw257	https://deeptools.readthedocs.io/en/latest/
meme-chip (version 5.3.3)	https://doi.org/10.1093/bioinformatics/btr189	https://meme-suite.org/meme/tools/meme-chip
R (version 4.2.2)	CRAN	https://cran.r-project.org
Other		
Odyssey® M	LI-COR	https://www.licor.com/bio/odyssey-m/
HiSeq 4000 sequencer	Illumina	https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/hiseq-3000-4000-specification-sheet-770-2014-057.pdf
LSM980 Airyscan 8Y	Zeiss	https://www.zeiss.com/microscopy/en/products/light-microscopes/confocal-microscopes/lsm-980-with-airyscan-2.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Eric Soler (eric.soler@igmm.cnrs.fr).

Materials availability

Plasmids generated in this study are available upon request.

Data and code availability

- CUT&RUN data for MYB and MYB-GFP have been deposited at GEO, and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- ChIP sequencing data for GATA1 and LDB1 are available at the Sequence Read Archive with accession ERA000161 (http://www.ebi.ac.uk/ena/data/view/ERA000161).
- This paper does not report original code.
- Microscopy data reported in this paper will be shared by the lead contact upon request.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines

Generation of CAS9, iCAS9, CAS9-NLS MEL cells

To select MEL cells expressing (constitutive or inducible) CAS9, transduced cells with pLentiCRISPRv2_Neo-short, pLX_311 or TLCV2-T2A-noGFP-puro virions were selected with G418 (Neomycin) at 500μg/ml for at least 10 days (MEL-CAS9), with Blasticidin at 10μg/ml for at least

10 days (MEL-CAS9-NLS) or with Puromycin at 1μg/ml for at least 4 days (MEL-iCAS9). Individual clones for each population were obtained following manual dilution. Induction of CAS9 expression in MEL-iCAS9 was achieved by culturing cells with Doxycyclin at 1μg/ml for at least 2 days.

CRISPR-mediated knock-in MEL and MEDEP cells

For MEL cells, CAS9-expressing cells were either i) transfected with *non-viral plasmids expressing sgRNAs* plus n*on-viral plasmids with HDR template* or ii) transduced with virions from *plasmids with HDR template and expressing sgRNAs*. For MEDEP cells, transfection with *plasmids expressing Cas9 and sgRNAs* plus n*on-viral plasmids with HDR template* was performed. In each case, at day 4, the presence of correct integration events was confirmed by PCR (see screening primers sequence are listed in the key resources table). Then, a pool of GFP+ or Cherry+ cells were sorted and expanded. At day 7, individual GFP+ or Cherry+ cells were sorted into single wells.

METHOD DETAILS

Plasmid constructions

Non-viral plasmids expressing sgRNAs with or without Cas9 were generated by inserting annealed sgRNA primers (key resources table) into BbsI of pSpCas9(BB)-2A-Puro (Addgene PX459) to obtain pSpCas9(BB)-2A-Puro-sgMyb, -sgZeb1 and -sgNuma. A pSp(BB)-2A-Puro version of each construct was obtained by deleting the CAS9 sequence to express sgRNAs only.

Plasmids bearing the HDR templates (pUC19-Myb-Linker-GFP, pUC19-Zeb1-Linker-GFP and pUC19-Numa-Linker-cherry) were obtained by cloning two HDR fragments of ~700 bp surrounding the stop codon (5'HR and 3'HR) separated by an in-frame protein linker[29] and corresponding fluorescent protein sequences. The same strategy was used with an additional in frame P2A-Hygro following GFP for *Myb* and *Zeb1* vector constructs to get pUC19-Myb-Linker-GFP-P2A-Hygro and pUC19-Zeb1-Linker-GFP-P2A-Hygro. Alternatively, a DTA (diphteria toxin gene) cassette has been added in the backbone of the vector to get pUC19-Myb-Linker-GFP-DTA and pUC19-Zeb1-Linker-GFP-DTA.

Viral plasmids expressing Cas9: pLentiCRISPRv2_Neo-short was obtained by deleting the Acc65I/EcoRI fragment from the pLentiCRISPRv2_Neo vector (Addgene 98292), and TLCV2-T2A-noGFP-puro by deleting the BamHI/NheI GFP fragment from TLCV2 (Addgene 87360). There were used together with pLX_311 (Addgene 96924) to test the level of Cas9 expression.

Viral vectors containing the HDR templates and sgRNAs expression cassettes were obtained by inserting annealed sgRNA primers into the BsmBI site of pLentiGuideMCS (derived from Addgene 117986 by deleting SmaI/MluI Puro fragment) to obtain pLentiGuideMCS-sgMyb et -sgZeb1. KI fragment of interest from the different pUC19 constructs were subsequently cloned to get pLentiGuideMCS-sgMyb-Myb-Linker-GFP, pLentiGuideMCS-sgZeb-Zeb-Linker-GFP, pLentiGuideMCS-sgMyb-Myb-Linker-GFP-P2A-Hygro, pLentiGuideMCS-sgZeb-Zeb-Linker-GFP-P2A-Hygro, pLentiGuideMCS-sgMyb-Myb-Linker-GFP-DTA and pLentiGuideMCS-sgZeb-Zeb-Linker-GFP-DTA.

Cell culture and cell handling procedures

HEK-293T and MEL cells were cultured in DMEM (41965-039 Gibco) +10% SVF + Penicillin (100 U/mL)/Streptomycin (100 μg/mL). MEDEP cells were cultured in IMDM (319800-022 Gibco) + 15% SVF + Penicillin (100 U/mL)/Streptomycin (100 μg/mL) + ITS (Insuline Transferrin Sodium selenite Media supplement) 1X (Gibco), ascorbic acid (50μg/mL, Sigma) + α-thioglycerol (45nM, Sigma) + Dexamethasone (10^{-7} M, Sigma) + mSCF (100 ng/mL). All cells were grown at 37°C in 5% CO2.

For growth curves, cells were plated at a density of 0.15 x 10^6 cells / ml and were counted every day (2 days for MEL, 4 days for MEDEP).

Differentiation was induced in MEL cells by adding 2% DMSO (Sigma) and in MEDEP by replacing Dexamethasone and mSCF by erythropoietin (4U/mL).

Transfection methods such as Neon (Invitrogen), *Trans*IT-X2® (Mirus), GenCellin™-HTC (BioCell Challenge) and TurboFect (ThermoScientific) were compared according to the manufacturer instructions using the pSpCas9(BB)-2A-GFP (PX458 Addgene). For Neon in MEL and MEDEP, 3 pulses, 1450V, 10ms on 20.10^6 cells/mL in 10 μL tips with 1 or 2 μg DNA were used, respectively.

Transductions: Virions from the different viral plasmids (*expressing Cas9* or *with HDR template and expressing sgRNAs*) were obtained by calcium phosphate transfection in HEK-293T cells and transduced in MEL or MEDEP cells.

Flow cytometry and cell sorting

Cell pools or individual clones were analyzed using a NovoCyte 1 or 2 (ACEA Biosciences Inc.) and data analyses were performed using the NovoExpress Software (Agilent). Cell sorting was performed with a FACS ARIA IIu (Becton Dickinson).

Screening for knock-in

Genomic DNA was extracted from cell pools or individual clones using QuickExtract™ DNA Extraction Solution (Lucigen®) according to the manufacturer instructions (screening primers listed in the key resources table).

RT-qPCR

Total RNAs were extracted using High Pure RNA Isolation Kit (Roche) and reverse transcribed with the SuperScript™ II (Invitrogen). cDNA were amplified in triplicate with the PowerUp™ SYBR™ Green Master Mix (Applied Biosystems) using *Myb*, *Epb4.2*, *Gypa* and *Prkag1* (as internal control) primers (qPCR primers listed in the oligonucleotides section of the key resources table).

Western blotting

Cells were lysed on ice with the Triple Detergent Lysis Buffer (50 mM Tris-HCl, 150 mM NaCl, 0,02% sodium azide, 0,1% SDS, 1 mM PMSF) containing complete EDTA-free protease inhibitor cocktail (Roche). Lysates were clarified by centrifugation at 14,000 rpm for 15 minutes at 4°C. Whole cell extracts (40μg) were boiled in western-loading buffer, resolved by Bolt™ Bis-Tris Plus gel (Invitrogen™) and transferred onto nitrocellulose membrane (Amersham™ Protran®). Western blot analyses were performed using the indicated antibodies (listed in the antibody section of the key resources table) and captured on an Odyssey® M (LI-COR).

CUT&RUN

CUT&RUN targeted chromatin profiling was performed for MYB and GFP on WT MEL cells and homozygous *Myb*-GFP MEL cells, as described by Henikoff and colleagues (https://doi.org/10.7554/eLife.46314). In brief, for each condition, 500.000 cells were harvested, after two washes, cells were bound to 10 μl of BioMag Plus Concanavalin A-coated beads (Bangs Laboratories, cat. no. BP531) diluted in 1ml of binding buffer and rotated 10 min at RT. Beads were cleared on a magnetic rack and resuspended in 100 μl of antibody solution (1/100 antibody dilution, key resources table) before an overnight incubation at 4°C on a thermomixer. Beads were then cleared on a magnetic rack, washed once in digitonin buffer and resuspended in 100 μl of digitonin buffer containing 700 ng/μl pA-MNase. After 1h rotating at 4°C, beads were cleared on a magnetic rack and washed twice in digitonin buffer and once in low salt buffer. To activate MNase activity, beads were resuspended in 200 μl of activation buffer containing CaCl2 and incubate 15 min at 0°C. Reaction was stopped by resuspending the beads in 200 μl of stop buffer. Chromatin was released from the beads for 30 min at 37°C. The cleared supernatant was then purified using the NucleoSpin Gel and PCR clean up kit (Macherey-Nagel, cat. no. 740609.250). DNA was stored at −20°C before library preparation.

CUT&RUN libraries were prepared from 0.1 to 10 ng of double-stranded DNA using the MicroPlex Library Preparation kit v3 (C05010001, Diagenode, Seraing, Belgium), according to manufacturer's instructions with some modifications. In the first step, the DNA was end-repaired. In the next step, stem-loop adaptors were ligated to the 5 prime end of the genomic DNA. In the final step, the 3 prime ends of the genomic DNA were extended to complete library synthesis and Illumina compatible indexes were added through a PCR amplification (3 min at 72°C, 2 min at 85°C, 2 min at 98°C; [20 sec at 98°C, 10 sec at 60°C] x 14 cycles). Amplified libraries were purified and size-selected using Agencourt SPRIselect beads (Beckman Coulter) by applying the following ratio: volume of beads / volume of libraries = 1,4 / 1. Libraries were sequenced on an Illumina HiSeq 4000 sequencer as paired-end 100 base reads.

CUT&RUN and ChIP-seq data analysis

Raw reads were subjected to quality control and adapter trimming using Trimgalore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ ; version 0.6.6). Trimmed reads were then aligned to the reference genome (mm9) using STAR aligner (version 2.7.9a)[48] for CUT&RUN and bowtie2 (version 2.3.4.3)[49] for ChIP-seq data. The aligned reads were converted from SAM to BAM format, sorted, and indexed using Samtools (version 1.11).[50] Duplicate reads were removed using Picard tool (version 2.17.11) (http://broadinstitute.github.io/picard/ only for CUT&RUN data. Peak calling was performed using HOMER (version 4.10.3),[51] with the following parameters: FDR rate threshold=0.001; FDR effective Poisson threshold=3.446e-06; fold over input required=3; Poisson p-value over input required=1.00e-04. The resulting peaks were annotated using ChIPseeker (version 1.34.1)[52,53] and ChIPpeakAnno (version 3.32.0) R packages.[54,55] Enrichment analysis was performed using clusterProfiler R package (version 4.6.0).[56,57] Heatmaps and meta-profiles were generated using the deepTools (version 3.4.2).[58] Motif enrichment analysis was performed using meme-chip from MEME suite (version 5.3.3).[59,60] For this analysis, we used the summit of the peak -/+ 500 bp.

Microscopy

Live visualization of GFP- and Cherry-KI cells was performed using an LSM980 Airyscan 8Y (Zeiss) confocal microscope. For fixed cells, GFP and Cherry were stained using GFP-Booster_ATTO488 and RFP-booster_ATTO594 (ChromoTek) respectively, according to the manufacturer instructions. Nuclei were stained with 10 μg/mL of 4',6 diamidino-2-phenylinolole dihydrochloride (Sigma-Aldrich). To visualize the fluorescence, a LSM980 Airyscan 8Y (Zeiss) confocal microscope was used, and images were acquired with ZEN Blue software (Zeiss).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical methods used for genome-wide studies are explained in detail under the respective method sub-section. Growth curves and gene expression measurements are represented as mean ± SEM of n=3 to 6 independent experiments as indicated in the figure legends.