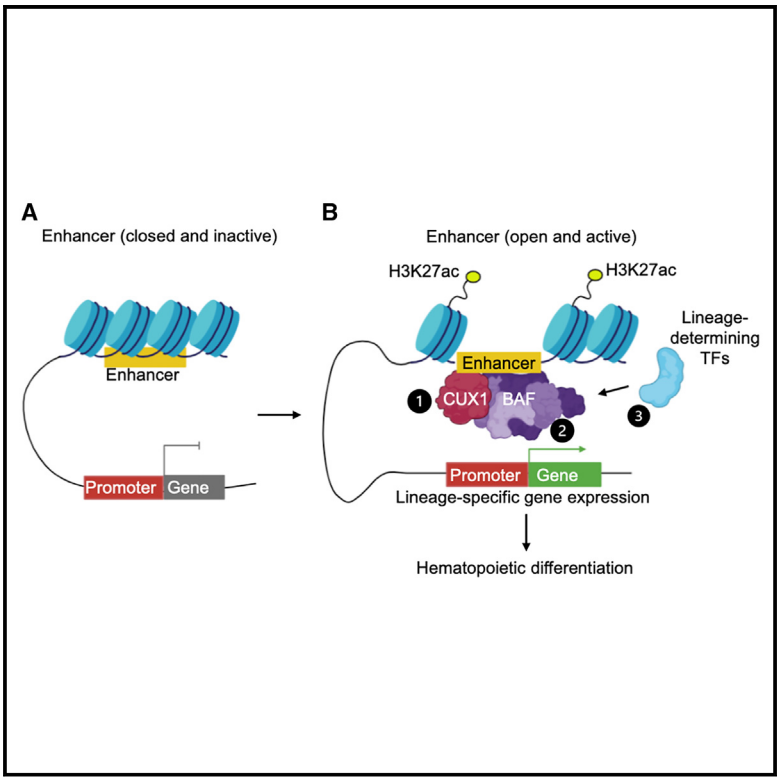
Cell Reports

CUX1 regulates human hematopoietic stem cell chromatin accessibility via the BAF complex

Graphical abstract



Highlights

- CUX1 recruits the BAF chromatin remodeling complex to enhancers to increase DNA accessibility
- BAF complex binding to chromatin is partially dependent on CUX1
- CUX1 maintains accessibility at lineage-specific enhancers in human hematopoietic stem cells
- CUX1-regulated genes are predictive of stem cell lineage determination *in vivo*

Authors

Weihan Liu, Jeffrey L. Kurkewich, Angela Stoddart, ..., Lia Jueng, Stephen J. Kron, Megan E. McNerney

Correspondence

megan.mcnerney@bsd.uchicago.edu

In brief

Liu et al. reveal that the hematopoietic transcription factor CUX1 possesses pioneer factor activity in hematopoietic stem and progenitor cells. CUX1 recruits the BAF chromatin remodeling complex to promote the DNA accessibility of enhancers associated with differentiation.

Liu et al., 2024, Cell Reports *43*, 114227
May 28, 2024
https://doi.org/10.1016/j.celrep.2024.114227

Cell Reports

CellPress
OPEN ACCESS

Report

CUX1 regulates human hematopoietic stem cell chromatin accessibility via the BAF complex

Weihan Liu,[1,3,6] Jeffrey L. Kurkewich,[1,6] Angela Stoddart,[1] Saira Khan,[1] Dhivyaa Anandan,[1] Alexandre N. Gaubil,[1] Donald J. Wolfgeher,[5] Lia Jueng,[1] Stephen J. Kron,[2,3,5] and Megan E. McNerney[1,2,3,4,7,*]

[1]Department of Pathology, The University of Chicago, Chicago, IL 60637, USA
[2]The University of Chicago Medicine Comprehensive Cancer Center, The University of Chicago, Chicago, IL 60637, USA
[3]Committee on Cancer Biology, The University of Chicago, Chicago, IL 60637, USA
[4]Department of Pediatrics, Section of Hematology/Oncology, The University of Chicago, Chicago, IL 60637, USA
[5]Department of Molecular Genetics and Cell Biology, The University of Chicago, Chicago, IL 60637, USA
[6]These authors contributed equally
[7]Lead contact
*Correspondence: megan.mcnerney@bsd.uchicago.edu
https://doi.org/10.1016/j.celrep.2024.114227

SUMMARY

CUX1 is a homeodomain-containing transcription factor that is essential for the development and differentiation of multiple tissues. *CUX1* is recurrently mutated or deleted in cancer, particularly in myeloid malignancies. However, the mechanism by which CUX1 regulates gene expression and differentiation remains poorly understood, creating a barrier to understanding the tumor-suppressive functions of CUX1. Here, we demonstrate that CUX1 directs the BAF chromatin remodeling complex to DNA to increase chromatin accessibility in hematopoietic cells. CUX1 preferentially regulates lineage-specific enhancers, and CUX1 target genes are predictive of cell fate *in vivo*. These data indicate that CUX1 regulates hematopoietic lineage commitment and homeostasis via pioneer factor activity, and CUX1 deficiency disrupts these processes in stem and progenitor cells, facilitating transformation.

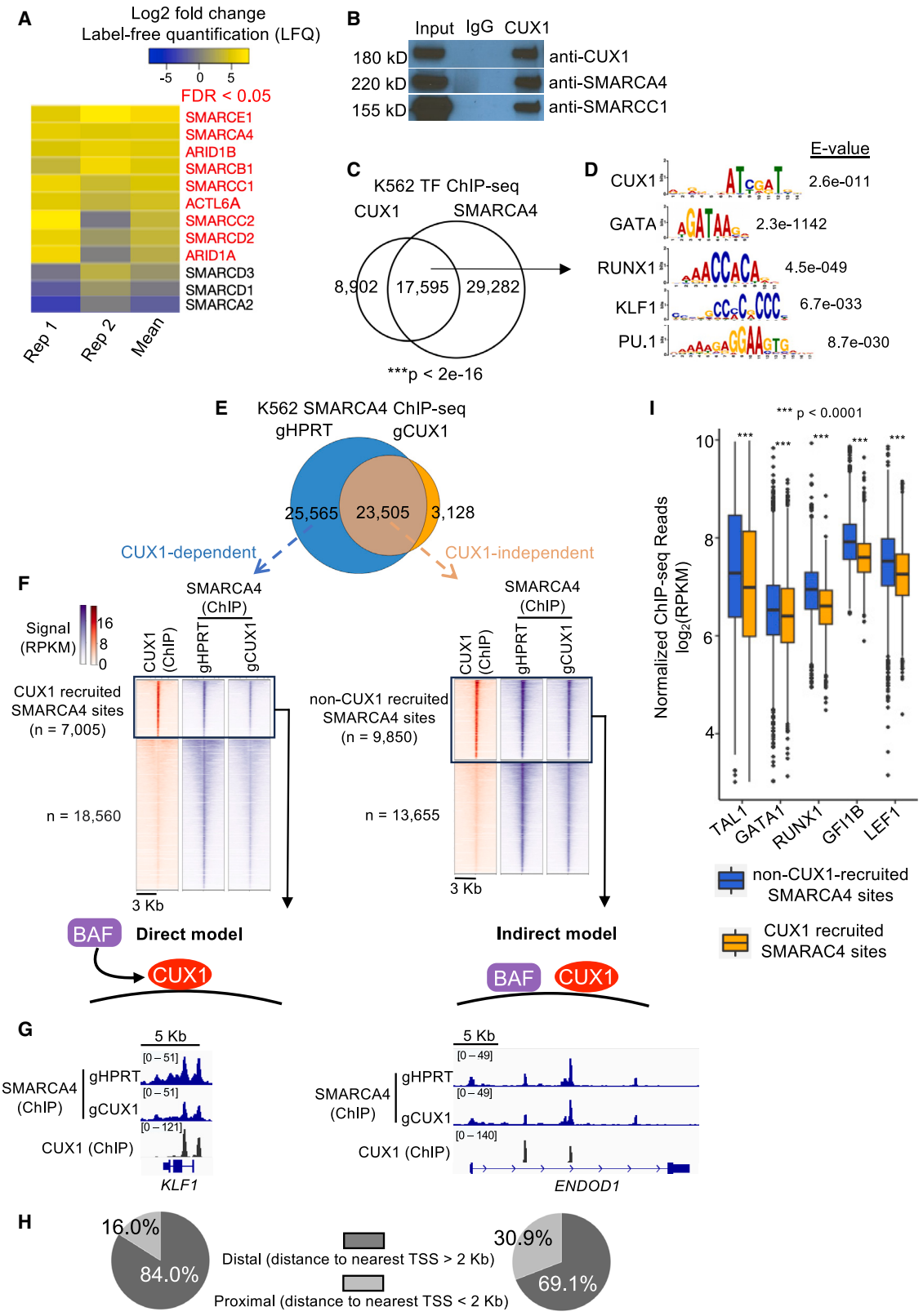
INTRODUCTION

Multipotent tissue-resident stem cells are essential for the maintenance and function of adult tissues. Defects in stem cell homeostasis and lineage commitment underly myriad human diseases, including cancer.[1] The mechanisms governing lineage determination are incompletely understood and remain a fundamental question in developmental biology. Insight into these processes is central to the identification of therapeutic interventions for diseases of stem cell dysfunction.

While cell fate decisions are influenced by extrinsic factors, such as cell-cell signaling and growth factors, intrinsic factors, such as transcription factors (TFs) and other epigenetic regulators, are ultimately responsible for integrating these cues to guide the genomic reprogramming required for cell-type-specific gene expression.[2] Here we define epigenetic regulators as proteins that can remodel chromatin or chemically modify DNA and chromatin without altering the DNA sequences. Upon DNA binding, TFs regulate transcription via the recruitment of proteins that physically remodel nucleosomes, enzymatically modify histones and DNA, and/or regulate RNA polymerase machinery directly.[3] TFs that bind nucleosomal DNA and promote *de novo* DNA accessibility for "settler" TF binding have been described as "pioneer" factors.[4] Pioneer factors mediate this process via recruitment of nucleosome remodeling enzymes such as the SWI/SNF (or BAF, BRG/BRM-associated factor) complex.[4]

BAF complexes are composed of 10–13 subunits and have essential roles in transcriptional activation, DNA repair, and development.[5] BAF activity is required for reconfiguring nucleosomes at enhancers vital for lineage-specific gene expression.[6] To alter nucleosome position and/or content, enzymatic BAF subunits hydrolyze ATP.[5] However, BAF proteins lack intrinsic DNA binding domains and depend on TFs for DNA targeting specificity. In hematopoiesis, important TFs such as RUNX1, PU.1 (SPI1), and KLF1 interact with BAF complex components to promote hematopoietic development.[7–9] Nonetheless, these TFs only account for a portion of BAF chromatin binding, implicating additional, yet unknown, hematopoietic pioneer TFs.

CUT-like homeobox 1 (CUX1, also referred to as CUTL1 or CDP/CUT), is a widely expressed, non-clustered homeodomain-containing TF. CUX1 is critical in a broad range of cellular functions in a variety of tissue types, including differentiation of neural, lung, and hematopoietic tissues.[10–13] Germline mutations in *CUX1* are associated with developmental delay, and somatic mutations are found in cancer.[14,15] In myeloid malignancies, such as acute myeloid leukemia and myelodysplastic syndrome, *CUX1* deletions or inactivating mutations are recurrent and carry a poor prognosis.[16] *CUX1* mutations or deletions are also recurrent in clonal hematopoiesis and carry an increased risk of subsequent transformation.[17,18] We reported that CUX1 regulates



(legend on next page)

hematopoietic stem and progenitor cell (HSPC) homeostasis and differentiation, including a role for CUX1 in promoting erythropoiesis while inhibiting myelopoiesis.[13] These data indicate that CUX1 exerts tumor suppressor activity via transcriptional regulation of HSPC functions, yet the mechanism by which CUX1 coordinates gene expression remains unclear.

CUX1 has four DNA binding domains: three CUT repeats and one homeodomain. Genome-wide, CUX1 binding is enriched at enhancers, particularly those in active contact with promoters, indicating long-distance looping of CUX1-bound enhancers to promoters.[19,20] Depending on the context, CUX1 has been found to have transcriptional activating and repressive capabilities, and multiple mechanisms have been proposed for each. In the repressive context, CUX1 has been reported to recruit histone deacetylases or alternately compete with transcriptional activators for DNA binding sites.[21–23] In the activating context, CUX1 can activate gene expression through collaborating with E2F TFs or co-activators such as the histone acyltransferase P300.[24,25] However, these studies were largely performed with reporter assays outside the native chromatin context, so the endogenous, genome-wide functions of CUX1 remain unclear. Together, these data indicate that CUX1 is an epigenetic modifier that interfaces with higher order chromatin structure, yet the molecular mechanism by which CUX1 controls transcription is incompletely understood. In this study, we address this question by identifying endogenous CUX1 interacting partners, CUX1 genomic targets, and the ensuing epigenetic consequences through unbiased proteomics and genome-wide functional genomics approaches in a human leukemia cell line and primary human HSPCs.

RESULTS

CUX1 recruits the BAF chromatin remodeling complex to enhancers

To determine the mechanism by which CUX1 governs gene expression, we identified CUX1 protein interaction partners by performing co-immunoprecipitation for endogenous CUX1 followed by mass spectrometry in the K562 human myeloid leukemia cell line. We chose K562 cells for this experiment as they are considered human leukemic representatives of multipotent progenitors, capable of differentiation into erythroid, megakaryocytic, and myeloid lineages.[26–28] This analysis revealed nine components of the BAF complex interacting with CUX1 (false discovery rate [FDR] < 0.05) (Figure 1A). Many of the protein subunits identified are shared across the three major BAF complexes; however, the detection of ARID1A and ARID1B suggests that CUX1 interacts with the canonical BAF complex (cBAF).[29] CUX1 interactions with two core BAF complex members, SMARCA4 (BRG1) and SMARCC1 (BAF155), were confirmed by western blot (Figure 1B).

We next tested if CUX1 and BAF bind to overlapping genomic loci. We performed chromatin immunoprecipitation with massively paralell DNA sequencing (ChIP-seq) for CUX1 and SMARCA4, the essential enzymatic BAF subunit.[30] Using the thresholded peak calling method by MACS2 and irreproducible discovery rate (IDR) analysis,[31,32] in total, 66.4% (17,595/26,497) of CUX1 binding sites overlapped with SMARCA4 peaks, revealing extensive overlap of CUX1 and SMARCA4 on DNA (Figure 1C). CUX1 and SMARCA4 overlapping sites were localized predominantly at enhancers (Figure S1A) and enriched for the hematopoietic TF motifs GATA, RUNX1, KLF1, and PU.1 (Figure 1D). Significant overlap of CUX1/SMARCA4 co-bound sites with published ChIP-seq data[33,34] shows that CUX1 and the BAF complex interact with other hematopoietic TFs at enhancers (Figure S1B).

We next tested the hypothesis that CUX1 recruits BAF to DNA. We performed ChIP-seq for SMARCA4 in K562 clones CRISPR-Cas9 edited for *CUX1* (gCUX1) or a control intronic region of *HPRT* (gHPRT[35]; Figure S1D). Among the 49,070 (IDR < 0.05)[32] SMARCA4 binding sites identified in gHPRT control cells, 52.1% (25,565) were reduced in gCUX1 cells (CUX1-dependent SMARCA4 sites) (Figure 1E). An example of the reduction of SMARCA4 binding after CUX1 knockout is shown at the *HMBS* gene, encoding the essential erythrocyte hydroxymethylbilane synthase enzyme (Figure S1C).[36] This experiment shows that CUX1 promotes recruitment of the BAF complex to bind certain loci.

Next, we interrogated if CUX1 directly recruits SMARCA4. An intersection of CUX1-dependent SMARCA4 sites with CUX1 binding sites revealed that 27.4% (7,005/25,565) of CUX1-dependent SMARCA4 sites are at loci directly bound by CUX1 (Figure 1F, left). In this "direct model," CUX1 promotes recruitment of SMARCA4 to a substantial fraction of DNA binding sites. We next examined CUX1 binding at the CUX1-independent SMARCA4 sites. To this end, we intersected CUX1-independent SMARCA4 sites with CUX1 ChIP-seq peaks. 41% (9,850/23,505) of these sites were bound to CUX1 (Figure 1F, right). This finding suggests that while CUX1 is not necessary for SMARCA4 binding at these loci, SMARCA4 may still be co-bound with CUX1, referred to herein as an "indirect model" of SMARCA4 binding. These 7,005 and 9,850 sites are referred to

Figure 1. CUX1 recruits the BAF chromatin remodeling complex to enhancers

(A) Co-immunoprecipitation for CUX1 in K562 cells was followed by mass spectrometry ($n = 2$ biological replicates). The heatmap indicates BAF members ranked by the mean label-free quantification fold enrichment compared to IgG controls, the range of which is indicated in the scale bar. (Significance calculated by Student's t test, FDR < 0.05.)

(B) Representative co-immunoprecipitation followed by immunoblot in K562 ($n = 2$ biological replicates).

(C) K562 CUX1 and SMARCA4 ChIP-seq overlap ($n = 2$ biological replicates, IDR < 0.05, significance of overlap calculated by hypergeometric test).

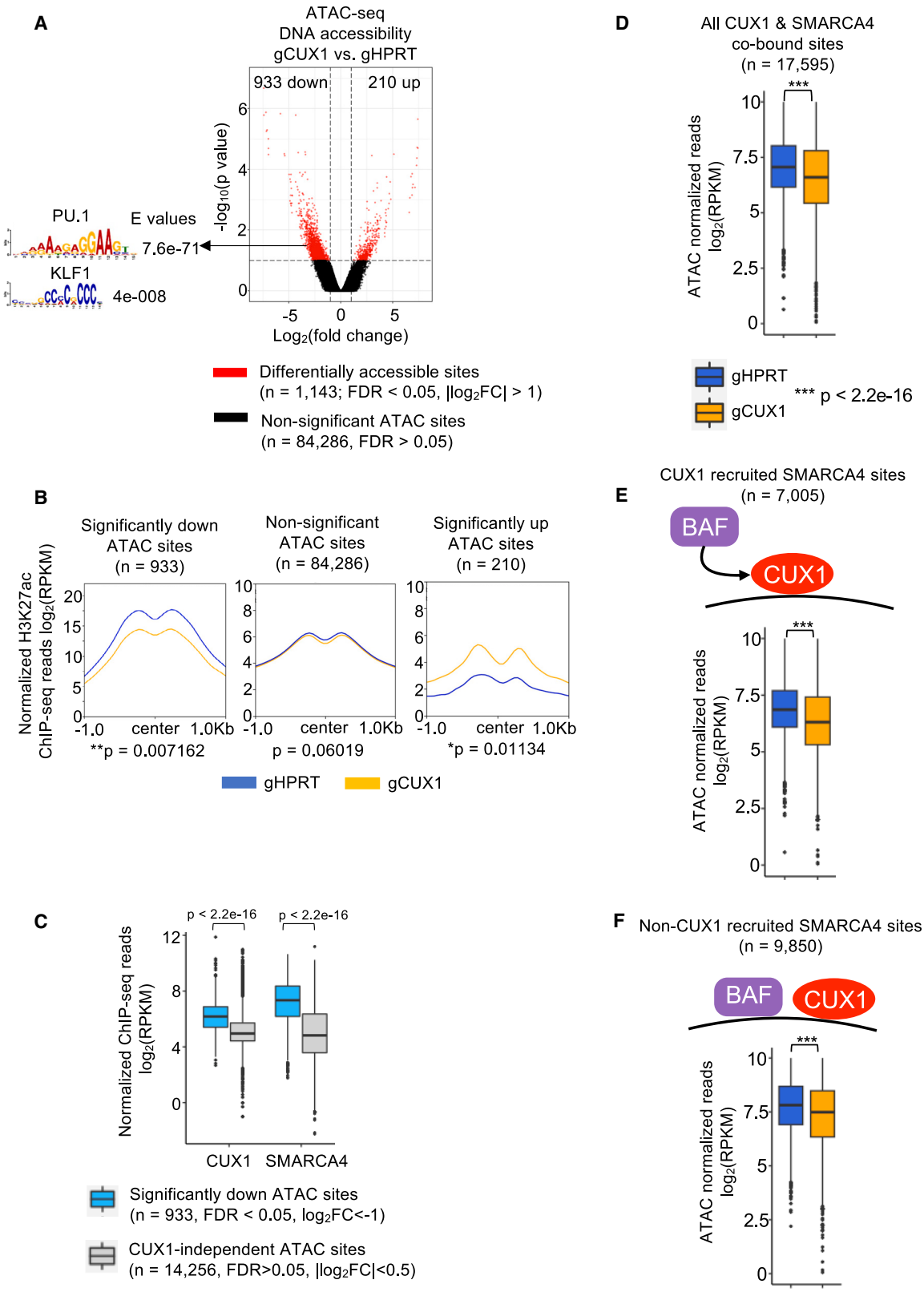
(D) Enriched motifs[84] at CUX1 and SMARCA4 co-occupied sites.

(E) Overlap of SMARCA4 peaks ($n = 2$ biological replicates, IDR < 0.05) in gHPRT and gCUX1 K562 cells.

(F) Heatmaps showing overlap between CUX1-dependent or CUX1-independent SMARCA4 sites with CUX1. The values are normalized ChIP-seq reads (RPKM). The direct model represents CUX1 recruitment of SMARCA4. The indirect model represents SMARCA4 sites bound but not recruited by CUX1.

(G) Example genome snapshots for each category are shown.[85]

(H and I) Distance to the nearest transcription start site (TSS) (H) of CUX1-recruited and non-CUX1-recruited SMARCA4 sites and hematopoietic TF occupancy (I).



(legend on next page)

hereafter as "CUX1-recruited SMARCA4" and "non-CUX1-recruited SMARCA4" sites, respectively. Example genome snapshots of these two categories are shown in Figure 1G (left: *KLF1*, encoding a TF essential for erythropoiesis[37,38]; right: *ENDOD1*, encoding a nucleic acid hydrolyzation nuclease).[39] To understand the differences between these two categories, we further characterized the underlying features of these sites. While 69.1% of non-CUX1-recruited SMARCA4 sites are at distal regulatory elements, this increases to 84% for CUX1-recruited SMARCA4 sites, suggesting that CUX1 recruits SMARCA4 to many distal enhancers (Figure 1H). Analyses with ENCODE ChIP-seq data show that the CUX1-recruited SMARCA4 sites are enriched for lineage-specifying TFs (Figure 1I).[33] These data are compatible with a model wherein CUX1 promotes BAF recruitment, particularly at enhancers occupied by lineage-directing TFs.

CUX1 with SMARCA4 promotes the establishment of accessible chromatin

As recruitment of BAF is one mechanism through which pioneer TFs remodel chromatin,[4] we next assessed the role of CUX1 in the regulation of DNA accessibility. We performed ATAC-seq[40] (the assay for transposaseaccessible chromatin with sequencing) on gCUX1 and control gHPRT K562 cells. To investigate the effect of CUX1 on DNA accessibility, we first applied a non-thresholded, quantitative approach. To this end, we performed genome-wide differential accessibility analysis using csaw[41] on the ATAC-seq data and observed more sites with significantly downregulated (n = 933) than upregulated (n = 210) accessibility after loss of CUX1 (FDR < 0.05, $|\log_2 FC| > 1$) (Figure 2A), indicating that CUX1 normally contributes to chromatin opening. Among the 933 significantly decreased ATAC sites, a considerable proportion (38.1%) are at predicted enhancers (Figure S2A) and are enriched for PU.1 and KLF1 motifs (Figure 2A). The changes in accessibility were accompanied by concordant changes in the activating chromatin mark H3K27ac, indicating that CUX1 maintains enhancer activation and accessibility in K562 cells (Figure 2B). Overall, these data are consistent with the model that CUX1 promotes chromatin accessibility at enhancers involved in hematopoietic differentiation.

We assessed CUX1 and SMARCA4 occupancy at differentially accessible sites. Only a small percentage of the significantly downregulated ATAC sites overlap CUX1 and/or SMARCA4 binding sites using strict peak calling thresholds (IDR = 0.05). This may stem from false-negative ChIP-seq data, as SMARCA4 does not bind DNA directly, and CUX1 has low DNA-binding affinity.[42] To address this, we used an approach that does not rely on thresholded peak calling. By this analysis, both CUX1 and SMARCA4 occupancy are significantly higher at the significantly downregulated ATAC sites (933 peaks) compared to the CUX1-independent ATAC sites (14,256 peaks with the least significant change in chromatin accessibility, FDR > 0.05, |log2FC| < 0.5) (Figure 2C). Further, SMARCA4 occupancy decreases significantly at these sites upon CUX1 loss (Figure S2B). The finding that sites normally opened by CUX1 have higher occupancy of both CUX1 and SMARCA4 implicates direct involvement of these factors in driving chromatin accessibility.

Analysis of both CUX1 and SMARCA4 co-bound sites and CUX1-recruited SMARCA4 sites demonstrated significant drops in accessibility after CUX1 knockout (p < 2.2e−16, Figures 2D and 2E). We performed ATAC-seq after SMARCA4 knockout and observed that 91% of the peaks lost in gCUX1 samples are also lost upon SMARAC4 loss (Figure S2C and S2D). SMARCA4 knockout decreased accessibility at both the significantly down ATAC sites (n = 933) and CUX1-recruited SMARCA4 sites (n = 7,005) (Figure S2E). These observations are consistent with a model in which CUX1 recruits BAF to enhancers and increases DNA accessibility. Unexpectedly, CUX1 also influences accessibility independent of its ability to directly recruit the BAF complex (i.e., the non-CUX1-recruited SMARCA4 sites) (Figure 2F). While not tested here, this later finding may be due to CUX1 recruitment of additional activating factors, such as HATs,[24,25] or downstream indirect effects of CUX1 loss.

In human HSPCs, CUX1 maintains DNA accessibility at enhancers associated with SMARCA4 and hematopoietic differentiation

To observe whether CUX1 co-occupies genomic loci with BAF components in primary human CD34$^+$ HSPCs, we used cleavage under targets and release using nuclease (CUT&RUN) in lieu of ChIP-seq as CUT&RUN requires fewer cells.[43] Compared to ChIP-seq in K562 cells, CUT&RUN in CD34$^+$ cells showed fewer peaks and a relative enrichment for CUX1 and SMARCA4 at promoter-proximal binding sites (Figure 3C). The decreased peak number and shift in peak location of CUX1 in CD34$^+$ cells compared to K562 (Figure 1C) is likely due to technical differences between the assays, as CUX1 CUT&RUN in K562 also shows fewer peaks and promoter enrichment compared to ChIP-seq (Figures S3A and S3B). Nonetheless, 52.1% (3,519/6,758) of CUX1 binding sites overlap those of SMARCA4 (Figure 3A), and CUX1 and SMARCA4 binding signals are highly correlated with each other genome-wide (Spearman's ρ = 0.71, p < 2.2e−16) (Figure 3B). CUX1 and SMARCA4 binding signals remain positively correlated with activating chromatin marks in HSPCs from the NIH Roadmap Epigenomics database,[44] with correspondingly higher correlations with H3K4me3, associated with promoters (Figure 3B). The CUX1/SMARCA4 co-bound

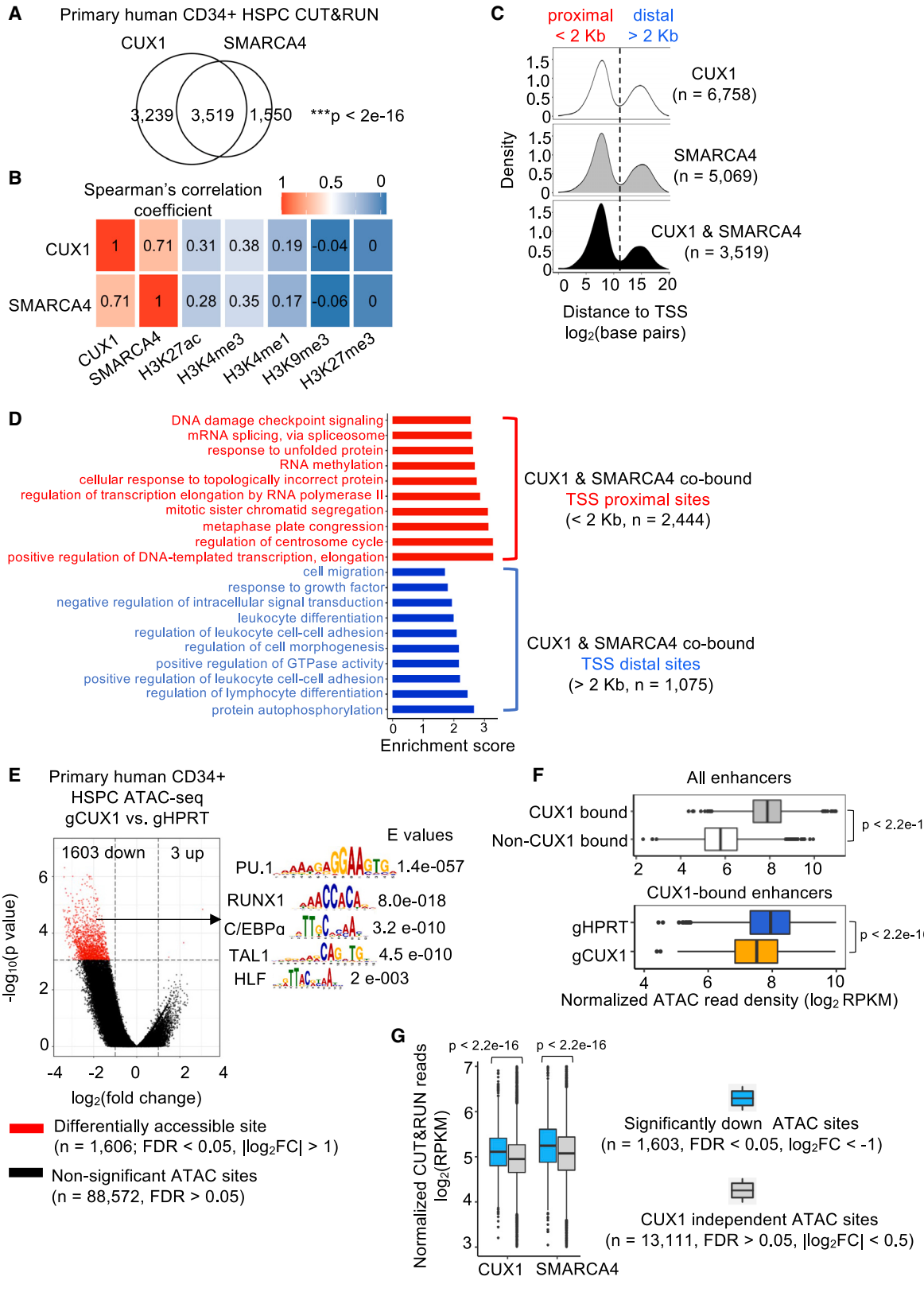
Figure 2. CUX1 with SMARCA4 promotes the establishment of accessible chromatin

(A) Volcano plot comparing ATAC-seq signal in gCUX1 vs. gHPRT K562 cells (n = 2 biological replicates). Significance calculated by csaw.[41] Top enriched motifs for the significant down sites are shown.

(B) H3K27ac ChIP-seq reads (n = 2 biological replicates) at significantly down, up, and non-significant ATAC sites.

(C) CUX1 and SMARCA4 occupancy at down (n = 933) vs. CUX1-independent ATAC sites (n = 14,256).

(D–F) ATAC-seq signal from gHPRT and gCUX1 cells for CUX1 and SMARCA4 co-occupied sites (D), CUX1-recruited SMARCA4 sites (E), and SMARCA4 sites bound but not recruited by CUX1 (F). Significance for (B)–(F) calculated by two-sided Wilcoxon rank-sum test.



(legend on next page)

sites at promoter-proximal (n = 2,444) and -distal (n = 1,075) regions were assigned to the single nearest gene using GREAT and functionally annotated using AMIGO.[44–48] Notably, the distal genes were enriched for processes involved in cellular differentiation and morphogenesis. In comparison, the proximal genes were enriched for more general cellular processes such as transcription and mitosis (Figure 3D). Examples of CUX1 and SMARCA4 co-occupancy at enhancers of genes important for multilineage hematopoietic cell differentiation, *FLT1* and *RUNX1*,[49–51] and at promoters of the mitosis and DNA transcription related genes, *TUBB* and *MED18*, are shown (Figures S4A and S4B).[52,53]

To assay accessibility following CUX1 loss in primary cells, we transfected human CD34$^+$ HSPCs with CRISPR gRNAs targeting *HPRT* and *CUX1* for ATAC-seq analysis 48 h post transfection. The mean editing efficiency of *CUX1* was 49.5% and 75.5% for *HPRT*. Differential accessibility analysis using csaw[41] showed that 1,603 sites were significantly lost (FDR < 0.05, $\log_2$FC < −1), and only 3 were gained (FDR < 0.05, $\log_2$FC > 1) after *CUX1* editing, confirming that CUX1 promotes open DNA accessibility in primary HSPCs (Figure 3E). Most of the significantly lost ATAC-seq sites are located at predicted enhancers (Figure S5A) and show an enrichment of multiple hematopoietic TF motifs including PU.1, RUNX1, C/EBPα, TAL1, and HLF (Figure 3E). These TFs play key roles in lineage commitment and maintaining hematopoietic stem cell (HSC) quiescence.[54,55] To further quantify the effect of CUX1 on enhancer accessibility, we obtained 3,902 genome-wide CUX1-bound enhancers by intersecting CUX1 binding sites from CUT&RUN and the human CD34$^+$ chromHMM track from the NIH Roadmap Epigenomics database.[44] Enhancers bound by CUX1 have significantly greater DNA accessibility than enhancers not bound by CUX1 (Figure 3F). Next, we focused on the enhancers directly bound by CUX1 (n = 3,902) and observed that upon CUX1 loss, there is a significant decrease of accessibility (Figure 3F), indicating CUX1 is required to promote open chromatin at enhancer regions.

Lastly, to examine the relationship of CUX1-mediated accessibility with the BAF complex, we quantified CUX1 and SMARCA4 occupancy at the significantly down ATAC sites. Compared to CUX1-independent sites, occupancies of both CUX1 and SMARCA4 are significantly higher at significantly down ATAC sites (Figure 3G). Examples for significant loss of chromatin accessibility following CUX1 knockout are shown at *NCOA4*, which promotes erythropoiesis by regulating ferritin turnover,[56] and *JARID2*, which regulates HSC homeostasis by collaborating with PRC2 and functions as a myeloid tumor suppressor (Figure S5B).[57,58] Taken together, in human HSPCs, CUX1 is directly involved in maintaining chromatin accessibility at enhancers associated with SMARCA4 occupancy and targeting genes regulating hematopoiesis.

CUX1 genomic targets are linked with genome architecture and *in vivo* lineage potential

Previous studies reported that CUX1 binding is highly predictive of enhancer-promoter interactions.[19,20] As we observed a substantial proportion of CUX1 binding at promoter-proximal regions in human CD34$^+$ cells (Figure 3C), we tested if CUX1 binding at these promoters influences accessibility at enhancers looped to those promoters. We intersected 2,684 looping DNA contact points, obtained from Hi-C (high-throughput chromosome conformation capture) analysis of human HSPCs,[59] with CUX1 CUT&RUN data and identified n = 272 DNA loops that contain distal elements in contact with CUX1-bound promoters. Integrating these sites with our ATAC-seq data revealed two findings. First, distal elements in contact with CUX1-bound promoters had overall increased DNA accessibility compared to non-CUX1-bound counterparts (Figure 4A). Second, distal elements in contact with CUX1-bound promoters trend toward decreased accessibility after CUX1 loss (p = 0.085), while there is no change in accessibility for loops not in contact with CUX1-bound promoters (p = 0.89) (Figure 4A). In the converse analysis, we did not observe a significant decrease in accessibility of promoters looped to CUX-bound distal elements (p = 0.32). An example genome snapshot of CUX1 promoting accessibility of enhancers looped to CUX1-bound promoters is shown at *KIT* (Figure 4B).[60] Other hematopoiesis genes targeted by CUX1-bound loops include *MEIS1,*[61,62] *ZFP36L1,*[63] and *LMO4.*[64] In summary, CUX1 binding to promoters is associated with increased accessibility of looped enhancers.

Heretofore, our data suggest that CUX1 with SMARCA4 promotes accessibility for recruitment of TFs that drive differentiation (Figure 3E). To explore the transcriptional consequences of CUX1 loss, we integrated the ATAC-seq with RNA-seq from CD34$^+$ HSPCs with 98 genes upregulated (FDR < 0.1, $\log_2$FC > 0.75) and 334 genes downregulated (FDR < 0.1, $\log_2$FC < −0.75) after CUX1 knockdown.[13] In total, 406/432 of the differentially expressed gene (DEGs) contain significantly

Figure 3. In human HSPCs, CUX1 and SMARCA4 maintain chromatin accessibility at enhancers associated with hematopoietic differentiation

(A) Overlap of CUX1 and SMARCA4 CUT&RUN peaks in primary human CD34$^+$ HSPCs (n = 2 biological replicates, IDR < 0.05, significance of overlap calculated by hypergeometric test).

(B) Genome-wide correlation of CUX1 and SMARCA4 CUT&RUN signals with histone marks from Roadmap Epigenomics.[44] All pairwise correlations have p < 0.001. The scale bar indicates Spearman's correlation coefficient from 0 to 1.

(C) CUX1 and SMARCA4 peaks' absolute distance ($\log_2$ transformed) to the nearest TSS. The dashed line indicates 2 kb.

(D) Top GO terms for TSS-proximal and -distal CUX1/SMARCA4 co-bound sites (Bonferroni corrected p value < 0.05).[45,46]

(E) Volcano plot of ATAC-seq changes in gCUX1 and gHPRT CD34$^+$ HSPCs (n = 2 biological replicates). Significance calculated by csaw.[41] Top motifs for the down sites are shown.

(F) Normalized ATAC reads at genome-wide CUX1-bound enhancers (n = 3,902) and a randomly sampled, size-matched list of enhancers not bound by CUX1 (top). Normalized ATAC reads at CUX1-bound enhancers (n = 3,902) comparing the control gHPRT and gCUX1 conditions (bottom).

(G) Normalized CUT&RUN reads of CUX1 and SMARCA4 in CD34$^+$ HSPC at down vs. CUX1-independent ATAC sites. Significance for (F) and (G) is by two-sided Wilcoxon rank-sum test.

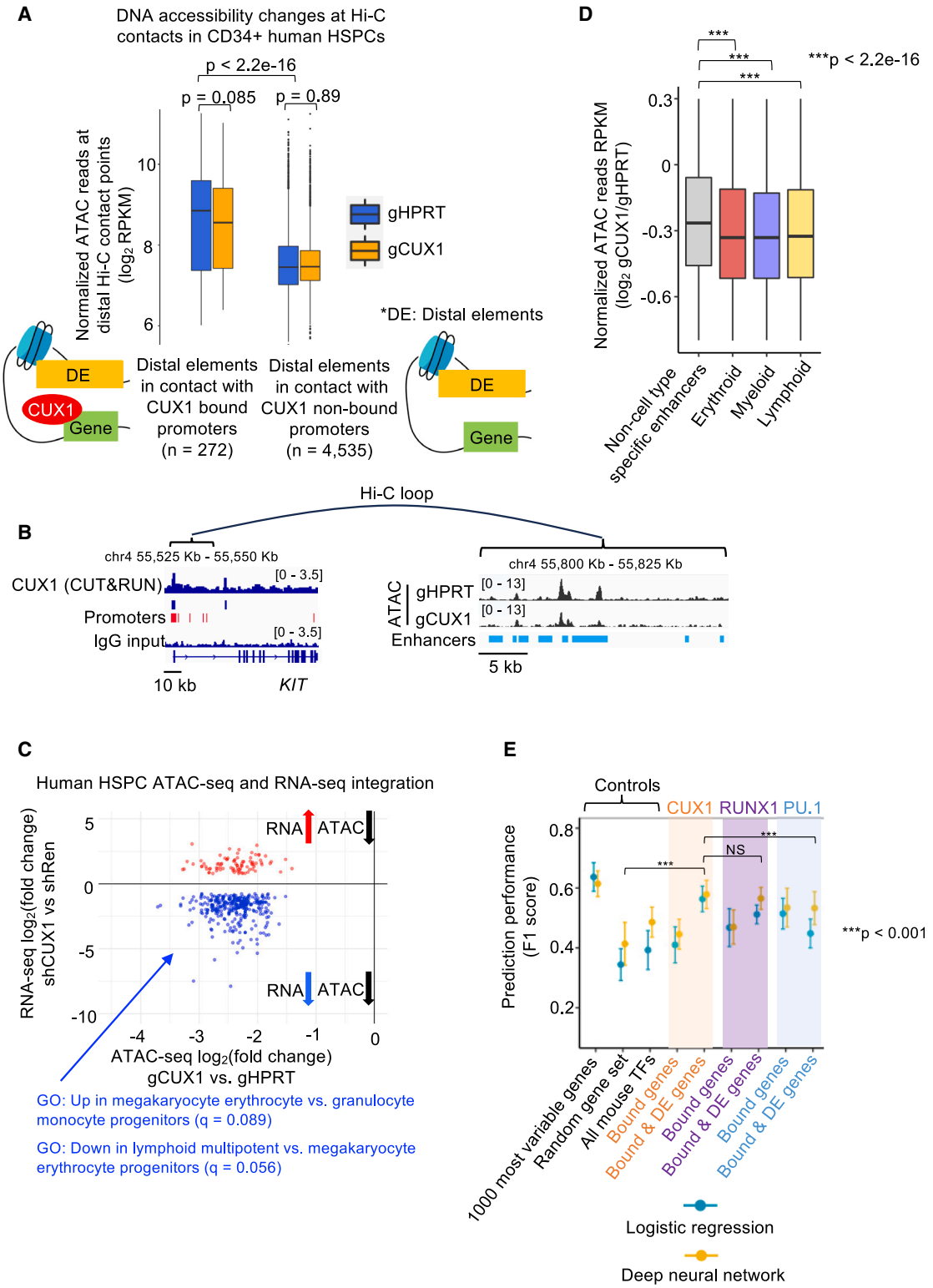


Figure 4. CUX1 genomic targets are linked with genome architecture and *in vivo* lineage potential

(A) ATAC-seq accessibility for gHPRT and gCUX1 CD34$^+$ HSPCs at distal 3D chromatin contact points looped to CUX1-bound promoters from published CD34$^+$ HSPC Hi-C data.[59]

(legend continued on next page)

decreased ATAC-seq sites. Of these 406 genes, 317 have decreased, while only 89 have increased expression (Figure 4C). The proportion of DEGs with simultaneously decreased RNA expression and DNA accessibility is significantly higher than random ($p < 2.2e{-}16$, chi-squared test). This finding links CUX1-dependent increased DNA accessibility with increased target gene expression, as expected. Notably, both CUX1 and SMARCA4 occupancies are higher at these 406 genes than in the background control, demonstrating a positive correlation between the presence of CUX1 and BAF in chromatin accessibility and RNA expression (Figure S6). While Gene Ontology (GO) enrichment analysis revealed no significantly enriched GO terms for the 89 genes with increased RNA levels, those genes that decreased were enriched for genes involved in lineage potential and transcriptional priming (Figure 4C). Therefore, our data indicate that the chromatin accessibility-promoting role of CUX1 in human HSPC is coupled to transcriptional changes in lineage potential.

Lineage-determining TFs bind enhancers to drive cell-type-specific gene expression and terminal differentiation.[65,66] Based on the evidence that CUX1 regulates HSPC cell fate in driving erythroid, myeloid, and lymphoid fate decisions,[13] we hypothesized that CUX1 promotes accessibility at cell-type-specific enhancers. We obtained a list of enhancer annotations specific for each human hematopoietic cell type from the Integrative and Discriminative Epigenome Annotation System database of the VISION project.[67] We then quantified the change in accessibility after *CUX1* editing at these enhancers (Figure 4D). Loss of CUX1 induced a significantly larger drop in accessibility at cell-type-specific enhancers for all hematopoietic lineages, compared to the control, which is a randomly sampled ($n = 10,000$) set of enhancers that did not appear in any cell-type-specific enhancer lists ($p < 2.2e{-}16$). These data suggest that CUX1 preferentially unmasks DNA at lineage-specific enhancers to facilitate hematopoietic maturation.

To examine the role of CUX1 gene regulation in cell fate decisions *in vivo*, we turned to a clonal lineage-tracing dataset coupling murine HSPC single-cell transcriptomic state to progeny cell fates.[68] To test the hypothesis that CUX1 target gene expression is predictive of lineage determination, we used logistic regression and deep neural network classifiers as described by Weinreb et al.[68] Comparable to their studies, a randomly sampled group of genes ($n = 1,000$) and a curated list of mouse TFs (negative controls) returned less than 50% prediction performance measured by F1 score, whereas the top 1,000 most variable genes (positive control) returned F1 scores of ~61%–63%, validating our machine learning models (Figure 4E). While all CUX1-bound genes we identified in HSPCs ($n = 6,758$) could not predict cell fate well (F1 < 45%), CUX1-bound genes with differential expression after CUX1 knockdown ($n = 923$) improved accuracy to 56%–58%. Notably, this performance is similar to the published equivalent gene sets from known HSPC fate-specifying pioneer factors PU.1 (45%–53%) and RUNX1 (51%–57%) (Figure 4E).[69–72] This analysis suggests that CUX1-regulated genes are predictive of HSPC cell fate *in vivo*.

DISCUSSION

Collectively, our data support a model wherein CUX1 recruits BAF to remodel nucleosomes and increase DNA accessibility. Nucleosomes generally act as a barrier to exclude TFs and RNA polymerase machinery from accessing DNA at inappropriate locations and times. A central feature of pioneer factors is the ability to bind to nucleosome-bound DNA. This capability has been ascribed to purified CUX1 *in vitro*.[73,74] While CUX1 binding destabilizes the nucleosome,[74] CUX1 alone does not cause nucleosome displacement.[73] Our results implicate the BAF complex in the subsequent nucleosomal remodeling that is observed in cells.

This sequence of events is apparent in a substantial portion of CUX1 DNA binding sites, exemplified by the "direct model" of CUX1-dependent SMARCA4 recruitment and increased DNA accessibility (Figures 1F and 2E). We also observed a similar number of CUX1 binding events that were not required for SMARCA4 recruitment in the "indirect model" (Figures 1F and 2F). In this latter category of sites, SMARCA4 is potentially recruited via alternate TFs such as SP1,[9,75] whose motif is enriched at the indirect sites. CUX1 binding may be independent of or might follow BAF recruitment to these sites. It is not obvious why chromatin accessibility also decreases at these indirect sites after CUX1 knockout (Figure 2F). Perhaps CUX1 also promotes an open chromatin state by recruitment of histone acetyltransferases at these sites.[24,25] Alternatively, the partial nucleosome destabilization mediated by CUX1 alone enables other factors to bind and stabilize the more open chromatin state.

As in any adult tissue, hematopoietic differentiation requires stem and progenitor cells to undergo epigenetic reprogramming to commission and decommission the appropriate enhancers while reorienting genomic architecture to implement the pertinent mature cell transcriptional program. To date, the central actors in this process in the apex of the hematopoietic hierarchy

(B) IGV snapshot of CUX1 binding at the promoter of *KIT* and the reduced accessibility of multiple enhancers looped to the promoter. Enhancer and promotor annotations are from Roadmap Epigenomics.[44]
(C) Integration of $CD34^+$ HSPC ATAC-seq and RNA-seq ($n = 2$ biological replicates).[13] Scatterplot shows the RNA $\log_2$FC vs. ATAC-seq $\log_2$FC for 406 DEGs (FDR < 0.1, $|\log_2 FC| > 0.75$). Enriched GO terms related to HSPC lineage commitment are shown.[86]
(D) $\log_2$FC of ATAC-seq signal comparing $CD34^+$ HSPC gCUX1 vs. gHPRT cells at the hematopoietic cell-type-specific enhancers from the VISION database[87] (9,657 myeloid enhancers, 11,653 erythroid enhancers, and 15,323 lymphoid enhancers) and 10,000 randomly sampled non-cell-type-specific enhancers.
(E) Performance score of cell fate prediction using the published murine HSPC scRNA-seq.[68] From left to right: positive control is the top 2,000 genes with the highest cell-cell variation; negative controls are a randomly sampled gene set ($n = 1,000$) and curated list of mouse TFs ($n = 1,636$)[88]; CUX1-bound genes from human $CD34^+$ HSPC CUT&RUN ($n = 6,758$); overlap of CUX1-bound and differentially expressed upon CUX1 knockdown in $CD34^+$ HSPC ($n = 923$). Equivalent gene sets were tested for PU.1 and RUNX1 as benchmarks ($n = 336$ and 325).[69–72] For all gene sets larger than 1,000, 50 bootstraps were performed to sample for 1,000 genes. Logistic regression and deep neural network were used to construct the classifier. Error bars indicate 1X standard deviation of the F1 score. Significance for (A), (D), and (E) is by two-sided Wilcoxon rank-sum test.

have remained unclear. With respect to chromatin remodelers in normal hematopoiesis, our mechanistic knowledge of these factors, including the BAF complex, remains incomplete.[76] Regarding the TFs that direct these complexes, a few pioneer factors have been identified, but these have largely been described in cell lines or to act in downstream progenitors.[77]

Although not measured here, a logical extension of our finding is that CUX1 regulates chromatin accessibility in other tissue types. Given the wide-ranging role of CUX1 in the homeostasis of diverse tissues, it seems improbable that CUX1 only regulates a stereotypical set of target genes. CUX1 is conceivably a more general regulator of enhancer receptivity to activation via ensuing lineage-specific TFs. In this paradigm, CUX1 is critical for initiating epigenetic remodeling in tissue-specific stem cells, and lineage-determining TFs drive subsequent differentiation.

In myeloid malignancies, developmental syndromes, and other developmental contexts, CUX1 has haploinsufficient phenotypes.[14,78] Likewise, mutations in the BAF complex are commonly heterozygous in cancer and developmental disorders.[5] It remains to be determined how CUX1 haploinsufficiency impacts genome-wide BAF recruitment and DNA accessibility. One possibility is that loss of one copy of CUX1 untethers a portion of BAF to enable promiscuous BAF recruitment to *de novo* sites via other interacting partners.[79] We did not convincingly identify such a "gain-of-function" effect in K562 cells, where few *de novo* SMARCA4 binding sites are acquired after CUX1 knockout (Figure 1E). More likely, CUX1 or BAF complex haploinsufficiency leads to either partial or complete loss of regulation at a subset of target sites. The tools to precisely address this important question and characterize dose-dependent binding sites for future studies are only recently emerging.[80]

The canonical model of pioneer factor activity posits that after DNA accessibility is increased, "settler" TFs can subsequently bind DNA and execute gene expression. While our study did not test the subsequent recruitment of "settler" TFs by CUX1, we find the motifs and TF occupancy of several key regulators of hematopoietic differentiation uncovered at sites regulated by CUX1 and BAF. Ostensibly and counterintuitively, several of these TFs independently harbor pioneer factor activity, including RUNX1, PU.1, and KLF1.[77] There are several potential explanations for this apparent redundancy. First, it is conceivable that more than one pioneer TF binds simultaneously to an enhancer to cooperatively establish the enhancer landscape during differentiation.[81] Second, and not mutually exclusive, a given TF does not necessarily have pioneer activity at all DNA targets, as we observed for CUX1 and was described for PU.1, as two examples.[7] In other words, CUX1 may be required for PU.1 binding at a subset of enhancers. A third possibility is that these factors are binding sequentially, as opposed to simultaneously, during differentiation. In this case, CUX1 is required in HSPCs while a subsequent pioneer factor maintains accessibility in more mature progenitors. Thus, while the pioneer model provides a framework for conceptualizing epigenetic regulation, like many biological models, there is likely more underlying complexity. In fact, the binary concept of pioneer vs. settler TFs has been drawn into question, and more TFs may be uncovered within a spectrum of pioneer-like activity.[82]

Nonetheless, our findings indicate that a central role for CUX1 in hematopoiesis is the epigenetic regulation of lineage-specific enhancer accessibility. Haploinsufficiency of CUX1 disrupts normal HSPC homeostasis and differentiation, resulting in clonal expansion, lineage biases, and multilineage dysplasia.[13,16] When combined with additional mutations, CUX1 deficiency promotes fulminant leukemic transformation.[15,83] BAF inhibitors are an active area of translational exploration, including in myeloid neoplasms.[76] Our results suggest a potential synthetic-lethal rationale for therapeutic targeting of the residual BAF complex activity in this disease subset. Going forward, it will be important to determine if CUX1-deficient tumors are more or less responsive to BAF inhibition.

Limitations of the study

Our study did not delineate the subsequent CUX1 recruitment events nor the destinations of unbound BAF complexes post CUX1 loss. Future research should chart the sequential recruitment of other TFs by CUX1 and probe the fate and impact of the displaced BAF complexes. Additionally, early consequences of CUX1 depletion warrant exploration through techniques such as nascent RNA sequencing.

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
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell line and cell culture
- METHOD DETAILS
  - Co-immunoprecipitation
  - Sample preparation for LC–MS/MS
  - LC–MS/MS via MaxQuant
  - Mass spectrometry database searching and analysis
  - Ribonucleoprotein (RNP) transfection
  - ChIP-seq library preparation and sequencing
  - CUT&RUN library preparation and sequencing
  - ATAC-seq sample preparation and sequencing
  - ChIP-seq and CUT&RUN analysis
  - ATAC-seq analysis
  - Analysis of chromatin accessibility at cell type specific enhancers
  - Annotation of peaks with chromatin state
  - Hi-C analysis
  - Murine HSPC fate prediction
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.114227.

ACKNOWLEDGMENTS

The authors thank Tanner Martinez, Xin He, and Barbara Kee for critical feedback on the manuscript. The authors are grateful for the services and assistance provided by the following University of Chicago core facilities supported by the Cancer Center Support Grant (P30 CA014599): the DNA Sequencing

Facility and the Genomics Facility (RRID: SCR_019196). We also thank the University of Chicago's Center for Research Informatics Bioinformatics Core. Human CD34$^+$ cell acquisition was supported by the Cooperative Centers of Excellence in Hematology NIDDK Grant #U54 DK106829. We thank the ENCODE consortium and the ENCODE production laboratories, Dr. Peggy Farnham's lab at USC for K562 GATA1 ChIP-seq ENCODE:ENCSR227OMH, and Dr. Michael Snyder's lab at Stanford for RUNX1 ChIP-seq ENCODE:ENCSR406YIU. We downloaded these two datasets from the ENCODE portal (https://www.encodeproject.org/).[33,34] The graphical abstract was generated using BioRender (https://www.biorender.com/).

This work was supported in part by the American Cancer Society Research Scholar grant 132457-RSG-18-171-01-LIB, the Cancer Research Foundation Fletcher Scholars Award, and the National Institutes of Health (R01 HL142782, R01 CA231880, R01 HL166184, and P30 CA014599). M.E.M. is a Scholar of The Leukemia & Lymphoma Society. W.L. is supported by the Women's Board of the Cancer Center and Fitch Scholarship Fund. J.K. was supported by NIH F32 HL152524.

This manuscript is dedicated to the memory of Dr. Christopher (Casey) Brown.

AUTHOR CONTRIBUTIONS

J.L.K., W.L., S.J.K., and M.E.M. designed and interpreted the research. J.L.K., S.K., L.J., and D.A. performed most of the experiments. D.J.W. performed the co-immunoprecipitation mass spectrometry experiment and analysis. W.L., J.L.K., and A.S. carried out the data analysis. A.N.G. and W.L. carried out the Hi-C integration and machine learning classifier analysis. W.L., J.L.K., A.S., and M.E.M. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 27, 2023
Revised: March 16, 2024
Accepted: April 26, 2024

REFERENCES

1. Zakrzewski, W., Dobrzyński, M., Szymonowicz, M., and Rybak, Z. (2019). Stem cells: past, present, and future. Stem Cell Res. Ther. *10*, 68. https://doi.org/10.1186/s13287-019-1165-5.
2. Stadhouders, R., Filion, G.J., and Graf, T. (2019). Transcription factors and 3D genome conformation in cell-fate decisions. Nature *569*, 345–354. https://doi.org/10.1038/s41586-019-1182-7.
3. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.
4. Balsalobre, A., and Drouin, J. (2022). Pioneer factors as master regulators of the epigenome and cell fate. Nat. Rev. Mol. Cell Biol. *23*, 449–464. https://doi.org/10.1038/s41580-022-00464-z.
5. Mittal, P., and Roberts, C.W.M. (2020). The SWI/SNF complex in cancer — biology, biomarkers and therapy. Nat. Rev. Clin. Oncol. *17*, 435–448. https://doi.org/10.1038/s41571-020-0357-3.
6. Alver, B.H., Kim, K.H., Lu, P., Wang, X., Manchester, H.E., Wang, W., Haswell, J.R., Park, P.J., and Roberts, C.W.M. (2017). The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers. Nat. Commun. *8*, 14648. https://doi.org/10.1038/ncomms14648.
7. Chambers, C., Cermakova, K., Chan, Y.S., Kurtz, K., Wohlan, K., Lewis, A.H., Wang, C., Pham, A., Dejmek, M., Sala, M., et al. (2023). SWI/SNF Blockade Disrupts PU.1-Directed Enhancer Programs in Normal Hematopoietic Cells and Acute Myeloid Leukemia. Cancer Res. *83*, 983–996, OF1–OF14. https://doi.org/10.1158/0008-5472.CAN-22-2129.
8. Bakshi, R., Hassan, M.Q., Pratap, J., Lian, J.B., Montecino, M.A., van Wijnen, A.J., Stein, J.L., Imbalzano, A.N., and Stein, G.S. (2010). The human SWI/SNF complex associates with RUNX1 to control transcription of hematopoietic target genes. J. Cell. Physiol. *225*, 569–576. https://doi.org/10.1002/jcp.22240.
9. Kadam, S., McAlpine, G.S., Phelan, M.L., Kingston, R.E., Jones, K.A., and Emerson, B.M. (2000). Functional selectivity of recombinant mammalian SWI/SNF subunits. Genes Dev. *14*, 2441–2451.
10. Ellis, T., Gambardella, L., Horcher, M., Tschanz, S., Capol, J., Bertram, P., Jochum, W., Barrandon, Y., and Busslinger, M. (2001). The transcriptional repressor CDP (Cutl1) is essential for epithelial cell differentiation of the lung and the hair follicle. Genes Dev. *15*, 2307–2319. https://doi.org/10.1101/gad.200101.
11. Grueber, W.B., Jan, L.Y., and Jan, Y.N. (2003). Different levels of the homeodomain protein cut regulate distinct dendrite branching patterns of Drosophila multidendritic neurons. Cell *112*, 805–818. https://doi.org/10.1016/s0092-8674(03)00160-0.
12. Cubelos, B., Sebastián-Serrano, A., Beccari, L., Calcagnotto, M.E., Cisneros, E., Kim, S., Dopazo, A., Alvarez-Dolado, M., Redondo, J.M., Bovolenta, P., et al. (2010). Cux1 and Cux2 regulate dendritic branching, spine morphology and synapses of the upper layer neurons of the cortex. Neuron *66*, 523–535. https://doi.org/10.1016/j.neuron.2010.04.038.
13. An, N., Khan, S., Imgruet, M.K., Gurbuxani, S.K., Konecki, S.N., Burgess, M.R., and McNerney, M.E. (2018). Gene dosage effect of CUX1 in a murine model disrupts HSC homeostasis and controls the severity and mortality of MDS. Blood *131*, 2682–2697. https://doi.org/10.1182/blood-2017-10-810028.
14. Platzer, K., Cogné, B., Hague, J., Marcelis, C.L., Mitter, D., Oberndorff, K., Park, S.-M., Ploos van Amstel, H.K., Simonic, I., van der Smagt, J.J., et al. (2018). Haploinsufficiency of CUX1 Causes Nonsyndromic Global Developmental Delay With Possible Catch-up Development. Ann. Neurol. *84*, 200–207. https://doi.org/10.1002/ana.25278.
15. Wong, C.C., Martincorena, I., Rust, A.G., Rashid, M., Alifrangis, C., Alexandrov, L.B., Tiffen, J.C., Kober, C., Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium; and Green, A.R., et al. (2014). Inactivating CUX1 mutations promote tumorigenesis. Nat. Genet. *46*, 33–38. https://doi.org/10.1038/ng.2846.
16. Jotte, M.R.M., and McNerney, M.E. (2022). The significance of CUX1 and chromosome 7 in myeloid malignancies. Curr. Opin. Hematol. *29*, 92–102. https://doi.org/10.1097/MOH.0000000000000699.
17. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P.V., Mar, B.G., Lindsley, R.C., Mermel, C.H., Burtt, N., Chavez, A., et al. (2014). Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. N. Engl. J. Med. *371*, 2488–2498. https://doi.org/10.1056/NEJMoa1408617.
18. Zink, F., Stacey, S.N., Norddahl, G.L., Frigge, M.L., Magnusson, O.T., Jonsdottir, I., Thorgeirsson, T.E., Sigurdsson, A., Gudjonsson, S.A., Gudmundsson, J., et al. (2017). Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood *130*, 742–752. https://doi.org/10.1182/blood-2017-02-769869.
19. Whalen, S., Truty, R.M., and Pollard, K.S. (2016). Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. Nat. Genet. *48*, 488–496. https://doi.org/10.1038/ng.3539.
20. Arthur, R.K., An, N., Khan, S., and McNerney, M.E. (2017). The haploinsufficient tumor suppressor, CUX1, acts as an analog transcriptional regulator that controls target genes through distal enhancers that loop to target promoters. Nucleic Acids Res. *45*, 6350–6361. https://doi.org/10.1093/nar/gkx218.
21. Li, S., Moy, L., Pittman, N., Shue, G., Aufiero, B., Neufeld, E.J., LeLeiko, N.S., and Walsh, M.J. (1999). Transcriptional repression of the cystic fibrosis transmembrane conductance regulator gene, mediated by CCAAT displacement protein/cut homolog, is associated with histone

deacetylation. J. Biol. Chem. *274*, 7803–7815. https://doi.org/10.1074/jbc.274.12.7803.

22. Coqueret, O., Bérubé, G., and Nepveu, A. (1998). The mammalian Cut homeodomain protein functions as a cell-cycle-dependent transcriptional repressor which downmodulates p21WAF1/CIP1/SDI1 in S phase. EMBO J. *17*, 4680–4694. https://doi.org/10.1093/emboj/17.16.4680.
23. Livingston, S., Carlton, C., Sharma, M., Kearns, D., Baybutt, R., and Vanden Heuvel, G.B. (2019). Cux1 regulation of the cyclin kinase inhibitor p27kip1 in polycystic kidney disease is attenuated by HDAC inhibitors. Gene. X *2*, 100007. https://doi.org/10.1016/j.gene.2019.100007.
24. Ueda, Y., Su, Y., and Richmond, A. (2007). CCAAT displacement protein regulates nuclear factor-kappa beta-mediated chemokine transcription in melanoma cells. Melanoma Res. *17*, 91–103. https://doi.org/10.1097/CMR.0b013e3280a60888.
25. Truscott, M., Harada, R., Vadnais, C., Robert, F., and Nepveu, A. (2008). p110 CUX1 Cooperates with E2F Transcription Factors in the Transcriptional Activation of Cell Cycle-Regulated Genes. Mol. Cell Biol. *28*, 3127–3138. https://doi.org/10.1128/MCB.02089-07.
26. Andersson, L.C., Jokinen, M., and Gahmberg, C.G. (1979). Induction of erythroid differentiation in the human leukaemia cell line K562. Nature *278*, 364–365. https://doi.org/10.1038/278364a0.
27. Tabilio, A., Pelicci, P.G., Vinci, G., Mannoni, P., Civin, C.I., Vainchenker, W., Testa, U., Lipinski, M., Rochant, H., and Breton-Gorius, J. (1983). Myeloid and megakaryocytic properties of K-562 cell lines. Cancer Res. *43*, 4569–4574.
28. Green, A.R., Rockman, S., DeLuca, E., and Begley, C.G. (1993). Induced myeloid differentiation of K562 cells with downregulation of erythroid and megakaryocytic transcription factors: a novel experimental model for hemopoietic lineage restriction. Exp. Hematol. *21*, 525–531.
29. Centore, R.C., Sandoval, G.J., Soares, L.M.M., Kadoch, C., and Chan, H.M. (2020). Mammalian SWI/SNF Chromatin Remodeling Complexes: Emerging Mechanisms and Therapeutic Strategies. Trends Genet. *36*, 936–950. https://doi.org/10.1016/j.tig.2020.07.011.
30. Mardinian, K., Adashek, J.J., Botta, G.P., Kato, S., and Kurzrock, R. (2021). SMARCA4: Implications of an altered chromatin-remodeling gene for cancer development and therapy. Mol. Cancer Ther. *20*, 2341–2351. https://doi.org/10.1158/1535-7163.MCT-21–0433.
31. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.
32. Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. *5*, 1752–1779. https://doi.org/10.1214/11-AOAS466.
33. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.
34. Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., et al. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. *48*, D882–D889. https://doi.org/10.1093/nar/gkz1062.
35. Imgruet, M.K., Lutze, J., An, N., Hu, B., Khan, S., Kurkewich, J., Martinez, T.C., Wolfgeher, D., Gurbuxani, S.K., Kron, S.J., and McNerney, M.E. (2021). Loss of a 7q gene, CUX1, disrupts epigenetically driven DNA repair and drives therapy-related myeloid neoplasms. Blood *138*, 790–805. https://doi.org/10.1182/blood.2020009195.
36. Elder, G.H. (1998). Genetic Defects in the Porphyrias: Types and Significance. Clin. Dermatol. *16*, 225–233. https://doi.org/10.1016/S0738-081X(97)00202-2.
37. Miller, I.J., and Bieker, J.J. (1993). A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Krüppel family of nuclear proteins. Mol. Cell Biol. *13*, 2776–2786. https://doi.org/10.1128/mcb.13.5.2776-2786.1993.
38. Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R., and Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. Nature *375*, 316–318. https://doi.org/10.1038/375316a0.
39. Lyu, Z.-Z., Zhao, B.-B., Koiwai, K., Hirono, I., and Kondo, H. (2016). Identification of endonuclease domain-containing 1 gene in Japanese flounder Paralichthys olivaceus. Fish Shellfish Immunol. *50*, 43–49. https://doi.org/10.1016/j.fsi.2016.01.017.
40. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218. https://doi.org/10.1038/nmeth.2688.
41. Lun, A.T.L., and Smyth, G.K. (2016). csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Res. *44*, e45. https://doi.org/10.1093/nar/gkv1191.
42. Moon, N.S., Bérubé, G., and Nepveu, A. (2000). CCAAT displacement activity involves CUT repeats 1 and 2, not the CUT homeodomain. J. Biol. Chem. *275*, 31325–31334. https://doi.org/10.1074/jbc.M002912200.
43. 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.
44. Roadmap Epigenomics Consortium; Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330. https://doi.org/10.1038/nature14248.
45. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501. https://doi.org/10.1038/nbt.1630.
46. Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., and Lewis, S.AmiGO Hub; Web Presence Working Group (2009). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288–289. https://doi.org/10.1093/bioinformatics/btn615.
47. Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. *49*, D325–D334. https://doi.org/10.1093/nar/gkaa1113.
48. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29. https://doi.org/10.1038/75556.
49. Nottingham, W.T., Jarratt, A., Burgess, M., Speck, C.L., Cheng, J.-F., Prabhakar, S., Rubin, E.M., Li, P.-S., Sloane-Stanley, J., Kong-a-San, J., and de Bruijn, M.F.T.R. (2007). Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. Blood *110*, 4188–4197. https://doi.org/10.1182/blood-2007-07-100883.
50. Bee, T., Ashley, E.L.K., Bickley, S.R.B., Jarratt, A., Li, P.-S., Sloane-Stanley, J., Göttgens, B., and de Bruijn, M.F.T.R. (2009). The mouse Runx1 +23 hematopoietic stem cell enhancer confers hematopoietic specificity to both Runx1 promoters. Blood *113*, 5121–5124. https://doi.org/10.1182/blood-2008-12-193003.
51. Lee, D., Shi, M., Moran, J., Wall, M., Zhang, J., Liu, J., Fitzgerald, D., Kyono, Y., Ma, L., White, K.P., and Gerstein, M. (2020). STARRPeaker: uniform processing and accurate identification of STARR-seq active regions. Genome Biol. *21*, 298. https://doi.org/10.1186/s13059-020-02194-x.
52. Sato, S., Tomomori-Sato, C., Banks, C.A.S., Sorokina, I., Parmely, T.J., Kong, S.E., Jin, J., Cai, Y., Lane, W.S., Brower, C.S., et al. (2003). Identification of mammalian Mediator subunits with similarities to yeast Mediator subunits Srb5, Srb6, Med11, and Rox3. J. Biol. Chem. *278*, 15123–15127. https://doi.org/10.1074/jbc.C300054200.

53. Ferreira, L.T., Figueiredo, A.C., Orr, B., Lopes, D., and Maiato, H. (2018). Dissecting the role of the tubulin code in mitosis. Methods Cell Biol. *144*, 33–74. https://doi.org/10.1016/bs.mcb.2018.03.040.

54. Wahlestedt, M., Ladopoulos, V., Hidalgo, I., Sanchez Castillo, M., Hannah, R., Säwén, P., Wan, H., Dudenhöffer-Pfeifer, M., Magnusson, M., Norddahl, G.L., et al. (2017). Critical Modulation of Hematopoietic Lineage Fate by Hepatic Leukemia Factor. Cell Rep. *21*, 2251–2263. https://doi.org/10.1016/j.celrep.2017.10.112.

55. Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell *132*, 631–644. https://doi.org/10.1016/j.cell.2008.01.025.

56. Mancias, J.D., Pontano Vaites, L., Nissim, S., Biancur, D.E., Kim, A.J., Wang, X., Liu, Y., Goessling, W., Kimmelman, A.C., and Harper, J.W. (2015). Ferritinophagy via NCOA4 is required for erythropoiesis and is regulated by iron dependent HERC2-mediated proteolysis. Elife *4*, e10308. https://doi.org/10.7554/eLife.10308.

57. Kinkel, S.A., Galeev, R., Flensburg, C., Keniry, A., Breslin, K., Gilan, O., Lee, S., Liu, J., Chen, K., Gearing, L.J., et al. (2015). Jarid2 regulates hematopoietic stem cell function by acting with polycomb repressive complex 2. Blood *125*, 1890–1900. https://doi.org/10.1182/blood-2014-10-603969.

58. Celik, H., Koh, W.K., Kramer, A.C., Ostrander, E.L., Mallaney, C., Fisher, D.A.C., Xiang, J., Wilson, W.C., Martens, A., Kothari, A., et al. (2018). JARID2 Functions as a Tumor Suppressor in Myeloid Neoplasms by Repressing Self-Renewal in Hematopoietic Progenitor Cells. Cancer Cell *34*, 741–756.e8. https://doi.org/10.1016/j.ccell.2018.10.008.

59. Zhang, X., Jeong, M., Huang, X., Wang, X.Q., Wang, X., Zhou, W., Shamim, M.S., Gore, H., Himadewi, P., Liu, Y., et al. (2020). Large DNA Methylation Nadirs Anchor Chromatin Loops Maintaining Hematopoietic Stem Cell Identity. Mol. Cell *78*, 506–521.e6. https://doi.org/10.1016/j.molcel.2020.04.018.

60. Thorén, L.A., Liuba, K., Bryder, D., Nygren, J.M., Jensen, C.T., Qian, H., Antonchuk, J., and Jacobsen, S.-E.W. (2008). Kit regulates maintenance of quiescent hematopoietic stem cells. J. Immunol. Baltim. Md *180*, 2045–2053, 1950. https://doi.org/10.4049/jimmunol.180.4.2045.

61. Unnisa, Z., Clark, J.P., Roychoudhury, J., Thomas, E., Tessarollo, L., Copeland, N.G., Jenkins, N.A., Grimes, H.L., and Kumar, A.R. (2012). Meis1 preserves hematopoietic stem cells in mice by limiting oxidative stress. Blood *120*, 4973–4981. https://doi.org/10.1182/blood-2012-06-435800.

62. Azcoitia, V., Aracil, M., Martínez-A, C., and Torres, M. (2005). The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. Dev. Biol. *280*, 307–320. https://doi.org/10.1016/j.ydbio.2005.01.004.

63. Vignudelli, T., Selmi, T., Martello, A., Parenti, S., Grande, A., Gemelli, C., Zanocco-Marani, T., and Ferrari, S. (2010). ZFP36L1 negatively regulates erythroid differentiation of CD34+ hematopoietic stem cells by interfering with the Stat5b pathway. Mol. Biol. Cell *21*, 3340–3351. https://doi.org/10.1091/mbc.E10-01-0040.

64. Fagnan, A., Drissen, R., Di Genua, C., Meng, Y., and Nerlov, C. (2023). Myelo-Erythroid Lineage Segregation Is Regulated By the GATA-2 Interactome. Blood *142*, 2687. https://doi.org/10.1182/blood-2023-190285.

65. Zaret, K.S., and Mango, S.E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr. Opin. Genet. Dev. *37*, 76–81. https://doi.org/10.1016/j.gde.2015.12.003.

66. Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. *16*, 144–154. https://doi.org/10.1038/nrm3949.

67. Zhang, Y., and Hardison, R.C. (2017). Accurate and reproducible functional maps in 127 human cell types via 2D genome segmentation. Nucleic Acids Res. *45*, 9823–9836. https://doi.org/10.1093/nar/gkx659.

68. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F.D., and Klein, A.M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science *367*, eaaw3381. https://doi.org/10.1126/science.aaw3381.

69. Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell *144*, 296–309. https://doi.org/10.1016/j.cell.2011.01.004.

70. Kim, K.M., Mura-Meszaros, A., Tollot, M., Krishnan, M.S., Gründl, M., Neubert, L., Groth, M., Rodriguez-Fraticelli, A., Svendsen, A.F., Campaner, S., et al. (2022). Taz protects hematopoietic stem cells from an aging-dependent decrease in PU.1 activity. Nat. Commun. *13*, 5187. https://doi.org/10.1038/s41467-022-32970-1.

71. Cai, X., Gao, L., Teng, L., Ge, J., Oo, Z.M., Kumar, A.R., Gilliland, D.G., Mason, P.J., Tan, K., and Speck, N.A. (2015). Runx1 Deficiency Decreases Ribosome Biogenesis and Confers Stress Resistance to Hematopoietic Stem and Progenitor Cells. Cell Stem Cell *17*, 165–177. https://doi.org/10.1016/j.stem.2015.06.002.

72. Wu, J.Q., Seay, M., Schulz, V.P., Hariharan, M., Tuck, D., Lian, J., Du, J., Shi, M., Ye, Z., Gerstein, M., et al. (2012). Tcf7 Is an Important Regulator of the Switch of Self-Renewal and Differentiation in a Multipotential Hematopoietic Cell Line. PLoS Genet. *8*, e1002565. https://doi.org/10.1371/journal.pgen.1002565.

73. Last, T.J., van Wijnen, A.J., de Ridder, M.C., Stein, G.S., and Stein, J.L. (1999). The homeodomain transcription factor CDP/cut interacts with the cell cycle regulatory element of histone H4 genes packaged into nucleosomes. Mol. Biol. Rep. *26*, 185–194. https://doi.org/10.1023/a:1007058123699.

74. Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M., et al. (2018). The interaction landscape between transcription factors and the nucleosome. Nature *562*, 76–81. https://doi.org/10.1038/s41586-018-0549-5.

75. Swift, M.L., Beishline, K., and Azizkhan-Clifford, J. (2021). Sp1-dependent recruitment of the histone acetylase p300 to DSBs facilitates chromatin remodeling and recruitment of the NHEJ repair factor Ku70. DNA Repair *105*, 103171. https://doi.org/10.1016/j.dnarep.2021.103171.

76. Andrades, A., Peinado, P., Alvarez-Perez, J.C., Sanjuan-Hidalgo, J., García, D.J., Arenas, A.M., Matia-González, A.M., and Medina, P.P. (2023). SWI/SNF complexes in hematological malignancies: biological implications and therapeutic opportunities. Mol. Cancer *22*, 39. https://doi.org/10.1186/s12943-023-01736-8.

77. Wang, Z., Wang, P., Li, Y., Peng, H., Zhu, Y., Mohandas, N., and Liu, J. (2021). Interplay between cofactors and transcription factors in hematopoiesis and hematological malignancies. Signal Transduct. Target. Ther. *6*, 24. https://doi.org/10.1038/s41392-020-00422-1.

78. Ramdzan, Z.M., and Nepveu, A. (2014). CUX1, a haploinsufficient tumour suppressor gene overexpressed in advanced cancers. Nat. Rev. Cancer *14*, 673–682. https://doi.org/10.1038/nrc3805.

79. McBride, M.J., Pulice, J.L., Beird, H.C., Ingram, D.R., D'Avino, A.R., Shern, J.F., Charville, G.W., Hornick, J.L., Nakayama, R.T., Garcia-Rivera, E.M., et al. (2018). The SS18-SSX Fusion Oncoprotein Hijacks BAF Complex Targeting and Function to Drive Synovial Sarcoma. Cancer Cell *33*, 1128–1141.e7. https://doi.org/10.1016/j.ccell.2018.05.002.

80. Naqvi, S., Kim, S., Hoskens, H., Matthews, H.S., Spritz, R.A., Klein, O.D., Hallgrímsson, B., Swigut, T., Claes, P., Pritchard, J.K., and Wysocka, J. (2023). Precise modulation of transcription factor levels identifies features underlying dosage sensitivity. Nat. Genet. *55*, 841–851. https://doi.org/10.1038/s41588-023-01366-2.

81. Arnosti, D.N., and Kulkarni, M.M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? J. Cell. Biochem. *94*, 890–898. https://doi.org/10.1002/jcb.20352.

82. Hansen, J.L., Loell, K.J., and Cohen, B.A. (2022). A test of the pioneer factor hypothesis using ectopic liver gene activation. Elife *11*, e73358. https://doi.org/10.7554/eLife.73358.

83. An, N., Khan, S., Imgruet, M.K., Jueng, L., Gurbuxani, S., and McNerney, M.E. (2023). Oncogenic RAS promotes leukemic transformation of CUX1-deficient cells. Oncogene *42*, 881–893. https://doi.org/10.1038/s41388-023-02612-x.

84. 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.

85. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

86. Ng, S.Y.-M., Yoshida, T., Zhang, J., and Georgopoulos, K. (2009). Genome-wide lineage-specific transcriptional networks underscore Ikaros-dependent lymphoid priming in hematopoietic stem cells. Immunity *30*, 493–507. https://doi.org/10.1016/j.immuni.2009.01.014.

87. Xiang, G., Keller, C.A., Heuston, E., Giardine, B.M., An, L., Wixom, A.Q., Miller, A., Cockburn, A., Sauria, M.E.G., Weaver, K., et al. (2020). An integrative view of the regulatory and transcriptional landscapes in mouse hematopoiesis. Genome Res. *30*, 472–484. https://doi.org/10.1101/gr.255760.119.

88. Hu, H., Miao, Y.-R., Jia, L.-H., Yu, Q.-Y., Zhang, Q., and Guo, A.-Y. (2019). AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res. *47*, D33–D38. https://doi.org/10.1093/nar/gky822.

89. Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210. https://doi.org/10.1093/nar/30.1.207.

90. Wolfgeher, D., Dunn, D.M., Woodford, M.R., Bourboulia, D., Bratslavsky, G., Mollapour, M., Kron, S.J., and Truman, A.W. (2015). The dynamic interactome of human Aha1 upon Y223 phosphorylation. Data Brief *5*, 752–755. https://doi.org/10.1016/j.dib.2015.10.028.

91. Truman, A.W., Kristjansdottir, K., Wolfgeher, D., Ricco, N., Mayampurath, A., Volchenboum, S.L., Clotet, J., and Kron, S.J. (2015). The quantitative changes in the yeast Hsp70 and Hsp90 interactomes upon DNA damage. Data Brief *2*, 12–15. https://doi.org/10.1016/j.dib.2014.10.006.

92. Baeten, J.T., Liu, W., Preddy, I.C., Zhou, N., and McNerney, M.E. (2022). CRISPR screening in human hematopoietic stem and progenitor cells reveals an enrichment for tumor suppressor genes within chromosome 7 commonly deleted regions. Leukemia *36*, 1421–1425. https://doi.org/10.1038/s41375-021-01491-z.

93. Brunetti, L., Gundry, M.C., Kitano, A., Nakada, D., and Goodell, M.A. (2018). Highly Efficient Gene Disruption of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. J. Vis. Exp. *134*, 57278. https://doi.org/10.3791/57278.

94. Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. *42*, e168. https://doi.org/10.1093/nar/gku936.

95. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

96. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

97. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci. Rep. *9*, 9354. https://doi.org/10.1038/s41598-019-45839-z.

98. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191. https://doi.org/10.1093/nar/gku365.

99. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

100. 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.

101. Tl, B., J, J., Ce, G., and Ws, N. (2015). The MEME Suite. Nucleic Acids Res. *43*. https://doi.org/10.1093/nar/gkv416.

102. McLeay, R.C., and Bailey, T.L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinf. *11*, 165. https://doi.org/10.1186/1471-2105-11-165.

103. Ernst, J., and Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat. Biotechnol. *28*, 817–825. https://doi.org/10.1038/nbt.1662.

104. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49. https://doi.org/10.1038/nature09906.

105. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

106. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

107. Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191. https://doi.org/10.1038/nprot.2009.97.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CUX1	rabbit	PUC, Poconos
SMARCA4	rabbit	Abcam Cat# ab110641, RRID:AB_10861578
SMARCA4	rabbit	Cell Signaling Technology Cat# 49360, RRID:AB_2728743
GATA1	rabbit	Abcam Cat# ab181544, RRID:AB_2920794
H3K27ac	rabbit	Abcam Cat# ab4729, RRID:AB_2118291
H3K4me1	rabbit	Abcam Cat# ab8895, RRID:AB_306847
Critical commercial assays		
Ovation Ultralow Library Kit	Tecan Genomics	Cat0344NB-32
MinElute PCR Purification Kit	Qiagen	Cat#28004
Deposited data		
Co-IP Mass Spec CUX1 Interaction Proteins(K562)	This paper	PRIDE: PXD037838
ChIP-sequencing (K562)	This paper	GEO: GSE235309
CUT&RUN (K562 and CD34$^+$ HSPC)	This paper	GEO: GSE235303
ATAC-seq (K562 and CD34$^+$ HSPC)	This paper	GEO: GSE235299
HSPC Hi-C interaction loops	Zhang et al. 2020[59]	N/A
Experimental models: Cell lines		
K562	Dr. Michelle Le Beau's lab, the University of Chicago	N/A
Primary human CD34$^+$ HSPCs	Fred Hutch Hematopoietic Cell Procurement and Resource Development Center	N/A
Oligonucleotides		
gRNA targeting intron 2 of *HPRT*	Imgruet et al. 2021[35]	N/A
gRNA targeting exon 4 of *CUX1*	Imgruet et al. 2021[35]	N/A
gRNA targeting exon 6 of *CUX1*	This manuscript	N/A
gRNA targeting exon 3 of *SMARCA4*	This manuscript	N/A
gRNA targeting exon 4 of *SMARCA4*	This manuscript	N/A
Software and algorithms		
MaxQuant	Max-Planck Institute of Biochemistry	https://www.maxquant.org/
R studio v2023.09.0 + 463	Posit	https://posit.co/download/rstudio-desktop/
IGV version 2.8.10	Broad Institute	https://igv.org/
GREAT version 4.0.4	Standford University	http://great.stanford.edu/public/html/
MEME-suite version 5.5.5	N/A	https://meme-suite.org/meme/
Customized code for machine learning cell fate prediction (Figure 4E)	McNerney Lab, the Univeristy of Chicago	https://github.com/AlexandreGaubil/mcnerney-cux1-ML
Customized code for Hi-C integration analysis (Figure 4A)	McNerney Lab, the Univeristy of Chicago	https://github.com/liuweihanty/CD34_HiC_CUX1_integration

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Megan McNerney (megan.mcnerney@bsd.uchicago.edu)

Materials availability

The study did not generate new mouse lines or unique reagents.

Data and code availability

- The mass spectrometry proteomic datasets (MMSK1, MMSK2) were uploaded to the ProteomeXchange consortium via the PRIDE partner repository with the dataset identifier PRIDE:PXD037838. ChIP-seq, CUT&RUN and ATAC-seq data are available at NCBI's Gene Expression Omnibus[89] and are accessible through GEO Series accession number GEO:GSE235309. The data will be publically accessible by the paper's publishment date.
- The code for machine learning cell fate prediction (Figure 4E) can be accessed on Github (https://github.com/AlexandreGaubil/mcnerney-cux1-ML). The code used for Hi-C integration analysis (Figure 4A) can be accessed on Github (https://github.com/liuweihanty/CD34_HiC_CUX1_integration)
- 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 line and cell culture

K562 cell lines were obtained from Dr. Michelle Le Beau's lab (University of Chicago) and were authenticated by STR analysis (ATCC). Primary human $CD34^+$ peripheral blood mononuclear-stem cells were obtained from the Fred Hutch Hematopoietic Cell Procurement and Resource Development Center (Seattle, WA). K562 cells were grown in RPMI 1640 media (Gibco 61870127) supplemented with 10% FBS and 1X Antibiotic-Antimycotic (Gibco 15240062). Primary $CD34^+$ cells were obtained from two independent healthy donors and grown in StemSpan SFEMII media (STEMCELL Technologies 09655) supplemented with 1X StemSpan CC110 cytokine cocktail (Stemcell Technologies 02697).

METHOD DETAILS

Co-immunoprecipitation

100×10^6 K562 cells were spun down for a CUX1 pulldown and a control IgG pulldown each. Cells were lysed in hypotonic buffer (5 mM EDTA, 5 mM EGTA, 5 mM Tris–Cl) with protease inhibitor added (Roche complete mini-EDTA free 11836170001). Pellets were passed through a 20-gauge needle 10 times, incubated on ice for 10 min and spun down at 600 g for 8 min at 4°C. The supernatant was removed, and the pellet was resuspended in RIPA buffer (Boston BioProducts BP115) with protease inhibitor added (Roche Complete 5892953001). Protein lysates were again passed through a 27-gauge needle, incubated on ice and subsequently spun down at 14000 rpm for 15 min at 4°C. The supernatant was collected, and RIPA buffer was added to a final volume of 30 mL. 12 μg of CUX1 antibody (B-10 Santa Cruz sc-514008) and mouse IgG (Santa Cruz sc-2025) antibody were added to the lysate and incubated overnight on a rocker at 4°C. 150 μL Protein A/G Plus agarose beads (Santa Cruz sc-2003) were added the next day and incubated at 4°C on a rocker for 1 h. The immunoprecipitated proteins were washed twice with cold RIPA buffer followed by a final wash with cold PBS. Proteins were eluted by resuspending the beads in 2X loading buffer and sent for mass spec analysis.

Sample preparation for LC–MS/MS

Co-immunoprecipitate samples were brought to 1X and 40 μL was loaded onto 12% MOPS buffered 1D SDS-PAGE gel (Invitrogen NP0341BOX) and run at ~ 200 V for ~ 10 min, resulting in a ~2 cm gel plug. The gel was stained with Imperial Stain (Thermo Fisher #24615) for 1 h at room temperature. Gel plug trypsin digestion was adapted from methods previously published[90,91]. Specifically, Gel sections were washed in dH 2 O and destained using 100 mM NH 4 HCO3 (Sigma #285099) pH7.5 in 50% acetonitrile (Fisher A998SK-4). A reduction step was performed by addition of 100μL 50 mM NH4HCO3 pH 7.5 and 10 μL of 200 mM tris(2-carboxyethyl) phosphine HCl (Sigma#C4706-2G) at 37°C for 30 min. The proteins were alkylated by addition of 100 μL of 50 mM iodoacetamide (Sigma #RPN6320V) prepared fresh in 50 mM NH4HCO3 pH 7.5 buffer and allowed to react in the dark at 20°C for 30 min. Gel sections were washed in Millipore water, then acetonitrile, and vacuum dried. Trypsin digestion was carried out overnight at 37°C with 1:50–1:100 enzyme–protein ratio of sequencing grade-modified trypsin (Promega #V5111) in 50 mM NH4HCO3 pH 7.5, and 20 mM CaCl2 (Sigma #C-1016). Peptides were extracted with 5%formic acid (Sigma #F0507-1L) in aqueous and 75% organic (ACN) combined and vacuum dried. Peptides were cleaned up using C18 spin columns (Thermo #89870).

LC–MS/MS via MaxQuant

LC–MS/MS was performed using adapted methods previously published.[90] Electrospray tandem mass spectrometry (LC–MS/MS) was performed at the Mayo Clinic Proteomics Core on a Thermo Q-Exactive Orbitrap mass spectrometer, using a 70,000 RP (70 K Resolving Power at 400 Da) survey scan in profile mode, m/z 340–1800 Da, with lockmasses, followed by 20 MS/MS HCD fragmentation scans at 17,500 resolutions on doubly and triply charged precursors. Single charged ions were excluded, and ions selected for MS/MS were placed on an exclusion list for 60 s.

Mass spectrometry database searching and analysis

Tandem mass spectra MS/MS samples were analyzed using MaxQuant (version 1.6.17.0). MaxQuant was set up to search the 211102_Uniprot_Human_5640.fasta database assuming the digestion enzyme strict trypsin. MaxQuant was searched with a fragment ion mass tolerance, and a parent ion tolerance of 20 PPM. MQ 1FDR results file (proteingroups.txt) was processed in Perseus (version 1.6.14.0). Proteins were filtered out which included "identified by site", "reversed", and "potential contaminants", $\log_2$ transformed, imputed via default settings, and annotated against the human database. P-values were determined by Student's t-test within Perseus and a significance cutoff was applied if CUX11/IgG ratios were above NegLog10 P-value $\geq$ 1.3 and fold-change above 20% or log2 $\geq$ 0.26. Proteins only detected in CUX1 immunoprecipitates were also determined significant.

Ribonucleoprotein (RNP) transfection

gHPRT and gCUX1 K562 cell lines were described previously.[35] We generated an additional single cell clone of K562 with new gRNA targeting exon 6 of *CUX1*. The gRNA sequence for exon 6 is 5′-CUGUUCCUUCUCAAGAGCUA-3'. For gSMARCA4 in K562, two gRNAs targeting exon 3 and exon 4 of the *SMARCA4* gene were designed with sequences 5′-AUGGAGUCCAUGCAUGAGAA-3′ and 5′-GGUCCUGUUGCGGACACCGA-3'. The editing efficiencies are: gCUX1 exon6 96%, gSMARCA4 exon3 99%, gSMARCA4 exon4 93%. For primary CD34$^+$ HSPCs, cells were transfected with ribonucleoprotein complexes carrying the same gRNA sequences as used in K562 for exon 4 of CUX1 (5′-UGCACUGAGUAAAAGAAGCA-3′)[92] or intron 2 of HPRT (5′-GCAUUUCUCAGUC CUAAACA-3′) (Synthego) using the Neon transfection device (Thermo Fisher) with the following parameters: 1600V, 10ms, 3 pulses.[93] Editing efficiency was determined 72 h post-transfection using TIDE (https://tide.nki.nl/).[94] The editing efficiencies for the transfected cell population replicates used for experiments are: replicate one 47%, replicate two 52%; replicate one 79%, replicate two 72%.

ChIP-seq library preparation and sequencing

Chromatin was fixed from 100x10^6 gHPRT and gCUX1 transfected K562 cells using 1% formaldehyde for 10 min at room temperature and stopped by the addition of 0.125 M glycine. For SMARCA4 ChIP, protein cross linking was performed first. Cells were washed 3 times with 1X PBS at room temperature. 10 mL of PBS/MgCl2 were added to the cells after final PBS wash. 80 μL of 0.25M DSG-disuccinimidyl glutarate (Thermo Fischer 20593) was added and incubated for 45 min at room temperature. Cells were washed 3 times with 1X PBS and followed by DNA crosslinking with 1% formaldehyde as described above. Fixed chromatin was then sonicated (Bioruptor) for 10 min in 30 s on/off pulses two times for a total of 20 min, with vortexing in between. CUX1-specific antibodies were generated, characterized and validated as described by Imgruet et al. 2022.[35] Immunoprecipitation was performed using dynabead protein G magnetic beads (Thermo Fischer) and 6 μg of anti-CUX1 (PUC, Poconos)/20E6 cells, 5 μg/20E6 anti-SMARCA4 (Abcam ab110641), anti-H3K27ac (Abcam, ab4729) or anti-H3K4Me1 (Abcam, ab8895). Following elution, samples were treated with RNase A and proteinase K before crosslink reversal. DNA was purified using a PCR purification kit (Qiagen). Libraries were prepped using the Ovation Ultralow Library Kit Tecan Genomics Inc (0344NB-32) and size selected using SPRIselect beads (Beckman Coulter B23317). Illumina HiSeq was used to perform 50 bp single-end sequencing on the libraries. Two biological replicates were performed for each sample.

CUT&RUN library preparation and sequencing

CUT&RUN was performed as described by Skene and Henikoff 2017[43] using the direct ligation method for mammalian cells. Briefly, 5x10^5 cells were harvested from CD34$^+$ HSPCs expanded for 48 h post-thawing and bound to ConA-coating beads by rotation for 10 min at room temperature. Cells were permeabilized (20 mM HEPES pH 7.5, 150 mM NaCL, 0.5 mM Sperimidine, Roche Complete EDTA free 5892953001, 0.05% w/v digitonin) and incubated overnight at 4°C with anti-CUX1 (PUC) 1:50, anti-SMARCA4 (Cell Signaling, 49360s) 1:100, or anti-GATA1 (Abcam, ab181544) 1:100 antibodies. Protein A/G-MNase beads were added and placed on a tube rotator for 1 h at 4°C. MNAse bound DNA was cleaved and released by adding 1X pA-MNAse mix containing CaCl2 at 0°C for 30 min, STOP buffer was added and CUT&RUN fragments were released by incubating for 30 min at 37°C. Library end repair, ligation, and amplification were performed using the Ovation Ultralow System V2 kit (Tecan Genomics Inc. 0344NB-32) and amplified by PCR with the following parameters: 1 cycle of 72°C 2 min, 95°C 3 min, followed by 13 cycles of 98°C 20 s, 65°C 30 s, 72°C 30 s, and a final extension at 72°C for 1 min. Libraries were cleaned up using MinElute PCR purification kit (Qiagen) and a left-sided size selection using SPRI beads (Beckman Coulter B23317). Final libraries were analyzed by Bioanalyzer (Agilent) prior to sequencing.

ATAC-seq sample preparation and sequencing

ATAC-sequencing was performed according to a published protocol.[40] For all experiments, K562 cells were harvested from cultures at ~60% confluency and primary CD34$^+$ HSPCs were harvested 48 h post transfection. For both K562 and primary CD34$^+$ HSPCs, 50,000 cells were lysed using the following buffer: 10 mM Tris-HCL, pH 7.4, 10 mM NaCL, 3 mM $MgCl_2$, and 0.1% IGEPAL CA-630. Cells were transposed using a 1X concentration of Nextera Tn5 Transposase (Illumina) for 30 min at 37°C with shaking at 500 rpm. Following transposition, DNA was purified using the MinElute PCR Purification Kit (Qiagen). DNA was amplified for 5 initial cycles using the custom Nextera barcoded PCR primers with the following parameters: 1 cycle of 72°C for 5 min and 98°C for 30 s, followed by 5 cycles of 98°C for 10 s, 63°C for 30 s, and 72°C for 1 min. Following the initial 5 cycles of PCR, the additional number of cycles needed was determined by qPCR as previously described.[40] Specifically, qPCR was performed to determine the additional number

of pcr cycles required. The total number of PCR cycles ranged between 7 and 12 for all samples. Following the additional PCR cycles, DNA was obtained using the MinElute PCR Purification Kit and analyzed by Bioanalyzer (Agilent) prior to sequencing.

ChIP-seq and CUT&RUN analysis

For the ChIP-seq analysis using K562 cells, sequenced samples were trimmed using Cutadapt (version 4.2.0).[95] We aligned single-end reads to hg19 using bwa (version 0.7.17) and called peaks using MACS2 (version 2.1.0) with input control.[31,96] All peak calling was performed according to the ENCODE standards using an irreproducible discovery rate (IDR) of 0.05.[32] Non-uniquely mapped reads and reads mapped to ENCODE blacklist region[97] composed of artificially high regions of the genome were discarded. Coverage files were generated using deepTools (version 3.5.1) and visualized using IGV (version 2.8.10).[85,98] ChIP-seq for CUX1, SMARCA4 and histone marks H3K27ac were performed at the McNerney lab. GATA1 (ENCSR000EWM) and RUNX1 (ENCSR414TYY) ChIP-seq data were obtained from ENCODE. We assigned peaks to the single nearest transcription start site (TSS) within 1 Mb using GREAT (version 4.0.4).[45] Bed files were analyzed using Bedtools (version 2.30.0).[99] Significance of overlap of binding sites between two ChIP-seq experiments was calculated using the hypergeometric test with makeVennDiagram() from ChIPpeakAnno package (version 3.32.0), with options: ''totalTest = totalTest, scaled = FALSE, euler.d = FALSE, method = "hyperG''.[100] We used MEME-ChIP for motif discovery using the classical mode.[84,101] Summits of CUX1 and SMARCA4 binding sites were calculated and extended in both direction by 250 base pairs as the sequence input. Accessible chromatin sites obtained from the K562 gHPRT ATAC-seq were used as the background model. Differential motif analysis was performed using AME.[102] For Figure 1I, published K562 ChIP-seq data for hematopoietic TFs were obtained from ENCODE.[33] As non-CUX1 recruited SMARCA4 sites have on average ~1.97X increased ATAC-seq signal than the CUX1-recruited SMARCA4 sites, the middle 25% quantiles of both groups, which have comparable ATAC-seq signal, were selected for comparison to control for the impact of differential DNA accessibility on TF ChIP-seq signals. For CUT&RUN analysis for CUX1, GATA1 and SMARCA4 in primary human CD34$^+$ HSPCs, all analysis methods and parameters are the same as in ChIP-seq, except the sequencing reads are paired end.

ATAC-seq analysis

For both K562 and human CD34$^+$ HSPC ATAC-seq analysis, sequenced samples were trimmed using Cutadapt (version 4.2.0).[95] We aligned paired-end reads to the human hg19 genome using bwa (version 0.7.17) and called peaks using MACS2(version 2.1.0) with ''–nomodel, – shift −75, and – extsize 150'' options.[31,96] Non-uniquely mapped reads, mitochondrial reads, and reads mapped to the ENCODE blacklist region[97] were discarded. Coverage files were generated using deepTools (version 3.5.1) and visualized using IGV(version 2.8.10).[85,98] Differentially accessible regions in gCUX1 (exon 4) vs. gHPRT samples were identified using csaw using a 2-fold enrichment threshold and FDR smaller than 0.05.[41] Bed files were analyzed using Bedtools (version 2.30.0).[99] For integration with RNA-seq in Figure 4C, ATAC-seq peaks are identified to be the significant ATAC peaks called by csaw[41] within ±1 Mb window from the TSS of the differentially expressed genes(FDR<0.1, $|Log_2FC|>0.75$) identified from RNA-seq in shCUX1 vs. shControl.[13] 406/432 DEGs have significant ATAC peaks within ±1 Mb window from their TSS and are thus retained for this analysis. The ATAC peak with highest Log_2FC for each gene was selected.

Figures 2D–2F were generated from K562 gCUX1 exon 4 and gCUX1 exon 6 targeted clones, which have Pearson correlation coefficients of 0.6–0.75 between replicates.

Analysis of chromatin accessibility at cell type specific enhancers

We downloaded the *cis*-regulatory element annotation map generated by Zhang and Hardison, 2017[67] for primary human hematopoietic cell types including HSC, MEP, GMP, CLP, erythrocyte, megakaryocytes, neutrophils, monocytes, B cell, NK cells, CD4$^+$ and CD8$^+$ T cells from the Validated Systematic IntegratiON of hematopoietic epigenomes (VISION) data portal. (https://usevision.org/). We retained all the genomic intervals identified as enhancers for each cell type, including E: enhancer like; EN: enhancer like, nuclease accessible; EN_A: enhancer like, nuclease accessible, active; E_A: active enhancers; BE: bivalent enhancers; CNE_T: CTCF bound, nuclease accessible, transcribed enhancers; TE_A: transcribed active enhancers; TE: transcribed enhancers. We eliminated all the enhancer elements that are annotated ambiguously as promoter-like. For each progenitor cell type, we eliminated the enhancer elements that are shared in HSCs in order to obtain a list of enhancers that are unique in each specific progenitor cell type. Then we calculated the normalized chromatin accessibility from our CD34$^+$ HSPC ATAC-seq data gHPRT and gCUX1 at the cell type specific enhancers. For plotting, the cell types are merged into lineages: Erythroid (MEP + megakaryocytes + erythrocytes), Myeloid (GMP + neutrophils + monocytes) and Lymphoid (B cells + NK cells + CD4$^+$ and CD8$^+$ T cells). The negative control is a list of 10,000 randomly sampled enhancers that did not appear in any of the cell type specific enhancer lists.

Annotation of peaks with chromatin state

ChIP-seq, CUT&RUN and ATAC-seq peaks were annotated with chromatin state using publicly available data. K562 chromatin state prediction was obtained from UCSC genome browser chromHMM track, which uses hidden Markov model analysis of eight chromatin marks and CTCF ChIP-seq data.[103,104] Primary human CD34$^+$ HSPC chromatin state data was obtained from NIH Roadmap Epigenomics Mapping Consortium (EP50 primary hematopoietic stem cells G-CSF-mobilized female chromHMM track).[44] The database also used hidden Markov model analysis of six chromatin marks and DNase I hypersensitivity data. To establish the chromatin

state of genomic sites, we used Bedtools intersect (version 2.29.0) to obtain the overlap of each ChIP site with chromHMM annotations.[99]

Hi-C analysis

Hi-C data from $CD34^{+}CD38^{-}$primary human HSPC was obtained from a published study.[59] The Hi-C interaction loops in HSPC we used was from Table S2 of that study. We intersected CUX1-bound promoters (defined as CUX1 binding sites in human $CD34^{+}$ primary HSPC CUT&RUN that fall within 2 kb from the TSS) with the 2,684 chromatin loops called by Zhang et al. 2020.[59] 272 loops were found to contain CUX1-bound promoters. We then found the interacting regions of these 272 loops and defined them as the regions that contain putative enhancers in contact with CUX1-bound promoters. Normalized ATAC seq reads (RPKM) in gHPRT and gCUX1 samples on these regions were calculated using deepTools (version 3.5.1).[98] As the negative control, we size-matched and randomly sampled 272 regions that are not in contact with any CUX1-bound promoters.

Murine HSPC fate prediction

We obtain the scRNA lineage tracing data from Weinreb et al. 2020,[68] where murine Lin-Sca^{high} Kit+ HSCs were clonally traced by expressed DNA barcodes so that the terminally differentiated daughter cell fates are linked with ancestor HSC single cell transcriptomes. We downloaded the *in vivo* normalized count matrix and metadata containing the single cell clonal identities from the GEO database (GEO:GSE140802). Seurat V4 was used to import, preprocess and analyze the data.[105] Ancestor HSCs and daughter cells were assigned to their clonal identities. We filtered the cells that do not belong to any clones and the HSCs that do not have any daughter cells. All terminal cell fate annotations were stored in a list. We then looped through this list and determined the most common (if there is one) cell fate. For example, a clonal lineage with the cell fates A, A, B will be determined as being a clonal lineage A, while one with the cell fates A, B will be listed as ambiguous. This gave us 1,523 cells with unique terminal identities after removing cells with ambiguous or undifferentiated cell fates. The remaining cells contained basophil, dendritic cells, monocytes, neutrophiles, B cells, and erythrocyte progenitors. After building datasets with the gene expression levels matrix on one side and the cell fate on the other, we ran different Python scikit-learn machine learning models and graded their accuracy (We chose F1 score as the measurement for prediction performance due to label inbalance and the better control on type I and II errors) to determine how informative different sets of genes were in determining cell fate.[106] The two models we used were ''LogisticRegression'' and deep neural network (implemented by MLPClassifier). For the MLPClassifier, hyper parameter tunning using ''GridSearchCV'' was performed to identify the best parameters on each dataset. We then ran the prediction model and compared the cell fate prediction accuracy for different gene sets. Since some of these gene sets came from experiments on human and our training data is from mouse, we had to convert the gene sets from human to the corresponding mouse gene names using the R package biomaRt.[107] For each gene set, to reduce sampling bias, we performed 50 bootstrap analyses and took the average and standard deviation of the scores. The average accuracy for each of our models was recorded. From previously published studies, we obtained genes corresponding to PU.1 ($n = 2,074$) and RUNX1 binding sites ($n = 5391$), and PU.1 and RUNX1 bound genes that are differentially expressed after they were lost in HSPCs ($n = 336$ and $n = 325$ respectively).[69–72] We performed two-sided Wilcoxon rank-sum test on the F1 scores between our experimental datasets and the most variable genes, randomly selected genes, and mouse transcription factors obtained from AnimalTFDB 3.0.[88]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical tests were performed with R (version 4.3.2, Posit). The statistical test performed include hypergeometric test, Wilcoxon rank-sum tests, and Spearman's correlation test. $p < 0.05$ was defined as statistically significant. Sample sizes and biological replicates are indicated in the figure legends and main texts.