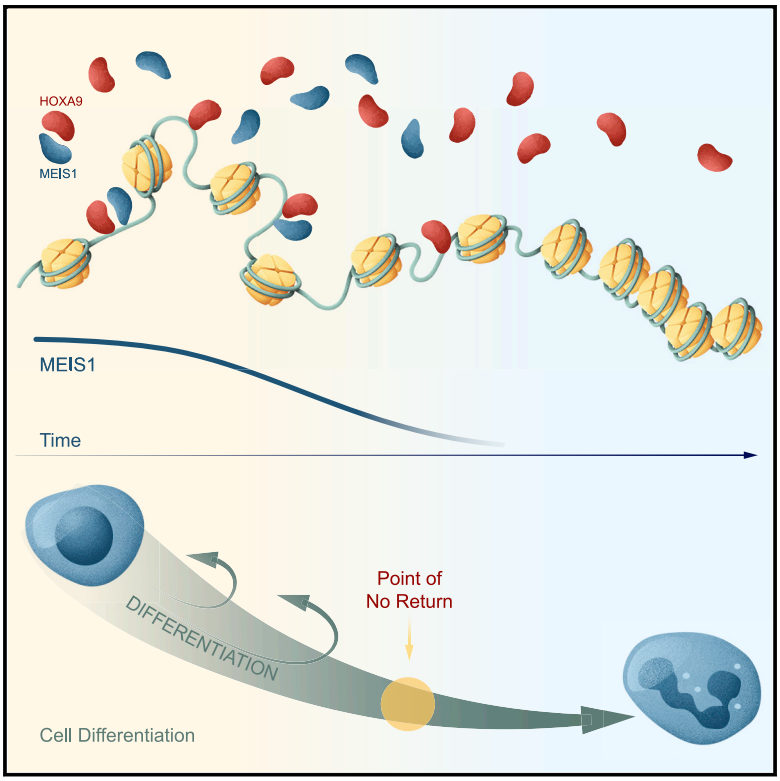
Cell Reports

Article

Chromatin-state barriers enforce an irreversible mammalian cell fate decision

Graphical abstract



Authors

M. Andrés Blanco, David B. Sykes, Lei Gu, ..., Konrad Hochedlinger, David T. Scadden, Yang Shi

Correspondence

ablanco@vet.upenn.edu (M.A.B.), david_scadden@harvard.edu (D.T.S.), yang.shi@ludwig.ox.ac.uk (Y.S.)

In brief

Blanco et al. investigate the balance between self-renewal and differentiation and the mechanism of irreversible cell fate commitment. Using a model of conditional myeloid development, they find that loss of accessible chromatin and disruption of a transcription factor feedback loop underlies the "point of no return" in terminal myeloid differentiation.

Highlights

- There exists a point of irreversible commitment in granulocytic differentiation
- Chromatin-state dynamics establish the transition from self-renewal to differentiation
- Reduced chromatin accessibility underlies irreversible loss of regulatory site access
- Restoration of a transcription factor feedback loop alters differentiation commitment

Blanco et al., 2021, Cell Reports *37*, 109967
November 9, 2021
https://doi.org/10.1016/j.celrep.2021.109967

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Chromatin-state barriers enforce an irreversible mammalian cell fate decision

M. Andrés Blanco,[1,2,12,15,19,*] David B. Sykes,[5,7,15] Lei Gu,[2,12,13,14,15] Mengjun Wu,[2,12,16] Ricardo Petroni,[1] Rahul Karnik,[6,7,8] Mathias Wawer,[9] Joshua Rico,[1] Haitao Li,[1] William D. Jacobus,[2,12,18] Ashwini Jambhekar,[2,12,17] Sihem Cheloufi,[4] Alexander Meissner,[6,7,8,10] Konrad Hochedlinger,[5,6,7,11] David T. Scadden,[5,7,8,*] and Yang Shi[2,3,*]

[1]Department of Biomedical Sciences, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, PA, USA
[2]Division of Newborn Medicine, Boston Children's Hospital, Boston, MA, USA
[3]Ludwig Institute for Cancer Research, Oxford Branch, Oxford University, Oxford, UK
[4]Department of Biochemistry, Stem Cell Center, University of California, Riverside, Riverside, CA, USA
[5]Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA, USA
[6]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[7]Harvard Stem Cell Institute, Cambridge, MA, USA
[8]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA
[9]Center for the Development of Therapeutics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[10]Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin, Germany
[11]Department of Molecular Biology and Cancer Center, Massachusetts General Hospital, Boston, MA, USA
[12]Department of Cell Biology, Harvard Medical School, Boston, MA, USA
[13]Cardiopulmonary Institute (CPI), Bad Nauheim, Germany
[14]Epigenetics Laboratory, Max Planck Institute for Heart and Lung Research, Bad Nauheim, Germany
[15]These authors contributed equally
[16]Present address: The Bioinformatics Centre, Department of Biology and Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark
[17]Present address: Department of Systems Biology, Harvard Medical School, Boston, MA, USA
[18]Present address: Kansas City University, Kansas City, MO, USA
[19]Lead contact
*Correspondence: ablanco@vet.upenn.edu (M.A.B.), david_scadden@harvard.edu (D.T.S.), yang.shi@ludwig.ox.ac.uk (Y.S.)
https://doi.org/10.1016/j.celrep.2021.109967

SUMMARY

Stem and progenitor cells have the capacity to balance self-renewal and differentiation. Hematopoietic myeloid progenitors replenish more than 25 billion terminally differentiated neutrophils every day under homeostatic conditions and can increase this output in response to stress or infection. At what point along the spectrum of maturation do progenitors lose capacity for self-renewal and become irreversibly committed to differentiation? Using a system of conditional myeloid development that can be toggled between self-renewal and differentiation, we interrogate determinants of this "point of no return" in differentiation commitment. Irreversible commitment is due primarily to loss of open regulatory site access and disruption of a positive feedback transcription factor activation loop. Restoration of the transcription factor feedback loop extends the window of cell plasticity and alters the point of no return. These findings demonstrate how the chromatin state enforces and perpetuates cell fate and identify potential avenues for manipulating cell identity.

INTRODUCTION

Differentiation is understood to be a process whereby a stem or progenitor cell divides, matures, and eventually exits the cell cycle as a "terminally" differentiated effector cell. An intriguing and fundamental aspect of differentiation is its unidirectionality; with rare exceptions (Murata et al., 2020; Schwitalla et al., 2013; Tata et al., 2013; Tetteh et al., 2016), mammalian cells do not revert or de-differentiate under physiological conditions (Doulatov et al., 2012; Nichols and Smith, 2012). This suggests that cell circuits embed critical stages of cell fate commitment, beyond which the differentiation program is irreversible. While the concept of differentiation commitment is well appreciated, the mechanistic basis of this process is not understood. What are the molecular "locking" factors that enforce cell fate along the pathway from progenitor to effector cell? This question has significant biomedical ramifications, as successful manipulation of cellular identity is critical for tissue regeneration and pro-differentiation-based cancer therapies.

Several hypotheses seek to explain the irreversibility of differentiation. Epigenomic analyses of the induced differentiation of pluripotent cells offer correlative support to the importance of chromatin dynamics and describe a program in which lineage-specific transcription factors (TFs) are induced and then

reinforced by stable, chromatin-based silencing of pluripotency gene expression programs (Gifford et al., 2013; Suelves et al., 2016; Xie et al., 2013). Proposed silencing mechanisms (Feldman et al., 2006; Nicetto and Zaret, 2019) include DNA methylation and repressive histone modifications such as H3K27me3 and H3K9me3. These modifications are antagonistic to transcription and are self-perpetuating (Hansen et al., 2008; Sharif et al., 2007), offering a means of permanently silencing transcription in the absence of continued stimulus. Studies in reprogramming also indirectly support this hypothesis. By exogenously expressing four key TFs, mature cells can be reverted to pluripotent cells (Takahashi and Yamanaka, 2006). Importantly, three of these four factors—OCT3/4, SOX2, and KLF4—are "pioneer" TFs, meaning that they can bind their DNA motifs even in the presence of nucleosomes and may override repressive chromatin states (Soufi et al., 2012, 2015). However, most TFs are not pioneer factors, allowing repressive chromatin states to enforce cellular identity under normal conditions.

While compelling, these hypotheses have been difficult to test. Time-course profiling of differentiation programs captures features that correlate with maturation, though it is not clear which of these features are causal in preventing reversion of the program (Stricker et al., 2017). Furthermore, the large-scale profiling studies that generate such models most commonly focus on the differentiation of embryonic stem cells (Feldman et al., 2006; Gifford et al., 2013; Gunne-Braden et al., 2020; Suelves et al., 2016; Xie et al., 2013). These models may be applicable for cell fate decisions in early embryogenesis, but it is not known whether they also apply to adult stem and progenitor cell developmental programs.

To understand mammalian cell differentiation commitment, we employed a system of conditional myeloid progenitor cell differentiation arrest that can be released from differentiation blockade under tight temporal control (Sykes et al., 2016). Using this system, we identified a critical differentiation commitment window, beyond which the cells are incapable of returning to the progenitor state. By performing epigenomic profiling before, during, and after this window of commitment, we identified the chromatin-state and transcriptomic processes that determine differentiation commitment.

Unexpectedly, we did not observe promoter DNA methylation or H3K27me3 as contributors of differentiation commitment. Instead, the most striking dynamics in this program involved chromatin remodeling, enhancer activation, and TF usage. Differentiation was accompanied by a progressive global loss of accessible chromatin and concomitant stable silencing of progenitor-state TFs. Here, we propose a parsimonious model in which the loss of access to active regulatory sites is sufficient to disable TF-driven re-acquisition of the progenitor state. This model highlights the essential interaction between TFs and chromatin and demonstrates how the enhancer landscape can act as a barrier to de-differentiation.

RESULTS

Myeloid differentiation is initiated by repression of progenitor-state transcriptional programs

Studies of myeloid cell differentiation commonly rely on sorted populations of primary cells, which are limited in abundance, or on cancer cell lines, which often differentiate aberrantly and in response to non-physiologic stimuli. To study the mechanism of normal differentiation commitment, we used a system that provides an abundant supply of diploid progenitor cells capable of synchronous and terminal differentiation. The ER-HOXA9 and ER-HOXB8 models of conditional differentiation arrest offer these benefits (Sykes et al., 2016; Wang et al., 2006). In the presence of estrogen (beta-estradiol, E2), the ER-HOXA9 TF is active and prevents differentiation beyond the granulocyte-monocyte precursor (GMP) stage, allowing for unlimited expansion. Withdrawal of estrogen rapidly inactivates the ER-HOXA9 protein via nuclear export and cytoplasmic sequestration. Following ER-HOXA9 inactivation, cells synchronously progress through normal myeloid differentiation over ~5 days to fully mature neutrophils and (less commonly) monocytes.

In utilizing HOXA9 (or HOXB8) to arrest differentiation, this model takes advantage of our understanding of normal hematopoietic development. Progenitors must downregulate *Hoxa9* expression as they commit to becoming mature monocytes and neutrophils. HOXA9 is a well-appreciated regulator of myeloid cell fate, and mice lacking *Hoxa9* have reductions in circulating white blood cell counts, myeloid colony-forming ability, and responsiveness to G-CSF (Lawrence et al., 1997). The dysregulated persistence of *Hoxa9* expression promotes GMP differentiation arrest and is a key oncogenic event seen in >70% of human acute myeloid leukemia (AML) (Golub et al., 1999).

These particular ER-HOXA9 GMPs were derived from adult bone marrow of a lysozyme-GFP knock-in mouse (Faust et al., 2000). In this context, expression of GFP acts as an endogenous reporter of differentiation status, as only maturing neutrophils activate expression of the secondary granule protein lysozyme.

To investigate the dynamics of differentiation, we analyzed gene expression (RNA sequencing [RNA-seq]) at 10 time points across the 120-h program. The differentiation program proceeded in graded fashion, with progressive and continuous global trends in gene expression (Figures 1A and S1A). We identified six main patterns of gene expression that differed primarily in the kinetics of induction or repression (Figure 1B).

Comparing the 0-h (progenitor state) time point to later time points, gene set enrichment analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005) revealed that the interferon-γ response and tumor necrosis factor (TNF)-α pathway were strongly induced throughout the time course, and dramatic cell-cycle repression and downregulation of Toll-like receptor X signaling emerged near the end of the program (Figures S1B and S1C). At the final time point, the most highly enriched gene sets (of 3,477) were (1) myeloid development and (2) genes downregulated upon overexpression of Hoxa9 and Meis1 (Figures 1C and S1B; Table S1), confirming the full execution of the differentiation program. Enrichment of these signatures showed unstable kinetics over the first 36 h, after which they progressed consistently (Figure S1C, left panel).

Analyses of TF and microRNA (miRNA) binding motif enrichment in promoters of differentially expressed genes identified the loss of MYC activity as an early event, induction of PU.1 and ELF1 targets as intermediate events, and loss of E2F1 as a terminal event (Figure S1C, right panel). Globally, concerted

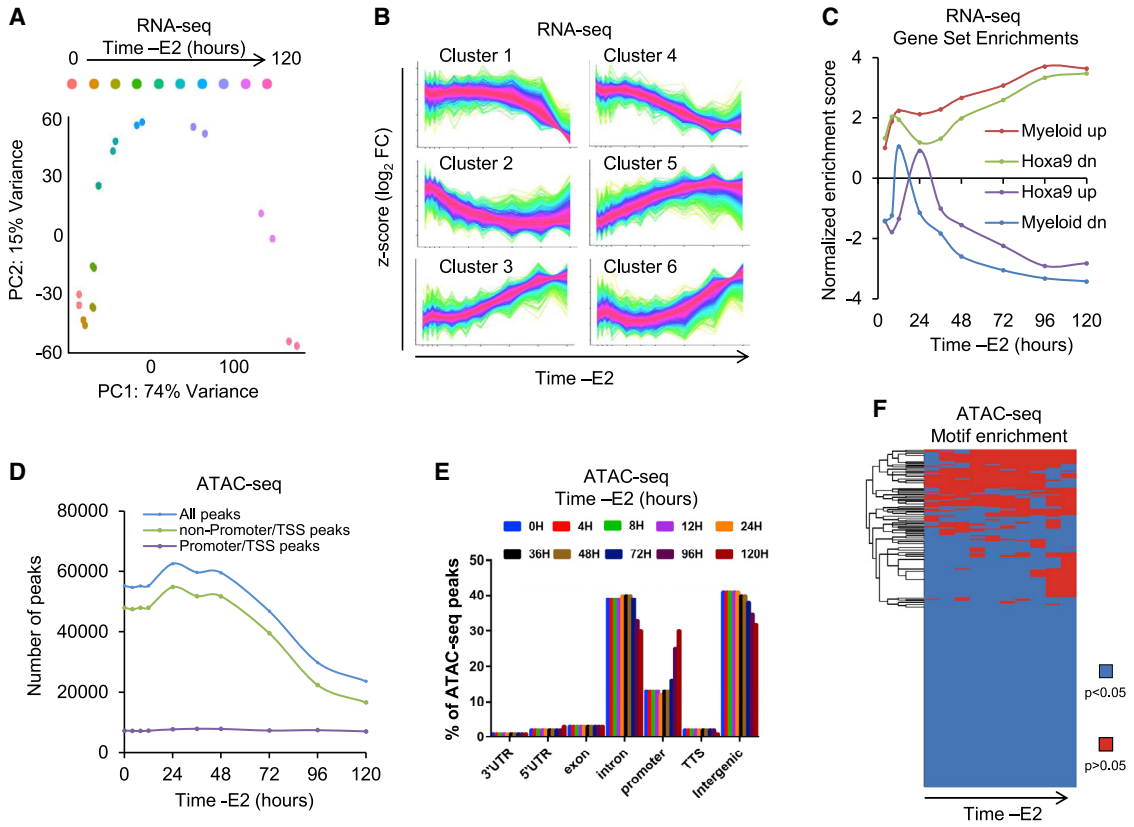


Figure 1. Transcriptomic and ATAC-seq profiling of the myeloid differentiation program

(A) Principal-component analysis (PCA) of RNA-seq at 10 time points in biological duplicate following inactivation of ER-HOXA9 (upon removal of estradiol) over 120 h.

(B) Identification of six main clusters of genes with distinct expression dynamics over the differentiation program. Individual transcripts in each cluster are colored according to degree of correlation to the cluster trend.

(C) Plot of GSEA normalized enrichment scores (NESs) of myeloid differentiation gene sets over the differentiation time-course.

(D) Plot of number of ATAC-seq peaks in ER-HOXA9 cells over the differentiation program.

(E) Distribution of ATAC-seq peaks at each time point according to genome annotation.

(F) Clustering of DNA motifs by degree of enrichment in ATAC-seq peaks of open chromatin at each time point. RNA-seq and ATAC-seq data represent the average of experiments performed in biological duplicate at each time point.

downregulation of active TF/miRNA programs occurs prior to the upregulation of a new set of TF/miRNA programs, suggesting that progenitor gene expression programs are shut down prior to induction of differentiation programs (Figure S1D).

ATAC-seq was performed to investigate epigenomic dynamics accompanying ER-HOXA9 differentiation. While patterns of chromatin accessibility evolved as cells differentiated (Figures S2A and S2C), the total amount of open chromatin appeared relatively constant in the first 2 days of the differentiation program. However, the number of ATAC-seq peaks dropped precipitously from 48 h until terminal differentiation, with fully mature cells harboring less than half the amount of open chromatin as progenitor-state cells (Figure 1D). Relatively few new regions of chromatin opened as cells differentiated, with the majority (75%) of accessible chromatin in differentiated cells already being open at the progenitor state (Figure S2B).

Concomitant with accessibility losses, the genomic distribution of ATAC-seq peaks showed a moderate relative shift from introns and intergenic regions to promoter/TSS regions (Figure 1E). Interestingly, this shift did not reflect an increase in global promoter accessibility, as the absolute number of promoter/TSS peaks was constant through the time course, but instead resulted from the large drop in the absolute number of intron/intergenic peaks starting around 48 h (Figures 1D and 1E).

Accessible chromatin in the progenitor state was enriched for binding sites of numerous TFs expressed in both the progenitor and the mature state (Figures 1F and S2D; Table S2). Motifs enriched specifically at early time points included binding sites for progenitor-state TFs such as c-MYC, GATA factors, and HOX TFs. The small number of motifs enriched only in mature cells included binding sites for the neutrophil master regulators GFI1b, HIF1A, and nulear factor (NF)-κB, suggesting the induction of these transcriptional programs late in differentiation.

Collectively, these data suggest that in the progenitor state, the chromatin landscape of ER-HOXA9 cells is already accessible to drivers of the differentiation program, but progenitor-state transcriptional programs must be silenced prior to the

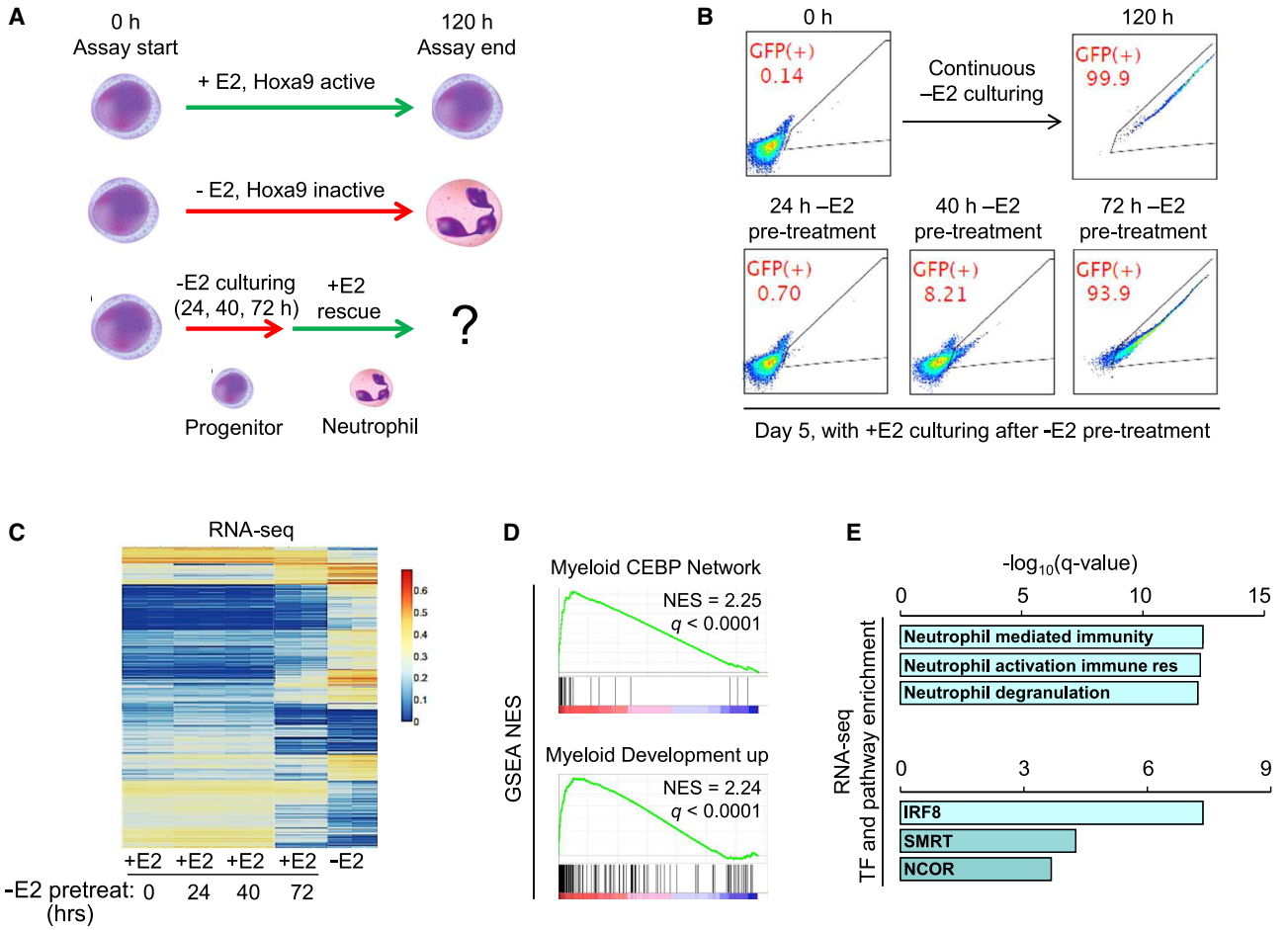


Figure 2. Identification of differentiation commitment "point of no return"

(A) Schematic of ER-HOXA9 inactivation and reactivation workflow.

(B) Flow cytometry of GFP differentiation reporter in ER-HOXA9 cells in the progenitor state (0 h –E2), in the terminally differentiated state (120 h –E2), or in cells receiving –E2 pretreatments followed by E2 add-back to reactivate ER-HOXA9. Images shown are representatives from biological triplicates.

(C) RNA-seq of cells cultured +E2 (progenitor state), 120 h –E2 (terminally differentiated state), or +E2 after –E2 pretreatments.

(D) GSEA indicates that cells receiving 72-h –E2 pretreatment stably maintain enrichment of the myeloid development maturation gene expression programs after ER-HOXA9 reactivation.

(E) GO biological process enrichment (top) and ChIP-seq-based TF target enrichment (bottom) in the 200 genes most overexpressed in cells receiving 72-h versus 0-h –E2 pretreatment. RNA-seq data represent the average of experiments performed in biological duplicate at each time point.

execution of terminal differentiation with concomitant losses in chromatin accessibility.

Identification of an irreversible "point of no return" in the myeloid differentiation program

In normal mammalian cells, terminal differentiation is thought to be irreversible. However, our transcriptional and epigenomic profiling did not identify any obvious inflection points or rapidly induced or lost gene expression programs that would suggest a critical commitment point in the program. To test whether the myeloid differentiation program is irreversible, we temporarily inactivated ER-HOXA9 by estrogen withdrawal ("–E2 pretreatment") for varying periods of time before reactivating ER-HOXA9 by the reintroduction of estrogen followed by monitoring the extent of differentiation (Figure 2A).

The inactivation of ER-HOXA9 for 24 or 40 h initiated early to intermediate steps of the differentiation program (Figures 1A–1C). However, these early events in differentiation were reversible, as reactivation of ER- HOXA9 overrode these changes and reverted cells to the starting progenitor state (Figure 2B). In contrast, after inactivating ER-HOXA9 for >72 h, the cells were stably committed to differentiation and could no longer be returned to the progenitor state upon ER-HOXA9 reactivation (Figure 2B); the cells had achieved a "point of no return" and inexorably progressed to terminal differentiation.

To confirm that this differentiation point of no return was not specific to a single ER-HOXA9 clone, we repeated the estrogen withdrawal experiments in other ER-HOXA9 clones as well as in the parallel system immortalized by ER-HOXB8. While the exact time of irreversibility varied slightly, all models demonstrated a clear commitment point from which only rare cells could

escape—even if the assay was extended to 7 days (Figures S3A and S3B).

In pre-commitment cells (24 or 40 h –E2), reactivation of ER-HOXA9 successfully re-established a pattern of gene expression that was nearly indistinguishable from the transcriptome of naive +E2 progenitor cells (Figure 2C). In contrast, in cells withdrawn from E2 for 72 h, reactivation of ER-HOXA9 was unable to reverse the global induction of differentiation. The transcriptomes of these cells remained markedly different from progenitors and were strongly enriched for myeloid development gene signatures and activated neutrophil gene ontologies (Figures 2C–2E). By 72 h –E2, the cells had passed a key commitment point, beyond which their phenotype was maintained in the absence of the originating stimulus. While not all cells proceeded to the very final stages of terminal differentiation (as seen in the 120-h –E2 samples), we reasoned that the transcriptional changes maintained after 72 h out of estrogen may represent the most critical processes for preventing reversion of the cells to the progenitor state.

Promoter DNA methylation and H3K27me3 silencing do not influence myeloid differentiation commitment

How was the differentiation gene expression program maintained following reactivation of ER-HOXA9? DNA 5mC methylation and PRC2-mediated H3K27me3 are repressive chromatin modifications that can be self-propagated and, during ESC differentiation, often localize to genomic loci that are stably silenced in the chosen lineage (Gifford et al., 2013; Suelves et al., 2016; Xie et al., 2013). We hypothesized that DNA methylation and/or H3K27me3 may facilitate commitment to myeloid differentiation.

We performed reduced representation bisulfite sequencing (RRBS) (Meissner et al., 2005) of 5mC as well as H3K27me3 chromatin immunoprecipitation sequencing (ChIP-seq) on progenitor cells (+E2) and on terminally differentiated cells (–E2 for 120 h), as well as after withdrawing E2 for varying time periods before ER-HOXA9 reactivation. Unexpectedly, very few promoters showed differentially methylated regions (DMRs). The methylomes were so similar that biological replicates did not always cluster together (Figure S3C), and the few promoter DMRs did not correlate with transcriptional silencing (Figure S3D). Similar results were found when considering DNA methylation in H3K27ac ChIP-seq peaks (discussed below) marking putative active regulatory sites beyond promoters (Figure S3E). This suggests that DNA CpG methylation at promoters and regulatory elements does not strongly influence the transcriptional dynamics in this differentiation program.

In progenitor cells, H3K27me3 peaks were enriched at genes regulating early embryonic developmental programs, with dramatic enrichment ($p < 1 \times 10^{-102}$) for homeodomain-containing genes (Table S3). However, global H3K27me3 enrichment patterns did not change appreciably in differentiated cells (Table S3; Figures S4A–S4C), and H3K27me3-marked genes were not enriched for Gene Ontology (GO) categories relevant to myeloid differentiation (data not shown). Furthermore, loss or gain of H3K27me3 from the progenitor to the differentiated state did not correlate with gene expression changes (Figure S4D). More specifically, genes comprising the myeloid development expression signature, highly differentially expressed over the differentiation program, had no appreciable changes in H3K27me3 dynamics (Figure S4E). We, therefore, conclude that similarly to DNA promoter CpG methylation, H3K27me3 dynamics are unlikely to drive myeloid differentiation commitment.

Chromatin remodeling is highly dynamic during differentiation

Given that we did not identify a role for DNA promoter CpG methylation or H3K27me3 in enforcing differentiation commitment, and considering the dramatic decline in chromatin accessibility observed via ATAC-seq (Figure 1D), we investigated whether chromatin remodeling and regulatory site access dynamics could lock cells into the maturation program. We compared ATAC-seq signatures among progenitor cells, mature cells, and cells following the estrogen removal and add-back described above. We performed ChIP-seq of H3K27ac and H3K4me3 to map regulatory elements. ATAC-seq trends were highly dynamic, with unbiased clustering identifying seven main patterns of loci accessibility during maturation (Figure S5A). Progenitor cells were observed to have roughly twice as many peaks as the terminally differentiated cells (Figures 3A and 3B). Intriguingly, longer –E2 pretreatments led to greater irreversible losses of ATAC-seq peaks up to the 56-h –E2 pretreatment sample. Defining "epigenetic" as a phenotype that is long-lasting, stably heritable across multiple cell divisions, and not resulting from DNA sequence changes, this suggests that there is significant epigenetic "memory" of the low chromatin accessibility state even after ER-HOXA9 reactivation.

Comparing chromatin accessibility in the 40-h and 72-h –E2 pretreatment populations, differentiation was initiated in both samples (Figures 1C and 2B), but only the 40-h sample could be reverted to the progenitor state by ER-HOXA9 reactivation. Three main clusters of accessible chromatin dynamics were identified: two regions that lost accessibility and one region that gained accessibility (Figure 3C). Motif enrichment of these genomic regions showed that many TF binding sites become irreversibly depleted from accessible chromatin in cells 72 h out of estrogen as compared to 40 h, while few motifs gained enrichment (Figures 3D and 3E).

TFs with binding motifs no longer enriched in open chromatin after the commitment point include many notable drivers of the progenitor/proliferation program, such as RUNX1 and 2, ERG, E2F1, and several ETS factors. Concordantly, loci of canonical transcriptional targets of these TFs—such as *Bcl2* and *Hmga2* (RUNX1; Lam et al., 2014), *Cdkn1a* and *Rfc1* (E2F1; Subramanian et al., 2005), and *Gata2* and *Arid3a* (ERG; Goldberg et al., 2013)—had regions of open chromatin in the +E2 and 40-h –E2 pretreatment samples, but not in the –E2 or 72-h –E2 pretreatment samples (Figure S5B). Thus, a stable, irreversible reduction of TF access to chromatin distinguishes cells that pass beyond the differentiation commitment point.

Altogether, this suggests a model in which transcriptional repression of pro-stemness TFs is followed by losses in chromatin accessibility at their target loci, blocking their access to chromatin upon re-expression. This offers a parsimonious explanation for why long-term repressive epigenetic modifications such as H3K27me3 and promoter DNA methylation may not be

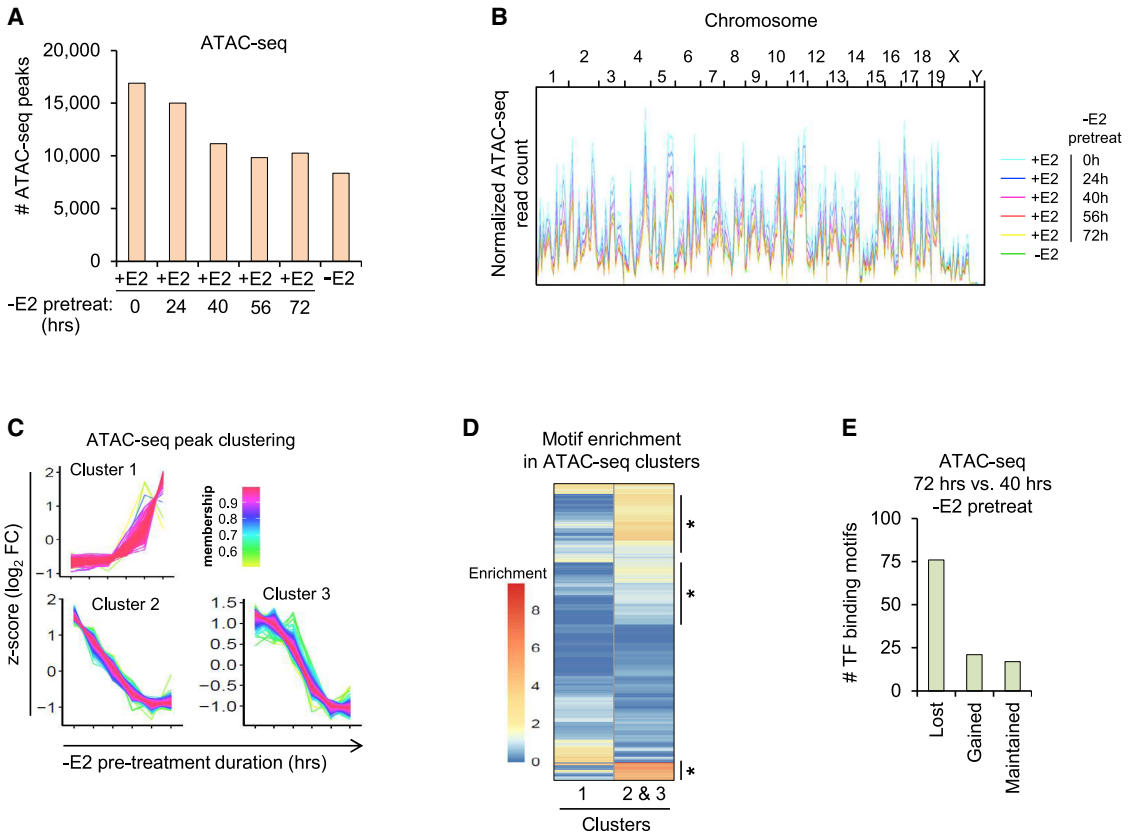


Figure 3. Chromatin remodeling dynamics in differentiation commitment

(A) Number of ATAC-seq peaks in ER-HOXA9 cells cultured +E2 (progenitor state), –E2 for 120 h (terminally differentiated state), or +E2 after –E2 pretreatments of varying duration.

(B) Normalized ATAC-seq read counts of samples across all chromosomes.

(C) Identified clusters of genomic loci undergoing the most dramatic ATAC-seq dynamics in samples receiving 40-h compared to 72-h –E2 pretreatment prior to E2 add-back.

(D) TF binding motif enrichment in ATAC-seq peaks found in clusters shown in (C). Heatmap is colored by motif enrichment score in each sample. Clusters of motifs that are enriched in 72-h –E2 pretreatments are noted with lines and asterisks.

(E) Quantification of binding motif enrichment trends in heatmap shown in (D). ATAC-seq was performed in biological duplicate at all time points, and results represent the average at each time-point.

pre-requisites for differentiation commitment in adult progenitor cells.

Identification of ER-HOXA9 direct target genes

To test our model, we focused on the most critical pro-stemness driver TF in this system, ER-HOXA9. Our model predicts that once the differentiation program is initiated, ER-HOXA9 can access its progenitor-state binding sites only if re-activated prior to the commitment point. To identify direct ER-HOXA9 binding sites, we turned to ChIP-seq performed in cell lines in which the ER-HOXA9 (or wild-type [WT] HOXA9) construct harbored both V5 (N-terminal) and AM (C-terminal) epitope tags. In the V5-HOX9-AM cell line, V5 ChIPs yielded more peaks than AM ChIPs (17,006 versus 6,773, respectively), while other results were highly consistent. V5 and AM ChIPs produced extremely similar global enrichment patterns with expected mapping to intronic and distal regions (Huang et al., 2012; Sun et al., 2018) and had highly overlapping gene targets, GO enrichments, and motif enrichments (Figures S6A–S6F; Table S4). However, in ChIPs of the ER-HOXA9 fusion protein, the V5 antibody outperformed the AM antibody and closely recapitulated binding patterns of the WT protein (Figures 4A–4D) and was therefore used for all subsequent ChIP experiments.

To identify direct regulatory targets of ER-HOXA9, progenitor-state (+E2) V5-ER-HOXA9-AM ChIP-seq yielded 1,417 peaks that mapped to 908 genes (Table S5). The set of the top 200 putative ER-HOXA9 target genes was rapidly globally downregulated as cells differentiated, reaching maximum repression by 36 h –E2 (Figures 4G and 4H). We defined ER-HOXA9 direct, functional targets as genes that are bound by ER-HOXA9 in progenitor cells and that are significantly downregulated in the first 48 h of differentiation. This yielded 216 genes highly enriched for GO categories involving the immune response and hematopoiesis (Figures 4E–4H; Table S5). Intriguingly, among these targets are several critical transcriptional regulators known to drive the progenitor/proliferative state and contribute to AML, including

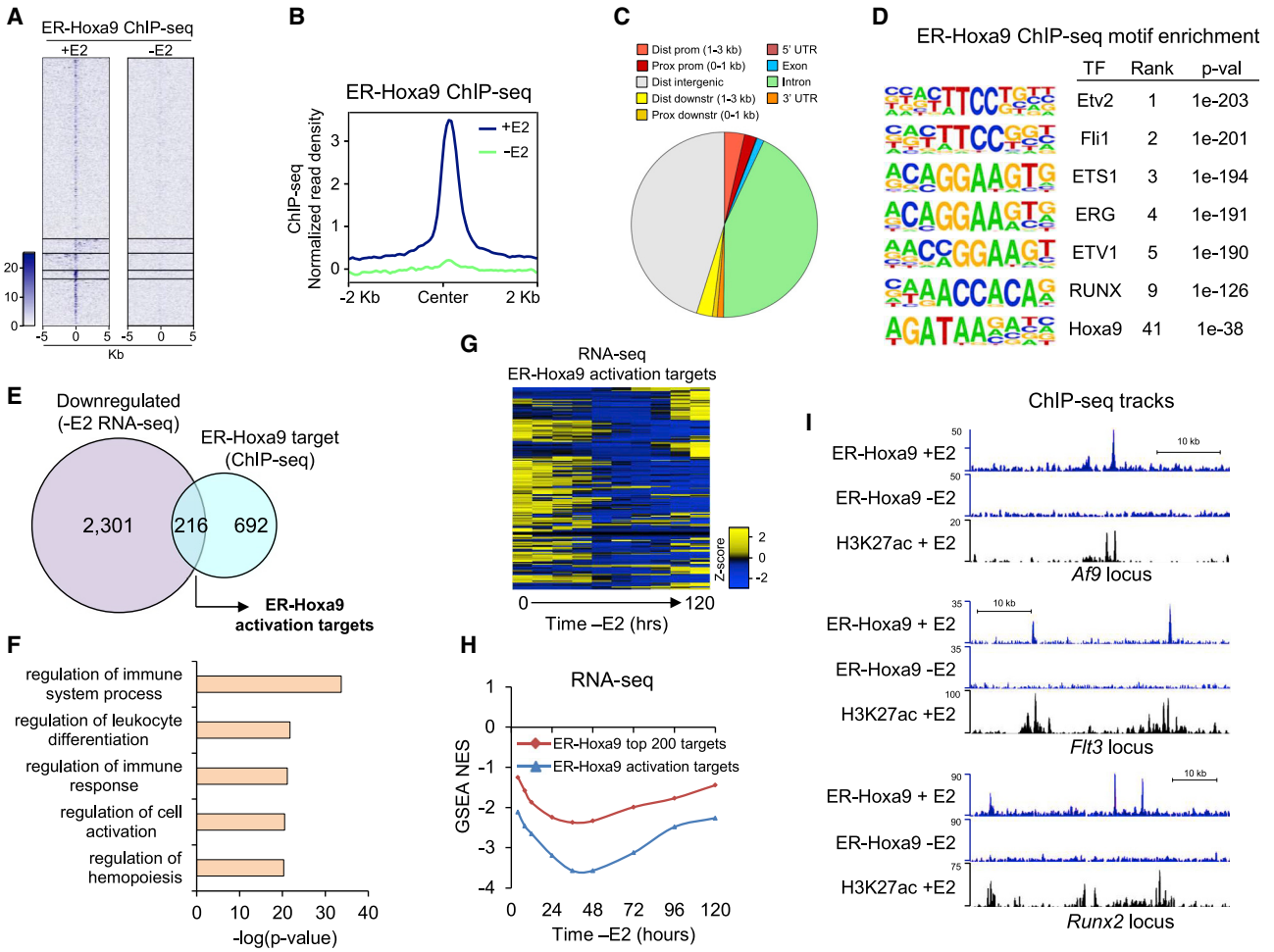


Figure 4. Identification of ER-HOXA9 direct targets

(A) Heatmap of merged peak regions from V5 ChIP-seq of V5-ER-HOXA9-AM cultured +E2 (nuclear V5-ER-HOXA9-AM) or 96 h –E2 (cytoplasmic V5-ER-HOXA9-AM).

(B) Plot of average V5-ER-HOXA9-AM ChIP-seq normalized read density from +E2 and –E2 samples.

(C) Distribution of +E2 V5-ER-HOXA9-AM peaks by genome annotation.

(D) MEME enrichment of known DNA motifs within peaks from +E2 V5-ER-HOXA9-AM ChIP-seq.

(E) Overlap of genes that are transcriptionally downregulated within 48 h of –E2 treatment and genes with proximal +E2 V5-ER-HOXA9-AM ChIP-seq peaks.

(F) Top five GO biological processes enriched in ER-HOXA9 direct activation targets.

(G) Heatmap of transcriptional dynamics of ER-HOXA9 activation targets over the 120-h –E2 differentiation program.

(H) GSEA NESs of V5-ER-HOXA9-AM top 200 gene targets (blue) and 216 direct activation targets (red) tested for enrichment in RNA-seq of genes ranked by expression at each –E2 time point relative to 0-h (+E2) time point.

(I) V5-ER-HOXA9-AM and H3K27ac ChIP-seq tracks at loci three direct activation targets of ER-HOXA9 from cells cultured +E2 or –E2.

Runx2, *Meis1*, *Gfi1b*, and *Af9*, as well as other AML oncogenes such as *Flt3* and *Msi2* (Figure 4I). The direct regulation of other master TFs by ER-HOXA9 suggests that orchestration of progenitor-state identity may utilize a hierarchical TF control model (Yu and Gerstein, 2006), in which a very small number of master regulators have disproportionately large control of gene regulation. This model predicts that TFs at or near the top of the hierarchy have a very small number of most critical regulatory targets.

We then investigated whether ER-HOXA9 can access its binding sites at progenitor-state genes when reactivated in cells that have passed the differentiation commitment point. We induced differentiation by withdrawing E2 for 72 h and then reactivating ER-HOXA9. Consistent with significant loss of chromatin access, reactivated ER-HOXA9 yielded reduced ChIP-seq signal (Figures 5A and 5B) and was bound to only 240 of its original 1,417 progenitor-state binding sites (16.9%), as well as to 187 new binding sites (Figure 5C; Table S5). Critically, only 32 ER-HOXA9 direct activation target genes were bound by reactivated ER-HOXA9 (Table S5), and these did not include any of the master TFs. Despite ER-HOXA9 reactivation, these essential drivers of the progenitor state remained irreversibly downregulated after the commitment point (Figure 5D). This is in line with the observation that the functional activation targets of ER-HOXA9 were among the earliest genes to be silenced upon E2 withdrawal

(Figures 4G and 4H) and, hence, the least likely to retain open chromatin late in the differentiation program. Of note, several of the original target loci that reactivated ER-HOXA9 could bind included genes highly upregulated in differentiated cells (*lysozyme*, *Mmp8*, *Cleck4n*, *Ccl4*).

Chromatin-state barriers restrict ER-HOXA9 activity in post-commitment cells

What features of chromatin precluded ER-HOXA9 binding after the commitment point? Progenitor-state binding sites were divided into those that ER-HOXA9 could (n = 240) or could not (n = 1,117) re-bind when reactivated after the commitment point. These sites were compared for areas of overlap with matching ATAC-seq, H3K27ac, H3K4me3, and H3K27me3 ChIP-seq datasets from post-commitment cells. In support of our model, the loci at which ER-HOXA9 failed to re-bind had dramatic reductions in ATAC-seq peaks, H3K27ac peaks, and super-enhancers compared to the loci with successfully re-bound ER-HOXA9 (Figures 5E and 5F). Peaks of H3K27ac overlapped with 86% of the re-bound ER-HOXA9 sites, but with just 33% of the loci where ER-HOXA9 failed to re-bind. In contrast, there were no significant differences between re-bound and non-re-bound loci in overlap with H3K4me3 or H3K27me3 (Figure 5E; data not shown). These results suggest that the progenitor-state enhancer landscape must be maintained for ER-HOXA9 to successfully drive progenitor cell identity.

Finally, we investigated the sequence content of the most high-confidence ER-HOXA9 target sites by intersecting ER-HOXA9 ChIP-seq and ATAC-seq peaks and performing motif enrichment analyses. As expected, open chromatin (ATAC-seq peaks) intersecting ER-HOXA9 peaks in the progenitor state was highly enriched for binding sites of TFs involved in progenitor biology, including ETS factors, HOXA9 itself, RUNX, MEIS, and PU.1 (Figure 5G; Table S6). Strikingly, fewer than half of these motifs were enriched in open chromatin intersecting ER-HOXA9 peaks in post-commitment cells, with binding sites for key TFs such as MEIS1 and PU.1:IRF notably absent. Accordingly, the loci of several potentially critical ER-HOXA9 targets, such as *Flt3*, *Aldh1b1*, *Pkca*, and *Angpt1*, were inaccessible in post-commitment cells (Figures 5H and 5I; data not shown). This suggests that following differentiation commitment, most of the original ER-HOXA9 binding sites are closed, and those that remain open harbor fewer options for coordinated binding between ER-HOXA9 and other transcriptional drivers of the progenitor state. Ultimately, these events compromise the ability of ER-HOXA9 to re-activate its activation targets, many of which remained irreversibly epigenetically silenced (Figures 5H and 5I).

In summary, in the progenitor state, ER-HOXA9 binds open chromatin at numerous genes and functionally regulates a subset of these. Upon inactivation of ER-HOXA9, its direct targets are rapidly downregulated with initiation of the differentiation transcriptional program, and as chromatin accessibility globally and progressively decreases, access to regulatory sites of these ER-HOXA9 targets declines. By 72 h, most binding sites are closed, and if ER-HOXA9 is re-activated, it can bind to only a small minority of sites. These sites do not include critical progenitor-state regulatory targets but instead correspond to genes that were initially bound by ER-HOXA9. However, they seem to play later roles in driving differentiation and have maintained open chromatin (Figure 7).

Enforced *Meis1* expression alters the point of no return by increasing cellular plasticity

While ER-HOXA9 binds at ~1,000 genes (Table S5), it appears to directly regulate one quarter of these. In this subset are transcriptional regulators of marked importance for GMP biology, including *Gfi1b*, *Meis1*, and *Runx2*. These may represent the most essential ER-HOXA9 targets, since they all have extensive gene regulatory potential. Several lines of evidence suggest that *Meis1* plays an especially critical role in driving the progenitor-state transcriptional program. *Meis1* is highly expressed in progenitor cells and is sharply downregulated with differentiation (Figure 6A). Loss of MEIS1 may strongly affect ER-HOXA9 binding patterns, as the *Meis1* binding site is highly overrepresented in the promoters of genes that ER-HOXA9 fails to re-bind when reactivated after the commitment point (Figure 6B). Importantly, the *Meis1* locus exhibits significant epigenetic control; it is among the most dramatically epigenetically inactivated in post-commitment ER-HOXA9 cells, losing all five of its proximal H3K27ac peaks and all four of its proximal ATAC-seq peaks (Figure 6D; data not shown). Accordingly, while *Meis1* is a robust ER-HOXA9 target in the progenitor state, epigenetic inactivation renders it inaccessible after the commitment point, and its ER-HOXA9 binding site is among the most markedly reduced of all peaks in ER-HOXA9 ChIP-seq of post- versus pre-commitment cells (Figures 6C and 6D). Consequently, *Meis1* downregulation is irreversible after the commitment point, and it remains silenced when ER-HOXA9 is reactivated after the differentiation point of no return (Figure 6E)

MEIS1 heterodimerizes with HOX9 and enhances its DNA-binding and *trans*-activation potential (Collins and Hess, 2016). It follows that direct regulation of *Meis1* by HOXA9 creates a positive feedback loop that would be irreversibly broken by ER-HOXA9 inactivation and subsequent closing of *Meis1* regulatory sites. This suggests that if *Meis1* could be expressed after the commitment point, this feedback loop would be restored, and ER-HOXA9 may be able to access and activate its progenitor-state gene regulatory targets.

To test the role of MEIS1 in differentiation commitment, we derived cell lines with constitutive expression of *Meis1* and *Pbx1*. We used the ER-Hoxb8 model, as they more stably expressed exogenous *Meis1* than ER-HOXA9 cells. Constitutive expression of *Meis1* and *Pbx1* did not affect the ability to differentiate upon inactivation of the ER-HOXB8 protein, and these cells upregulated CD11b along the same time course as control cells expressing ER-HOXB8 only (data not shown).

Commitment point assays were performed in which we inactivated ER-HOXB8 (E2 withdrawal) for varying periods of time before reactivation (E2 add-back). As expected, most ER-HOXB8 cells could not be reverted to the progenitor state after 72 h –E2, and effectively all cells were fully committed to terminal differentiation after 96 h out of estrogen (Figure 6F). In contrast, though CD11b levels of live ER-HOXB8-MEIS1-PBX1 cells were near their maximum at 72–96 h out of

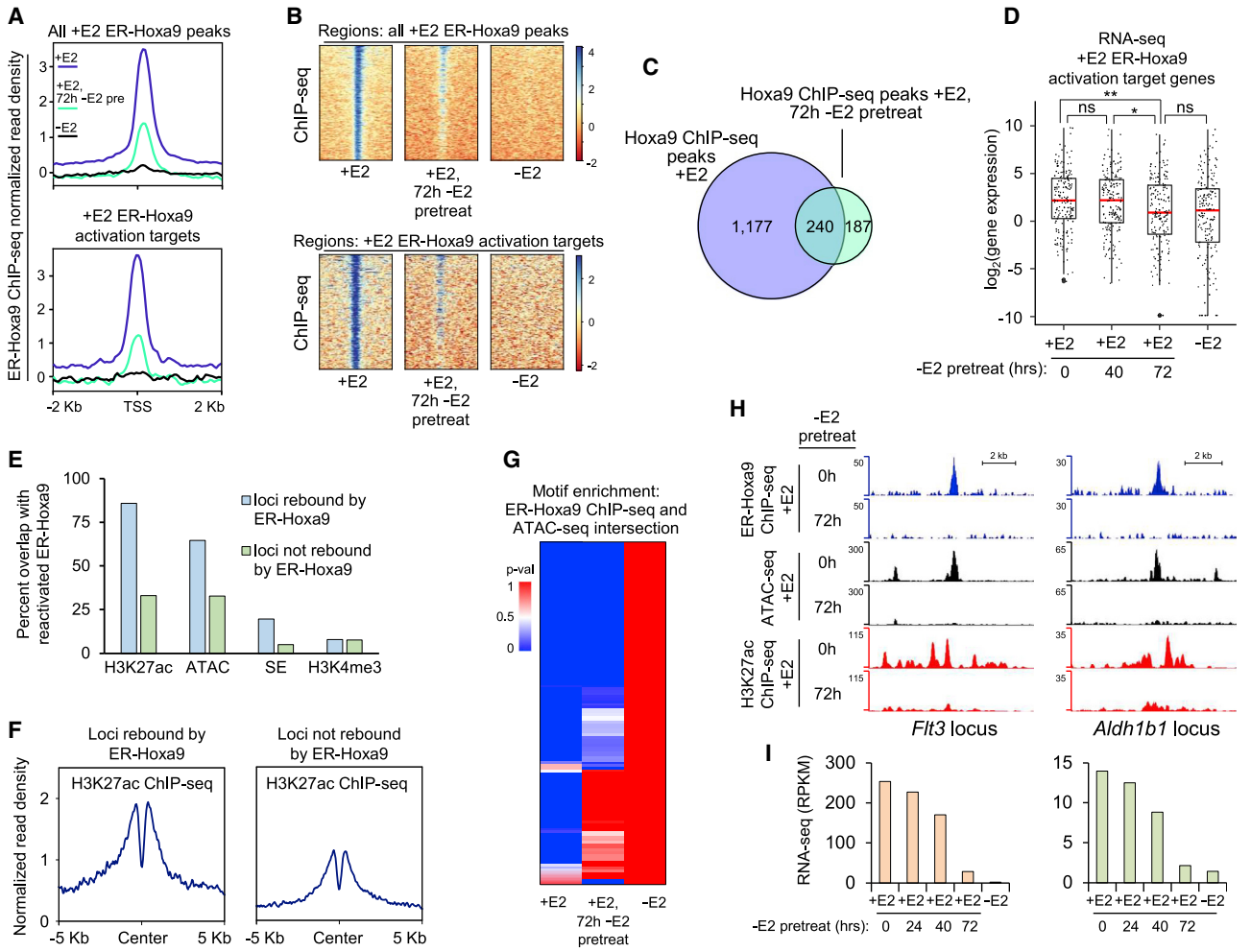


Figure 5. Integrative epigenomics analysis of ER-HOXA9 binding patterns in pre- and post-commitment cells

(A) Plots of normalized read density from ChIP-seq of V5-ER-HOXA9-AM in cells cultured +E2 (purple), –E2 (black), and +E2 after 72-h –E2 pretreatment (light green). Top plot is averaged over all +E2 V5-ER-HOXA9-AM peaks, and bottom plot is averaged over the +E2 direct activation targets of V5-ER-HOXA9-AM.

(B) Heatmap visualization of ChIP-seq data shown in (A).

(C) Overlap of V5-ER-HOXA9-AM peaks from cells cultured +E2 or +E2 after 72-h –E2 pretreatment.

(D) Box-and-whisker plot representation of expression of V5-ER-HOXA9-AM activation target genes in cells cultured +E2, 120 h –E2, or +E2 after –E2 pre-treatments. Red line indicates median, and whiskers represent ± 1.5*IQR (interquartile range). *$p < 0.05$; **$p < 0.01$; t test. ns, not significant.

(E) Percent of V5-ER-HOXA9-AM peaks overlapping an H3K27ac peak, ATAC-seq peak, super-enhancer, or H3K4me3 peak. V5-ER-HOXA9-AM peaks are split into groups of +E2 binding sites that ER-HOXA9 can (light blue) or cannot (light green) re-bind when reactivated after 72-h –E2 pretreatment.

(F) Plots of H3K27ac ChIP-seq average read density in loci that ER-HOXA9 can (left) or cannot (right) re-bind when reactivated after 72-h –E2 pretreatment.

(G) Heatmap of DNA motif enrichments in the intersection of ATAC-seq and V5-ER-HOXA9-AM ChIP-seq peaks in cells cultured +E2, 120 h –E2, or +E2 after 72-h –E2 pretreatment. Heatmap is colored according to p value of motif enrichment.

(H) Tracks of V5-ER-HOXA9-AM and H3K27ac ChIP-seq along and matching ATAC-seq at the *Flt3* (left) and *Aldh1b1* (right) loci in cells cultured +E2, or with or without 72-h –E2 pretreatment.

(I) Expression of *Flt3* (left) and *Aldh1b1* (right) in cells cultured +E2, 120 h –E2, or +E2 after –E2 pretreatments.

estrogen, the reactivation of ER-HOXB8 even at 96 h induced a significant portion to decrease CD11b expression and to return toward the progenitor state (Figure 6F). In control ER-HOXB8 cells lacking *Meis1* expression, >99.9% of cells had terminally differentiated and died by this point. These data suggest that co-expression of ER-HOXB8, MEIS1, and PBX1 endows ER-HOXB8 cells with the ability to go further into the myeloid differentiation program before irreversibly committing to differentiation.

Chromatin accessibility and HOXA9 activity human and mouse myeloid differentiation programs *in vivo*

As the ER-HOXA9 model faithfully recapitulates *in vivo* myeloid differentiation (Sykes et al., 2016), we hypothesized that the key features of our model (Figure 7) would also be observed in epigenomic analyses of normal primary human and mouse myeloid cell populations. We analyzed ATAC-seq and RNA-seq data from sorted human and mouse GMPs and mature monocytes (Corces et al., 2016; Lal et al., 2021; Xiang et al.,

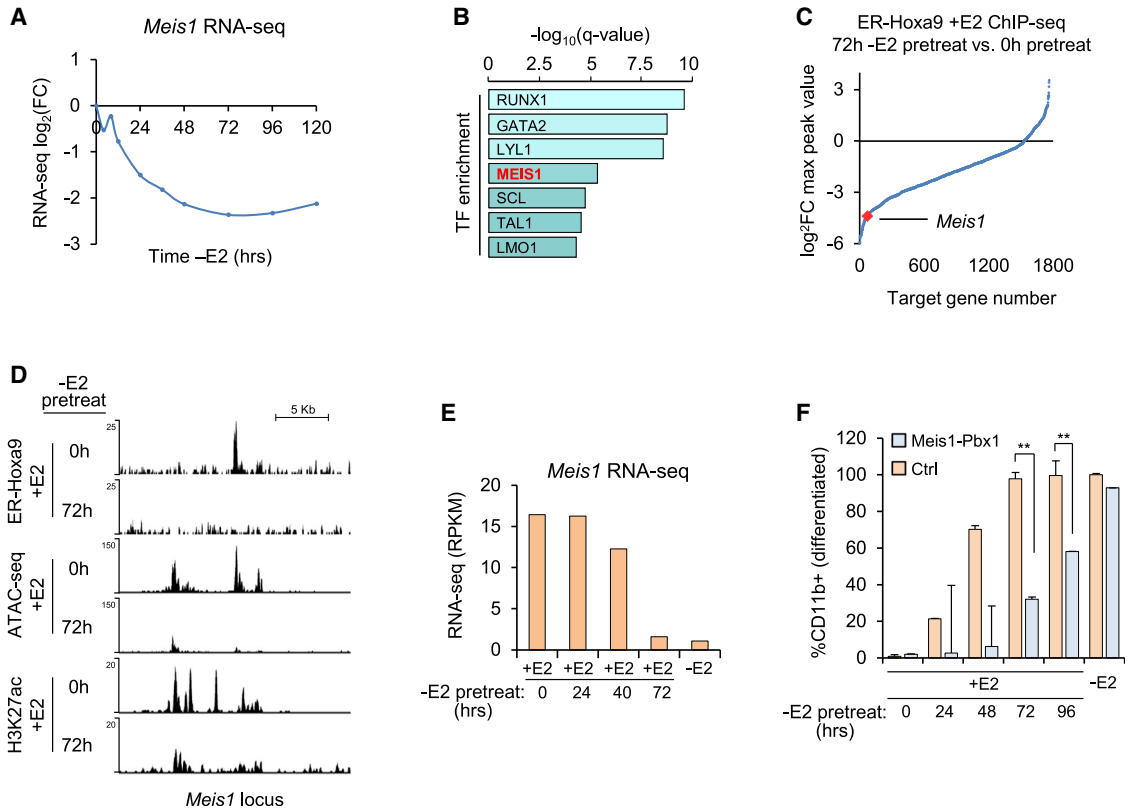


Figure 6. Re-expression of *Meis1* co-factor TF extends the pre-commitment window

(A) Downregulation of *Meis1* over the –E2 differentiation time-course.

(B) Top ChIP-seq-based TF target enrichments in the set of genes with the most reduced V5-ER-HOXA9-AM binding in cells cultured with versus without 72-h –E2 pretreatment.

(C) Plot of genes with proximal V5-ER-HOXA9-AM peaks, ordered by $\log_2$(fold-change) of maximum ChIP-seq peak value in cells cultured with versus without 72-h –E2 pretreatment.

(D) Tracks from V5-ER-HOXA9-AM ChIP-seq, H3K27ac ChIP-seq, and ATAC-seq of cells cultured with or without 72-h –E2 pretreatment.

(E) Expression of *Meis1* in cells cultured +E2, 120 h –E2, or +E2 after –E2 pretreatments.

(F) Live-cell CD11b+ percentages of control or Meis1/Pbx1-overexpressing ER-Hoxb8 cells on day 7 of commitment point assay. Cells were cultured +E2 for 7 days, –E2 for 7 days, or –E2 for the indicated durations followed by +E2 for the remainder of the experiment. *$p < 0.05$; **$p < 0.01$; t test. Graph indicates average ± standard deviation from experiments performed in experimental triplicate.

2020) (neutrophil data were not available) as comparators to the ER-HOXA9 cells in the progenitor and fully differentiated states. We assessed whether the *in vivo* progression from immature GMP to mature monocyte featured the same hallmark dramatic global reduction in chromatin accessibility as proposed in our commitment point model.

Consistent with our data, 130,325 ATAC-seq peaks are lost/reduced in the transition from human GMPs to monocytes, while only 25,758 peaks (~5-fold fewer) are gained/increased (Figure S7A). Interrogating this program in a parallel dataset of sorted murine hematopoietic cells, 141,465 peaks are lost/reduced in monocytes compared to GMPs, while only 35,432 peaks (~4-fold fewer) are gained/increased (Figure S7A). We also investigated whether, as in ER-HOXA9 cells, regions of chromatin that are open in differentiated cells are already open in the progenitor state. In line with our model, the majority of accessible chromatin regions in mouse monocytes are already accessible in GMPs (73,455 of 124,901 [59%]; Figure S7B).

In considering the distribution of peaks during myeloid differentiation, as ER-HOXA9 cells differentiate, the percentage of promoter ATAC-seq peaks increases, while the percentage of intron and intergenic peaks decreases (Figure 1E). Similarly, in human GMPs, 93% of ATAC-seq peaks that are lost/reduced in monocytes map to introns and intergenic regions, while only 2.6% map to promoters. Among peaks that are gained/increased in monocytes, only 60% map to introns and intergenic regions, while a striking 32% map to promoters (Figure S7C; Table S7).

We next compared putative HOXA9 targets and their transcriptional trends in the human datasets. We defined putative direct activation targets of ER-HOXA9 as genes that (1) have a proximal ER-HOXA9 ChIP-seq peak in the progenitor state and (2) are significantly downregulated within the first 48 h of the differentiation program, yielding 216 genes (Figure 4E). As the human datasets lack HOXA9 ChIP-seq, we defined candidate direct targets of HOXA9 as genes that (1) had a HOXA9

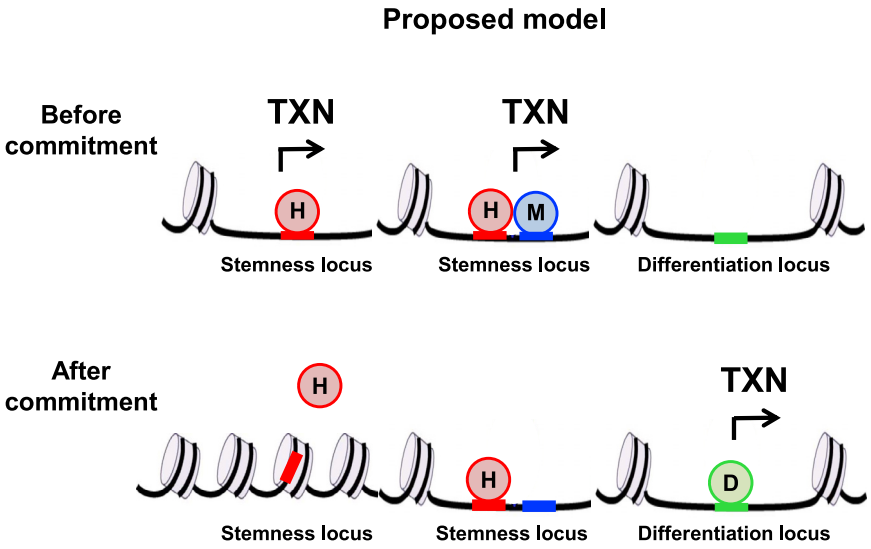


Figure 7. Model for molecular basis of point of no return

Chromatin-focused model of TF expression, chromatin accessibility, and cell identity fate in the pre-commitment (top) and post-commitment (bottom) states.

consensus binding site within a GMP ATAC-seq peak that was lost/reduced in monocytes and (2) were significantly transcriptionally downregulated in monocytes compared to GMPs, yielding 1,258 genes. Despite the differences in species (mouse versus human), differentiated cell types (neutrophils versus monocytes), datasets (with or without HOXA9 ChIP-seq), and contexts (*in vitro* versus *in vivo*), we observed a statistically significant overlap in the direct activation targets of HOXA9 in these two settings. Of the 216 direct activation targets in ER-HOXA9 cells, 31—roughly 3-fold more than expected by chance—were conserved in human GMPs ($p = 7.04 \times 10^{-7}$, hypergeometric distribution). Among conserved targets were several genes expected to be critical drivers of the ER-HOXA9 progenitor state, including *Flt3* and *Runx2* (Figure S7D).

Taken together, these data suggest that salient epigenomic features of our ER-HOXA9-based model of differentiation commitment are conserved *in vivo* in human hematopoietic cells.

DISCUSSION

Models of cell fate commitment

Lineage choice and cellular differentiation programs proceed unidirectionally during development and in the maintenance of adult tissue. This process has been conceptualized in "Waddington's Landscape" as a progressive loss of pluripotency accompanying each cell identity decision (Waddington, 1957). The molecular underpinnings of cell identity barriers remain understudied. Repressive, self-perpetuating chromatin modifications such as H3K27me3, H3K9me3, and DNA methylation are proposed to establish these barriers by facilitating irreversible silencing of stem or progenitor cell expression programs during differentiation. However, these models are typically derived from correlative studies in embryonic stem cells (Gifford et al., 2013; Suelves et al., 2016; Xie et al., 2013), and studies have not yet experimentally tested which of these chromatin features play a causal role in preventing de-differentiation.

Identification of the "point of no return" in the myeloid differentiation program

Here, we addressed the question of cell fate irreversibility using a model in which conditional activity of ER-HOXA9 governs the differentiation of GMP cells. GMPs with active ER-HOXA9 remain in an undifferentiated and self-renewing state; upon inactivation of ER-HOXA9, cells initiate their myeloid differentiation program to terminally differentiated neutrophils. By inactivating and then reactivating ER-HOXA9, we mapped a window of commitment, prior to which the cells can revert to the self-renewing progenitor state and beyond which they are destined to terminal differentiation.

While this program proceeds rapidly, the early stages are reversible. However, after ~72 h, reactivation of ER-HOXA9 fails to revert cells to their progenitor state. Surprisingly, dynamics in promoter DNA methylation or H3K27me3 did not contribute to this cell identity commitment. Rather, we found that chromatin remodeling, loss of active regulatory sites, and co-factor silencing collectively form an epigenetic barrier to prevent ER-HOXA9 from re-establishing the progenitor-state transcriptome after the commitment point. This contrasts with proposed models of ESC differentiation commitment, in which repressive histone modifications and DNA methylation can irreversibly silence pluripotency factors such as *Oct-3/4* and prevent dedifferentiation (Feldman et al., 2006; Nicetto and Zaret, 2019).

Our study focused on the choice of myeloid progenitor commitment from a self-renewing cell to a terminally differentiated effector cell. Here, chromatin accessibility and enhancer dynamics were sufficient to define a "point of no return." After inactivating ER-HOXA9, many of its target enhancers are also inactivated as chromatin accessibility decreases. Lacking active enhancers, the reactivated ER-HOXA9 protein—which is not thought to have pioneering factor activity at most of its binding targets (Choe et al., 2014; Porcelli et al., 2019)—is unable to reactivate the majority of its original target genes. While enhancers are known to define cell identity, our work highlights the role of enhancer accessibility in preventing de-differentiation. This model does not preclude a role for DNA methylation, H3K27me3, or other chromatin features in differentiation commitment in other contexts, especially in early embryonic development.

It is not clear why chromatin and enhancer accessibility are so dramatically reduced as the ER-HOXA9 cells differentiate. Epigenomic features can be "passively" lost via mitosis (Margueron et al., 2009). However, ER-HOXA9 cells exit the cell cycle within 72 h of differentiation (Sykes et al., 2016), and most losses in

chromatin accessibility occur beyond this point (Figure 1D). A likely possibility is that the open chromatin at enhancers and promoters closes by default once transcription at the locus ceases. This would be consistent with models holding that the default state of chromatin is nucleosomal and that regulatory loci are closed unless actively kept open (Becker and Workman, 2013; Escobar et al., 2019).

Addressing heterogeneity

As we followed the ER-HOXA9 cells during myeloid differentiation, we observed subtle cellular heterogeneity in the time required to reach the differentiation commitment point. One likely variable is the cell-cycle status, as terminal differentiation programs are intimately linked to the cell cycle (Kueh et al., 2013), and it is possible that the "point of no return" occurs after a fixed number of cell cycles. In this case, the relatively broad commitment window (40–72 h –E2) may be tightened in future experiments with synchronization of the cell cycle.

It is also possible that there is significant progenitor cell state heterogeneity in the starting population, as is proposed in pluripotent cells in which cycling through sub-states may enable more rapid differentiation responses. (Chambers et al., 2007; Rodriguez-Terrones et al., 2018). Whether GMPs cycle through analogous sub-states is an open question, and single-cell transcriptomic and epigenomic approaches may address this question.

Relevance to AML

AML is characterized by hallmark differentiation arrest that maintains the leukemic blasts in a proliferative, self-renewing, progenitor-like state. "Differentiation therapy" with all-*trans* retinoic acid (ATRA) and arsenic trioxide offers curative therapy to patients with the promyelocytic AML subtype (Huang et al., 1987). Inhibition of mutant IDH2 in another small subset of AML with IDH mutations may also induce differentiation (Losman et al., 2013). Unfortunately, differentiation therapy has not been available for patients of other AML subtypes.

In addition to its use in studies of normal myeloid differentiation, the ER-HOXA9 model system effectively models the differentiation block of AML. *Hoxa9* is upregulated in the majority (70%) of AMLs (Golub et al., 1999) and is considered a critical driver of oncogenic differentiation arrest (Collins and Hess, 2016). By delineating the mechanisms of differentiation irreversibility, our work helps to define the molecular processes underlying leukemic differentiation arrest and to identify those processes that may be amenable to therapeutic targeting in the development of new AML differentiation therapies.

Our model predicts that factors regulating chromatin remodeling and accessibility represent key vulnerabilities that AML cells need to overcome in order to avoid irreversible differentiation. Identification of such factors will be important to developing AML therapeutic strategies, and the ER-HOXA9 model of conditional differentiation arrest offers a tractable discovery (Mercier et al., 2016; Wickham et al., 2019) system. Targeting differentiation-suppressive proteins or associated complex members may identify new approaches to differentiation therapy. Future work will determine the generalizability of this endeavor as well as the genetic and mutational backgrounds in which it might be most effective.

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 SUBJECT DETAILS
  - Cell culture
- METHOD DETAILS
  - Establishing variants of ER-Hoxa9 cell lines
  - Flow cytometry
  - ChIP-seq
  - ChIP-seq data analysis
  - ATAC-seq
  - ATAC-seq data analysis
  - ATAC-seq motif enrichment analysis
  - ATAC-seq time course clustering analysis
  - RNA-seq
  - Gene set enrichment analysis
  - Reduced representation bisulfite seq (RRBS)
  - RNA extraction
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109967.

ACKNOWLEDGMENTS

M.A.B. was supported by a Special Fellow Award from the Leukemia & Lymphoma Society (3353-15) and an NCI K22 award. D.B.S. was supported by a Scholar Award from the American Society of Hematology and an NCI K08 award. L.G. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), EXC 2026 and Cardio-Pulmonary Institute (CPI), Project ID 390649896. S.C. and K.H. were supported by the NIH grant 5R01HD058013-10, and S.C. is currently supported by the City of Hope/UCR biomedical research initiative (CUBRI) and UC Cancer Research Coordinating Committee (CRCC) seed grants. A.M. was supported by the NIH grant 5P01GM099117 and the Max Planck Society. This work was also supported by grants to Y.S. from the Samuel Waxman Cancer Research Foundation, funds from Boston Children's Hospital, and a grant to Y.S. and D.T.S. from the Harvard Epigenetics Initiative. This project was also partially supported by funds from the Ludwig Institute for Cancer Research to Y.S. Y.S. is an American Cancer Society Research Professor.

AUTHOR CONTRIBUTIONS

M.A.B. and D.B.S. conceived and performed experiments, performed ChIP-seq and RNA-seq data analysis, interpreted all data sets, and wrote the manuscript. L.G. analyzed all ATAC-seq data and clustering of RNA-seq data and assisted with ChIP-seq and RNA-seq bioinformatics analyses. M.Wu worked with L.G. in bioinformatics analyses. M.Wawer contributed to RNA-seq analyses. S.C. performed the ATAC-seq experiments, intellectually contributed to overall study design, and assisted with data interpretation and manuscript writing. R.P. and W.D.J. assisted with ChIP-seq experiments and library preparations. J.R. and H.L. performed differentiation assays in additional ER-Hox cell systems. R.K. performed the RRBS methylation experiments and data analysis. A.J. assisted with study design and data interpretation. A.M. assisted

with RRBS data analysis and interpretation. K.H. assisted with ATAC-seq data analysis and interpretation. Y.S. and D.T.S. helped with study design, data interpretation, and manuscript writing.

DECLARATION OF INTERESTS

D.B.S. is a co-founder and holds equity in Clear Creek Bio and SAFI Biosolutions and is a consultant for Keros Therapeutics. D.T.S. is a co-founder and equity holder in Fate Therapeutics, Clear Creek Bio, and LifeVault Bio; he is a director, co-founder, and equity holder in Magenta Therapeutics; a director and equity holder in Agios Pharmaceuticals and Editas Medicine; a consultant for FOG Pharma and VCanBio; a DSMB member for Alexion; and a sponsored research recipient from Novartis. Y.S. is a co-founder and equity holder of K36 Therapeutics, a consultant for Active Motif, and holds equity in Imago Biosciences.

Received: December 30, 2020
Revised: May 12, 2021
Accepted: October 19, 2021
Published: November 9, 2021

REFERENCES

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., and Wishart, D.S. (2016). Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res. *44*, W147-53.

Becker, P.B., and Workman, J.L. (2013). Nucleosome remodeling and epigenetics. Cold Spring Harb. Perspect. Biol. *5*, a017905.

Boyle, P., Clement, K., Gu, H., Smith, Z.D., Ziller, M., Fostel, J.L., Holmes, L., Meldrim, J., Kelley, F., Gnirke, A., and Meissner, A. (2012). Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. Genome Biol. *13*, R92.

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.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234.

Cheloufi, S., Elling, U., Hopfgartner, B., Jung, Y.L., Murn, J., Ninova, M., Hubmann, M., Badeaux, A.I., Euong Ang, C., Tenen, D., et al. (2015). The histone chaperone CAF-1 safeguards somatic cell identity. Nature *528*, 218–224.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Choe, S.-K., Ladam, F., and Sagerström, C.G. (2014). TALE factors poise promoters for activation by Hox proteins. Dev. Cell *28*, 203–211.

Collins, C.T., and Hess, J.L. (2016). Deregulation of the HOXA9/MEIS1 axis in acute leukemia. Curr. Opin. Hematol. *23*, 354–361.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. *48*, 1193–1203.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Doulatov, S., Notta, F., Laurenti, E., and Dick, J.E. (2012). Hematopoiesis: a human perspective. Cell Stem Cell *10*, 120–136.

Escobar, T.M., Oksuz, O., Saldaña-Meyer, R., Descostes, N., Bonasio, R., and Reinberg, D. (2019). Active and Repressed Chromatin Domains Exhibit Distinct Nucleosome Segregation during DNA Replication. Cell *179*, 953–963.e11.

Faust, N., Varas, F., Kelly, L.M., Heck, S., and Graf, T. (2000). Insertion of enhanced green fluorescent protein into the lysozyme gene creates mice with green fluorescent granulocytes and macrophages. Blood *96*, 719–726.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006). G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. Nat. Cell Biol. *8*, 188–194.

Futschik, M.E., and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. J. Bioinform. Comput. Biol. *3*, 965–988.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell *153*, 1149–1163.

Goldberg, L., Tijssen, M.R., Birger, Y., Hannah, R.L., Kinston, S.J., Schütte, J., Beck, D., Knezevic, K., Schiby, G., Jacob-Hirsch, J., et al. (2013). Genome-scale expression and transcription factor binding profiles reveal therapeutic targets in transgenic ERG myeloid leukemia. Blood *122*, 2694–2703.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science *286*, 531–537.

Gunne-Braden, A., Sullivan, A., Gharibi, B., Sheriff, R.S.M., Maity, A., Wang, Y.-F., Edwards, A., Jiang, M., Howell, M., Goldstone, R., et al. (2020). GATA3 Mediates a Fast, Irreversible Commitment to BMP4-Driven Differentiation in Human Embryonic Stem Cells. Cell Stem Cell *26*, 693–706.e9.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. *10*, 1291–1300.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Horikoshi, M., and Tang, Y. (2018). ggfortify: Data Visualization Tools for Statistical Analysis Results (R Foundation).

Huang, M.E., Ye, Y.C., Chen, S.R., Zhao, J.C., Gu, L.J., Cai, J.R., Zhao, L., Xie, J.X., Shen, Z.X., and Wang, Z.Y. (1987). All-trans retinoic acid with or without low dose cytosine arabinoside in acute promyelocytic leukemia. Report of 6 cases. Chin. Med. J. (Engl.) *100*, 949–953.

Huang, Y., Sitwala, K., Bronstein, J., Sanders, D., Dandekar, M., Collins, C., Robertson, G., MacDonald, J., Cezard, T., Bilenky, M., et al. (2012). Identification and characterization of Hoxa9 binding sites in hematopoietic cells. Blood *119*, 388–398.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Kueh, H.Y., Champhekar, A., Nutt, S.L., Elowitz, M.B., and Rothenberg, E.V. (2013). Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. Science *341*, 670–673.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90-7.

Lal, A., Chiang, Z.D., Yakovenko, N., Duarte, F.M., Israeli, J., and Buenrostro, J.D. (2021). Deep learning-based enhancement of epigenomics data with AtacWorks. Nat. Commun. *12*, 1507.

Lam, K., Muselman, A., Du, R., Harada, Y., Scholl, A.G., Yan, M., Matsuura, S., Weng, S., Harada, H., and Zhang, D.-E. (2014). Hmga2 is a direct target gene of RUNX1 and regulates expansion of myeloid progenitors in mice. Blood *124*, 2203–2212.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lawrence, H.J., Helgason, C.D., Sauvageau, G., Fong, S., Izon, D.J., Humphries, R.K., and Largman, C. (1997). Mice bearing a targeted interruption of the homeobox gene HOXA9 have defects in myeloid, erythroid, and lymphoid hematopoiesis. Blood *89*, 1922–1930.

Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. *18*, 1851–1858.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425.

Losman, J.-A., Looper, R.E., Koivunen, P., Lee, S., Schneider, R.K., McMahon, C., Cowley, G.S., Root, D.E., Ebert, B.L., and Kaelin, W.G.J., Jr. (2013). (R)-2-hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible. Science *339*, 1621–1625.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. *40*, 4288–4297.

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.

McLeay, R.C., and Bailey, T.L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics *11*, 165.

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. *33*, 5868–5877.

Mercier, F.E., Sykes, D.B., and Scadden, D.T. (2016). Single Targeted Exon Mutation Creates a True Congenic Mouse for Competitive Hematopoietic Stem Cell Transplantation: The C57BL/6-CD45.1(STEM) Mouse. Stem Cell Reports *6*, 985–992.

Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273.

Murata, K., Jadhav, U., Madha, S., van Es, J., Dean, J., Cavazza, A., Wucherpfennig, K., Michor, F., Clevers, H., and Shivdasani, R.A. (2020). Ascl2-Dependent Cell Dedifferentiation Drives Regeneration of Ablated Intestinal Stem Cells. Cell Stem Cell *26*, 377–390.e6.

Nicetto, D., and Zaret, K.S. (2019). Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Curr. Opin. Genet. Dev. *55*, 1–10.

Nichols, J., and Smith, A. (2012). Pluripotency in the embryo and in culture. Cold Spring Harb. Perspect. Biol. *4*, a008128.

Porcelli, D., Fischer, B., Russell, S., and White, R. (2019). Chromatin accessibility plays a key role in selective targeting of Hox proteins. Genome Biol. *20*, 115.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

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*, W187-91.

R. C. Team (2005). R: A Language and Environment for Statistical Computing. In R Foundation for Statistical Computing (R Foundation).

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rodriguez-Terrones, D., Gaume, X., Ishiuchi, T., Weiss, A., Kopp, A., Kruse, K., Penning, A., Vaquerizas, J.M., Brino, L., and Torres-Padilla, M.-E. (2018). A molecular roadmap for the emergence of early-embryonic-like cells in culture. Nat. Genet. *50*, 106–119.

Schwitalla, S., Fingerle, A.A., Cammareri, P., Nebelsiek, T., Göktuna, S.I., Ziegler, P.K., Canli, O., Heijmans, J., Huels, D.J., Moreaux, G., et al. (2013). Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. Cell *152*, 25–38.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908–912.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994–1004.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555–568.

Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA *100*, 9440–9445.

Stricker, S.H., Köferle, A., and Beck, S. (2017). From profiles to function in epigenomics. Nat. Rev. Genet. *18*, 51–66.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Suelves, M., Carrió, E., Núñez-Álvarez, Y., and Peinado, M.A. (2016). DNA methylation dynamics in cellular commitment and differentiation. Brief. Funct. Genomics *15*, 443–453.

Sun, Y., Zhou, B., Mao, F., Xu, J., Miao, H., Zou, Z., Phuc Khoa, L.T., Jang, Y., Cai, S., Witkin, M., et al. (2018). HOXA9 Reprograms the Enhancer Landscape to Promote Leukemogenesis. Cancer Cell *34*, 643–658.e5.

Sykes, D.B., Kfoury, Y.S., Mercier, F.E., Wawer, M.J., Law, J.M., Haynes, M.K., Lewis, T.A., Schajnovitz, A., Jain, E., Lee, D., et al. (2016). Inhibition of Dihydroorotate Dehydrogenase Overcomes Differentiation Blockade in Acute Myeloid Leukemia. Cell *167*, 171–186.e15.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Tang, Y., Horikoshi, M., and Li, W. (2016). ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. R J. *8*, 478–489.

Tata, P.R., Mou, H., Pardo-Saganta, A., Zhao, R., Prabhu, M., Law, B.M., Vinarsky, V., Cho, J.L., Breton, S., Sahay, A., et al. (2013). Dedifferentiation of committed epithelial cells into stem cells in vivo. Nature *503*, 218–223.

Tetteh, P.W., Basak, O., Farin, H.F., Wiebrands, K., Kretzschmar, K., Begthel, H., van den Born, M., Korving, J., de Sauvage, F., van Es, J.H., et al. (2016). Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. Cell Stem Cell *18*, 203–213.

Waddington, C.H. (1957). The Strategy of the Genes (George Allen & Unwin).

Wang, G.G., Calvo, K.R., Pasillas, M.P., Sykes, D.B., Häcker, H., and Kamps, M.P. (2006). Quantitative production of macrophages or neutrophils ex vivo using conditional Hoxb8. Nat. Methods *3*, 287–293.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. J. Open Source Software *4*, 1686.

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.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell *153*, 1134–1148.

Yu, H., and Gerstein, M. (2006). Genomic analysis of the hierarchical structure of regulatory networks. Proc. Natl. Acad. Sci. USA *103*, 14724–14731.

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383.

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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD11b PE (M1/70)	Biolegend	101207; RRID:AB_11159243
Anti-Trimethyl Histone H3 (Lys4)	Millipore	CMA304; RRID:AB_1977251
Anti-trimethyl-Histone H3 (Lys27)	Millipore	07-449; RRID:AB_310624
Anti-Histone H3K27ac	Active Motif	39-135; RRID:AB_2614979
Anti-V5 tag	Abcam	ab15828; RRID:AB_443253
AbFlex anti-AM tag	Active Motif	91111; RRID:AB_2793779
Chemicals, peptides, and recombinant proteins		
IL-3	MedChemExpress	HY-102060
IL-6	MedChemExpress	HY-102058
Stem Cell Factor (SCF)	Cayman	10009929
Fetal Bovine Serum (FBS), USDA tested	EMSCO/Fisher	SH3091003
Formaldehyde	Thermo Fisher	BPF79-1
Human Plasma Fibronectin	GIBCO	33016015
Protein G dynabeads	Thermo Fisher	10004D-INV
RNase A	Fisher	EN0531
Proteinase K	Fisher	BP1700100
β-estradiol (E2)	Sigma	E2758
Polybrene	Millipore-Sigma	TR-1003-G
Critical commercial assays		
Ficoll-Paque-Plus gradient	Sigma	GE17-1440-02
NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina	New England Biolabs	E6240
Agencourt AMPure XP magnetic beads	Beckman Coulter	A63880
Tagment DNA Buffer	Illumina	15027866
Tagment DNA Enzyme 1 (TDE1)	Illumina	15027865
RNeasy kit	QIAGEN	74104
Deposited data		
ATAC-seq data from 10-sample differentiation ER-Hoxa9 differentiation time course	N/A	GEO: GSE178863
ATAC-seq profiling of ER-Hoxa9 cells before and after inducing irreversible commitment to myeloid differentiation	N/A	GEO: GSE178392
Reduced representation bisulfite sequencing (RRBS) profiling of ER-Hoxa9 cells before and after inducing irreversible commitment to myeloid differentiation	N/A	GEO: GSE183536
RNA-seq profiling of ER-Hoxa9 cells before and after inducing irreversible commitment to myeloid differentiation	N/A	GEO: GSE178866
ChIP-seq profiling of ER-Hoxa9 cells before and after inducing irreversible commitment to myeloid differentiation	N/A	GEO: GSE178433
ATAC-seq data from *in vivo* mouse GMPs and monocytes	Xiang et al., 2020	GEO: GSE143270

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ATAC-seq data from *in vivo* human GMPs and monocytes	Corces et al., 2016	GEO: GSE74912
ATAC-seq data from *in vivo* human GMPs and monocytes	Corces et al., 2016	N/A
RNA-seq data from *in vivo* human GMPs and monocytes	Corces et al., 2016	GEO: GSE7424
Experimental models: Cell lines		
ER-HOXA9-Lys-GFP	Sykes et al., 2016	N/A
ER-HOXB8-Lys-GFP	Wang et al., 2006	N/A
V5-ER-HOXA9-AM	This study	N/A
V5-HOXA9-AM	This study	N/A
CHO-SCF	Wang et al., 2006	
293T	Yang Shi's laboratory	N/A
Experimental models: Organisms/strains		
CD45.1STEM mice, female	Mercier et al., 2016	N/A
C57BL/6 mice with *lys-EGFP* knock in.	Faust et al., 2000	N/A
Recombinant DNA		
MSCVneo-EE-ER-HOXA9	Sykes et al., 2016	N/A
MSCVneo-HA-ER-HOXB8	Wang et al., 2006	N/A
MSCVdeltaNEO-V5-HOXA9-AM	This paper	N/A
MSCVdeltaNEO-V5-ER-HOXA9-AM	This paper	N/A
MSCVdeltaNEO-ER-HOXB8-MEIS1-PBX1	This paper	N/A
Software and algorithms		
BWA v0.7.17	Li et al., 2008	https://github.com/lh3/bwa
Deeptools v3.0.2	Ramírez et al., 2014	https://deeptools.readthedocs.io/en/develop/
Python v2.7.12	N/A	https://www.python.org/
Bedtools v2.26.0	Quinlan and Hall. 2010	https://bedtools.readthedocs.io/en/latest/
R v4.1.1, v3.2.1	R. C. Team, 2005	https://www.R-project.org/
MACS v2	Zhang et al., 2008	https://github.com/macs3-project/MACS
GREAT v4.0	McLean et al., 2010	http://great.stanford.edu/public/html/
EnrichR	Chen et al., 2013; Kuleshov et al., 2016	https://maayanlab.cloud/Enrichr/
MEME v4.12.0	McLeay and Bailey, 2010	https://meme-suite.org/meme/
UCSC Genome Browser	Kent et al., 2002	https://genome.ucsc.edu/
Tidyverse v1.3.1	Wickham et al., 2019	https://www.tidyverse.org/
HOMER v4.11	Heinz et al., 2010	http://homer.ucsd.edu/homer/
STAR v2.7.9a	Dobin et al., 2013	https://github.com/alexdobin/STAR/releases
GSEA v2.0	Mootha et al., 2003; Subramanian et al., 2005	https://www.gsea-msigdb.org/gsea/index.jsp
MAQ v0.7.1	Li et al., 2008	https://sourceforge.net/projects/maq/files/maq/
ChIPseeker v1.28.3	Yu et al., 2015	https://guangchuangyu.github.io/software/ChIPseeker
Bowtie v2.0	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Heatmapper	Babicki et al., 2016	http://heatmapper.ca
edgeR v3.34.1	McCarthy et al., 2012, Robinson et al., 2010	http://bioinf.wehi.edu.au/edgeR https://bioconductor.org/packages/release/bioc/html/edgeR
DESeq v1.8.1	Love et al., 2014	http://www.huber.embl.de/users/anders/DESeq
qvalue v2.24.0	Storey and Tibshirani, 2003	https://github.com/jdstorey/qvalue
FlowJo v9	N/A	https://www.flowjo.com/

RESOURCE AVAILABILITY

Lead contact

Request for more information about this manuscript and any reagents or codes used therein will be fulfilled upon request to the Lead Contact, M. Andrés Blanco (ablanco@vet.upenn.edu).

Materials availability

Materials generated in this study will be provided upon request.

Data and code availability

- All ChIP-seq, ATAC-seq, RNA-seq, and RRBS (Reduced Representation Bisulfite Sequencing) have been deposited at the GEO and are publicly available as the date of publication. Accession numbers are listed in the Key Resources Table. This paper also analyzes existing, publicly available data. The accession numbers for the datasets are listed in the Key Resources Table. All flow cytometry data reported in this paper will be shared by the lead contact upon request.
- The published article does not report novel code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

The ER-HOXA9 and ER-HOXB8 cell lines have been previously described in detail(Sykes et al., 2016; Wang et al., 2006). In brief, female murine bone marrow was harvested by crushing the femur and tibia bones. The cells were filtered through a 40-micron filter and layered over a Ficoll-Paque-Plus gradient [Sigma] to collect live mononuclear cells. Cells were cultured for 48 hours in RPMI supplemented with 10% fetal bovine serum, penicillin/streptomycin, SCF (10 ng/ml), IL-3 (10 ng/ml) and IL-6 (10 ng/ml). Tissue culture-treated 12-well plates were pre-coated with human plasma fibronectin (GIBCO, stock 10 mg/ml, use at 1 mL per well, overnight at 37°C, at a final concentration of 10 μg/ml in phosphate buffered saline (PBS)). Fibronectin was aspirated prior to the addition of cells. At ~48 hours, 2.5×10^5 cells in a volume of 500 μl were transferred to each well. Polybrene (Millipore-Sigma, stock of 10 mg/ml) was added to a final concentration of 8 μg/ml and the cells were transduced via spinoculation (1000 g, 90 min, 22-degrees) with 1 mL of ecotropic retrovirus. The viruses used in this manuscript include: MSCVneo-EE-ER-HOXA9; the EE denoting a GLU-GLU epitope tag, MSCVneo-HA-ER-HOXB8; the HA denoting a hemagluttinin epitope tag, and MSCVdNEO-V5-ER-HOXA9-AM; the dNEO denoting removal of the NEO cassette, V5 and AM denoting epitope tags on the N- and C-termini, respectively. The transduction volume was 1.5 mL with a polybrene concentration of 8 μg/ml. Following the transduction, 3 mL of fresh media was added to each well to dilute the polybrene to a less toxic concentration. A half media change was performed the next morning. These GMP progenitors were maintained in RPMI supplemented with 10% fetal bovine serum, penicillin/streptomycin, and stem cell factor (SCF). The source of stem cell factor was conditioned media generated from a Chinese hamster ovary (CHO) cell line that stably secretes *SCF.* Conditioned medium was added at a final concentration of 1%–2% (depending on the batch, final concentration of SCF approximately 100 ng/ml as measured by ELISA). Beta-estradiol (abbreviated E2, Sigma, E2758) was added to a final concentration of 0.5 μM from a 10 mM stock dissolved in 100% ethanol. The media was stable for at least 4 weeks when maintained at 4°C.

To inactivate ER-HOXA9 or ER-HOXB8, the cells were washed (2X PBS) free of beta-estradiol and re-plated in the same base media without estradiol.

METHOD DETAILS

Establishing variants of ER-Hoxa9 cell lines

Bone marrow mononuclear cells were isolated from female CD45.1STEM mice [https://pubmed.ncbi.nlm.nih.gov/27185283/] by Ficoll-Paque-Plus density gradient centrifugation and transduced as described above. The V5HOXA9AM, V5-ER-HOXA9-AM and ER-HOXB8MEIS1PBX1 constructs were cloned into the backbone of MSCVdeltaNEO (MSCVneo in which the PGK promoter and neomycin resistance cassette was removed by BglII + BamH1 digestion and ligation. This was done to decrease the size of the final MSCV construct and to allow for the generation of higher titer virus.

Gene synthesis and cloning of the constructs into the MSCVdeltaNEO backbone was done by VectorBuilder and ecotropic retrovirus generated by transient transfection of 293T cells using the Lipofectamine 2000 reagent.

Flow cytometry

Antibodies were purchased from BioLegend. Cells were suspended in FACS (fluorescence activated cell sorting) buffer (PBS + 2% FBS + 1 mM EDTA) and stained for 45 minutes at 4°C in the dark. 7-AAD or propidium iodide was included as a viability dye to help identify and gate out dead cells. Flow cytometry data was collected on either a BD FACSCalibur or BD LSR2 flow cytometer and analyzed using FlowJo software.

ChIP-seq

Cells were cross-linked with 1% formaldehyde at 37°C for 10 minutes, quenched with 125 mM glycine at room temperature for 5 minutes, and homogenized via douncing in swelling buffer (25mM HEPES pH 7.9, 1.5 mM $MgCl_2$. 10 mM KCl, 0.1% NP-40, and protease and phosphatase inhibitors (5 mM NaF, 1mM PMSF, and cOmplete Mini Protease Inhibitor Cocktail (Roche 11836153001)). Nuclei were isolated via centrifugation and fragmented in via probe sonication in sonication buffer (50 mM HEPES pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS, and proteinase and phosphatase inhibitors (same is in swelling buffer)). Insoluble material was removed via centrifugation and 500 ug chromatin (histone modification ChIPs) or 30 μg (V5-ER-HOXA9-AM) ChIPs was immunoprecipitated overnight at 4°C with 5-10 μg of antibody (for histone modification ChIPs) or 4 μg (for V5-ER-HOXA9-AM) ChIPs and 50 μl protein G magnetic dynabeads. Immunoprecipitated material was washed 1X in sonication buffer, 1X in wash buffer A (50 mM HEPES, 500 mM NaCl, 1mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, and 0.1% SDS), 1X in wash buffer B (20 mM Tris pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40, and 0.5% Na-deoxycholate), 2X in TE, and was then eluted by heating twice at 65°C for 5 minutes in elution buffer (50 mM Tris pH 7.5, 1 mM EDTA, and 1% SDS). Cross-links were reversed by incubating at 65°C overnight in 160 mM NaCl and 20 μg/ml RNase A. Samples were then treated with 200 μg/ml Proteinase K for two hours at 45°C and DNA recovered by phenol-chloroform extraction. 10 ng of DNA per sample was used for next generation sequencing library preparation using an NEBNext ChIP-seq Library Prep Master Mix Set for Illumina kit (New England Biolabs) as per manufacturer's guidelines using AMPure XP beads for purification (Beckmann Coulter). 50 bp single end sequencing of DNA was performed on an Illumina HiSeq 2500 sequencer (histone modification ChIP-seq) or an Illumina NextSeq 500 sequencer (V5-ER-HOXA9-AM ChIP-seq). The following antibodies were used for chromatin immunoprecipitation: H3K4me3: Millipore CMA304; H3K27me3: Millipore 07-449; H3K27ac: Active Motif AM39135; V5: Abcam ab15828; AM: Active Motif 91111.

ChIP-seq data analysis

Reads were aligned to the genome using Bowtie 2.0 (Langmead and Salzberg, 2012). Peaks were called using MACS 2.0 (Zhang et al., 2008), with NarrowPeak setting used for H3K27ac and BroadPeak setting used for H3K27me3, and with mfold minimum set to 10 and FDR < 0.01. GO category enrichment analyses were performed with Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010) and EnrichR (Chen et al., 2013; Kuleshov et al., 2016), and motif enrichment analyses were performed with Analysis of Motif Enrichment (AME) from the MEME Suite 4.12.0 (McLeay and Bailey, 2010). Peaks were visualized using the UCSC Genome browser (Kent et al., 2002). Superenhancer peaks were called as described previously (Whyte et al., 2013). Briefly, H3K27ac peaks were first called by using MACS2 and peaks within 12.5 kb of one another were stitched together as a super enhancer. ChIP-seq heatmaps were generated using deepTools (Ramírez et al., 2014) as follows. bamCompare was used to generate paired bigwig files using 50 bp bins, computeMatrix was used (parameters:–beforeRegionStartLength 2000;–binSize 10;–sortUsing mean;–averageTypeBins mean) to generate scores for heatmaps, and heatmaps were generated with with plotHeatmap (parameters: –sortRegions descending order; –sortUsing mean; –averageTypeSummaryPlot mean). Genomic distributions of ChIP-seq peaks were generated using ChIPseeker (Yu et al., 2015) with default parameters. For analysis of V5-ER-HOXA9-AM ChIP-seq data, 75-nt single-end sequence reads were mapped to the mm10 genome using the BWA algorithm with default settings. Reads that passed Illumina's purity filter, had no more than 2 alignment mismatches, and mapped uniquely to the genome were de-duplicated and maintained for subsequent analysis using MACS for peak calling. For peak-to-gene mapping, peaks were assigned to the closest gene within a gene margin of 10,000 bp upstream to 10,000 bp downstream. If the peak was not within the margin for any genes, it was not given a gene association. Differential peak values for a given gene when comparing any two time points were defined as the log2 fold change between the fragment densities of the gene-associated peak interval at the two time points. For genes with multiple peaks within the gene margin, the peak interval with the highest fragment density was used.

ATAC-seq

To generate ATAC-seq libraries, 50,000 cells were used and libraries were constructed as previously described (Buenrostro et al., 2013; Cheloufi et al., 2015). Briefly, cells were washed in PBS twice, counted and nuclei were isolated from 100,000 cells using 100 μl hypotonic buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% NP40) to generate two independent transposition reactions. Nuclei were split in half and treated with 2.5 μl Tn5 Transposase (Illumina) for 30 mins. at 37°C. DNA from transposed nuclei was then isolated and PCR-amplified using barcoded Nextera primers (Illumina). Library QC was carried out using high sensitivity DNA biolanalyzer assay and qubit measurement and sequenced using paired-end sequencing (PE50) on the Illumina Hi-Seq 2500 platform.

ATAC-seq data analysis

Raw data were mapped to mouse reference genome version mm9 by BWA (Li et al., 2008). Potential PCR duplicates and ambiguously mapped reads were further filtered out for downstream analysis. Peaks were called by MACS 2.0 (Zhang et al., 2008) with parameters callpeak–nomodel–extsize 200–shift −100 -q 0.01. Consistency of peak calling results between replicates was examined using IDR analysis. IDR consistency threshold of 0.05 was applied to truncate the peak lists, peaks that passed the threshold were used for subsequent analysis. PCAs of ATAC-seq peak files were generated using ggplot2 (Wickham, 2009) and ggfortify (Horikoshi and Tang, 2018; Tang et al., 2016), with MultiCovBed (Quinlan and Hall, 2010) used to generate input files from peak bed files (default settings used for all parameters). For analysis of *in vivo* samples, ATAC-seq data were obtained from GSE74912 (human GMPs), direct transfer from Jason D. Buenrostro (human monocytes) and GSE143270 (mouse GMPs and monocytes). Briefly, raw fastq data were downloaded from SRA, reads passed QC were then mapped to hg19 and mm10 by BWA, and peaks were called by MACS2.

ATAC-seq motif enrichment analysis

ATAC-seq peak regions were analyzed for the occurrences of known motifs in HOMER motif databases (Heinz et al., 2010) using HOMER. -log10 (motif P value) was used for the motif enrichment heatmap plot.

ATAC-seq time course clustering analysis

To obtain a reference peak set, ATAC-seq peaks of all time points were pooled and overlapping peaks regions ($\geq$1bp) were merged into a single region which was the widest region that covers all merged regions. Raw reads were counted in the reference peaks set for each time point using Rsubread and differential analysis was performed using edgeR (McCarthy et al., 2012; Robinson et al., 2010). Regions that showed no significant changes ($|\log_2 FC| > 2$, FDR < 0.01) between any two time points were eliminated and remaining regions used for following clustering analysis. Fuzzy clustering cmeans (Futschik and Carlisle, 2005) was applied to recognize temporal patterns of the time course data.

RNA-seq

Normalized 10 time point, 120-hour RNA-seq data were published previously (Sykes et al., 2016) and are available from the gene expression omnibus (GEO) under accession number GSE84874. E2 withdrawal pretreatment and add-back RNA-seq datasets were analyzed similarly. Briefly, Reads were aligned to the mouse GRCm38 primary assembly extended with the sequence of eGFP (https://www.ebi.ac.uk/ena/browser/view/AHK23750) using STAR (Dobin et al., 2013) and the GENCODE annotation M5 (https://www.gencodegenes.org). Differential expression analysis was performed with DESeq2 version 1.8.1 (Love et al., 2014) in R version 3.2.1 (R. C. Team, 2005). Significantly differentially expressed genes, when comparing any one time point to any other time point, were defined as genes with $\log_2$ fold change > 1 and Benjamini Hochberg adjusted p value < 0.05. Principal component analysis was performed in R and plotted using the ggplot2 package version 2.2.0 (Wickham, 2009). Clustering analyses of RNA-seq datasets was performed as described for ATAC-seq clustering analyses. Heatmaps were generated either using Heatmapper (Babicki et al., 2016) using average clustering and Euclidean distance metric, or using GSEA (below). For RNA-seq analysis of human *in vivo* samples, data were obtained from GSE74246, and DESeq2 was used to identify differential expression genes.

Gene set enrichment analysis

Normalized time-course RNA-seq data were analyzed via GSEA 2.0 (Mootha et al., 2003; Subramanian et al., 2005) after condensing the dataset to one value per gene by retaining the gene reading with the highest median RPKM value across all time points. Enrichment analyses were performed using 1,000 gene set permutations, a weighted enrichment statistic, the Signal2Noise ranking metric, and gene set minimum and maximum sizes of 15 and 500, respectively. Gene sets were obtained from the Molecular Signatures Database (MSigDB) (Liberzon et al., 2015, 2011; Subramanian et al., 2005) or from ChIP-seq datasets generated in this study.

Reduced representation bisulfite seq (RRBS)

Reduced Representation Bisulfite Sequencing (RRBS) Illumina libraries were prepared according to a standard gel-free pipeline (Boyle et al., 2012).

Reads were aligned to build mm9 of the mouse genome with MAQ (Li et al., 2008) in bisulfite alignment mode. For each CpG, the methylation level was computed as the number of reads with unconverted cytosines divided by the number of total reads covering

that CpG. Mean methylation for each genomic feature was calculated as the mean methylation level of the CpGs within the feature weighted by the sequence coverage at each CpG, and only CpGs covered at 5x or higher were used in this calculation. Differentially methylation testing was performed by using methylation levels of CpGs covered at 5X or higher in a two-sample weighted coverage t test. FDR q-values were calculated from the t test p values using the R q-value package (Storey and Tibshirani, 2003). Differentially methylated features were required to have a methylation difference of 0.2 at a FDR q-value threshold of 0.05.

RNA extraction

RNA was purified using the RNeasy kit (QIAGEN) as per manufacturer's instructions. In brief, ~5 million cells were pelleted and re-suspended in RLT buffer containing beta-mercaptoethanol as recommended. The optional on-column DNase digestion protocol was performed, and the final RNA was eluted in 30 ul of RNase-free water. Quantification of RNA was done using the Nanodrop.

QUANTIFICATION AND STATISTICAL ANALYSIS

The analysis, software, and quantification methodology that are specific to ChIP-seq, ATAC-seq, RNA-seq, and RRBS experiments are included under the relevant subsections of the Methods details section. Information regarding replicate numbers is provided in figure legends. In all cases, biological replicates are defined as cells cultured in independent tissue culture plates throughout the duration of the experiment and processed independently after cell harvest. If error bars are used in figures, information about what error bars represent is also provided in the figure legend. If asterisks are used to denote statistical significance in figures, the degree of significance is provided in the figure legend, and further details regarding statistical test used are provided in the relevant subsections of the Methods details that are specific to the analysis being performed.

In general, results were considered statistically significant if the given test yielded a multiple hypothesis testing adjusted p value (ie. FDR q-value or Bonferroni-corrected p value) was < 0.05, though alternative criteria are listed if greater stringency was used. In cases where t tests were used, data distributions were first tested for normality using Shapiro-Wilk tests. If the Shapiro-Wilk test did not yield $p < 0.05$, datasets were considered normal and evaluated by t test.

In flow cytometry experiments, experiments were performed in biological triplicate. When presented via bar graph, bars represent the mean of the replicates, with error bars representing the standard deviation.