

SET1A/COMPASS and shadow enhancers in the regulation of homeotic gene expression

Kaixiang Cao,¹ Clayton K. Collings,¹ Stacy A. Marshall,¹ Marc A. Morgan,¹ Emily J. Rendleman,¹ Lu Wang,¹ Christie C. Sze,¹ Tianjiao Sun,¹ Elizabeth T. Bartom,¹ and Ali Shilatifard^{1,2}

¹Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA; ²Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA

The homeotic (Hox) genes are highly conserved in metazoans, where they are required for various processes in development, and misregulation of their expression is associated with human cancer. In the developing embryo, Hox genes are activated sequentially in time and space according to their genomic position within Hox gene clusters. Accumulating evidence implicates both enhancer elements and noncoding RNAs in controlling this spatiotemporal expression of Hox genes, but disentangling their relative contributions is challenging. Here, we identify two *cis*-regulatory elements (E1 and E2) functioning as shadow enhancers to regulate the early expression of the HoxA genes. Simultaneous deletion of these shadow enhancers in embryonic stem cells leads to impaired activation of HoxA genes upon differentiation, while knockdown of a long noncoding RNA overlapping E1 has no detectable effect on their expression. Although MLL/COMPASS (complex of proteins associated with Set1) family of histone methyltransferases is known to activate transcription of Hox genes in other contexts, we found that individual inactivation of the MLL1-4/COMPASS family members has little effect on early Hox gene activation. Instead, we demonstrate that SET1A/COMPASS is required for full transcriptional activation of multiple Hox genes but functions independently of the E1 and E2 *cis*-regulatory elements. Our results reveal multiple regulatory layers for Hox genes to fine-tune transcriptional programs essential for development.

[*Keywords:* Hox genes; enhancer; COMPASS; SET1A; chromatin structure; epigenetic marks]

Supplemental material is available for this article.

Received December 15, 2016; revised version accepted April 12, 2017.

In mammals, the 39 Hox genes, organized into four clusters on different chromosomes, encode helix–turn–helix DNA-binding transcription factors (TFs) that are critical for specifying the anterior–posterior body plan during animal development (Alexander et al. 2009; Mallo et al. 2010). Knockout experiments in mouse models indicate that Hox genes are important for the development of the hindbrain, axial skeleton, and limbs (Wellik 2007; Zakany and Duboule 2007; Alexander et al. 2009). Hox genes in each cluster are activated sequentially in both time and space in the developing embryo according to their genomic position within the cluster (Kmita and Duboule 2003). Despite being extensively studied, the mechanisms underlying the spatial and temporal colinearity of Hox genes remain mysterious.

Accumulating evidence demonstrates the impact of *cis*-regulation on the activation of Hox gene clusters.

Chromatin decondensation coincides with the activation sequences of HoxB and HoxD clusters during embryonic stem cell (ESC) differentiation and embryonic development (Chambeyron and Bickmore 2004; Chambeyron et al. 2005; Morey et al. 2007). The genome architecture at the HoxD cluster is also dynamically remodeled at different development stages (Noordermeer et al. 2011, 2014). Moreover, multiple regulatory sequences are found to be critical for the activation of Hox clusters. Retinoic acid (RA) response elements (RAREs) located at the 3' of HoxA and HoxB clusters are responsible for the activation of 3' Hox genes triggered by RA, but genetic studies and single-cell analyses suggest that additional regulatory sequences are involved (Studer et al. 1994; Popperl et al. 1995; Dupe et al. 1997; Maamar et al. 2013). Indeed, a group of regulatory regions located

Corresponding author: ash@northwestern.edu

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.294744.116>.

© 2017 Cao et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

>500 kb upstream of the HoxD cluster are necessary for the expression of 5' HoxD genes in developing mouse digits and genitals (Montavon et al. 2011; Lonfat et al. 2014). On the other hand, distal enhancers downstream from the HoxD cluster regulate the expression of 5' HoxD genes in the forelimb (Andrey et al. 2013). Furthermore, DNA elements within the Hox gene clusters, which include CTCF-binding sites, have been shown to serve as topological boundaries separating active and repressive domains during stem cell differentiation (Narendra et al. 2015). Nevertheless, distal regulatory elements required for the activation of the 3' Hox genes are still unidentified.

Epigenetic mechanisms are associated with establishing Hox gene expression patterns during embryonic development and stem cell differentiation (Soshnikova and Duboule 2009). The antagonism between trithorax group proteins (TrxG) and polycomb group (PcG) proteins is known to enforce changes in H3K4me3 and H3K27me3 levels at Hox clusters, resulting in activation or repression, respectively, during development (Ringrose and Paro 2004; Schuettengruber et al. 2007; Piunti and Shilatifard 2016). In mammals, the COMPASS (complex of proteins associated with Set1) family of six protein complexes is responsible for the implementation of the vast majority of H3K4 methylation. The MLL1 branch of the COMPASS family, which is related to *Drosophila* Trx, is required for proper expression of many genes within the HoxA and HoxC clusters in mouse embryonic fibroblasts (MEFs) (Wang et al. 2009). MLL2 depletion in ESCs leads to loss of H3K4me3 at promoters of bivalently marked genes, which includes the Hox genes, and at numerous intergenic regions bearing signatures of transcriptional enhancers (Hu et al. 2013b, 2017). Nevertheless, none of the members of the COMPASS family has been demonstrated to participate in the early activation of anterior Hox genes during ESC differentiation to date.

To understand the relative contributions of *cis*-regulatory sequences and the COMPASS family, we individually and combinatorially examined the contribution of two shadow enhancers, a long noncoding RNA (lncRNA), and members of the COMPASS family to early activation of HoxA genes during ESC differentiation. Our study unveils novel *cis*- and *trans*-regulatory mechanisms that suggest a multilayered regulation model for Hox gene transcriptional activation.

Results

Identification of enhancers functioning in transcriptional regulation of the HoxA cluster

Hox genes are silent in pluripotent mouse ESCs but could be activated rapidly through RA-induced differentiation. Expression of 3' HoxA genes occurs in <4 h upon RA induction (De Kumar et al. 2015) in ESCs, making it an ideal model for deciphering mechanisms of early Hox gene activation. To examine the effects of RA-induced transcriptional activation in an unbiased fashion, we first per-

formed RNA sequencing (RNA-seq) in undifferentiated and RA-treated ESCs at two different time points. Upon RA treatment, genes within the HoxA cluster, including the lncRNA genes *Halr1* and *Hotairm1*, were highly elevated, while the expression of *Skap2*, a neighboring gene of the HoxA cluster, remained unchanged (Fig. 1A). Enhancers harboring RAREs contribute to the activation of Hox genes (Studer et al. 1994; Frasch et al. 1995; Morrison et al. 1996); however, deletion of the *Hoxa1* proximal enhancer with RAREs significantly reduces, but does not silence, the expression of *Hoxa1* in vivo (Dupe et al. 1997), suggesting the requirement of additional *cis*-regulatory sequences. To identify novel regulatory elements of HoxA genes, we performed ChIP-seq (chromatin immunoprecipitation [ChIP] combined with high-throughput sequencing) of the enhancer-associated histone marks H3K27ac and H3K4me1 and the active promoter mark H3K4me3 in ESCs treated with RA. In addition to proximal enhancers of *Hoxa1*, two putative enhancer regions (referred to here as E1 and E2), located in the gene desert between *Skap2* and *Hoxa1*, are enriched with H3K27ac upon RA treatment (Fig. 1B). These regions are marked initially by H3K4me1 in undifferentiated ESCs, and, upon RA-induced differentiation, this histone mark spreads to form a broad domain covering the majority of the intergenic region (Fig. 1B). Additionally, moderate levels of H3K4me3 were found at E1 upon differentiation (Fig. 1B). The E1 region overlaps with the *Halr1* gene; however, unlike the H3K4me3 peak that mainly covers the transcription start site (TSS) region, the H3K27ac-coated E1 is located within the *Halr1* gene body (Supplemental Fig. S1A). These data suggest that E1 and E2 may gain enhancer activity upon RA induction and can potentially impact transcriptional activation of HoxA genes.

Looping between enhancers and promoters plays an instructive role in transcriptional activation (Deng et al. 2012, 2014). To identify regions that gain interaction with the HoxA cluster during RA-induced differentiation, we generated chromatin interaction profiles of the *Hoxa1* promoter in ESCs using circularized chromosome conformation capture (4C) combined with high-throughput sequencing (4C-seq). Silent *Hoxa1* interacts with the previously identified RARE-containing proximal enhancer (Frasch et al. 1995) and a region 20 kb downstream from *Hoxa1* in undifferentiated ESCs (Fig. 1C, left). Upon RA treatment, these aforementioned interactions were strengthened. Interestingly, the contact frequency of *Hoxa1* with E1, E2, and their surrounding regions was moderately elevated upon differentiation (Fig. 1C, left). 4C-seq with the viewpoint at E1 indicates that the interactions between E1 and the HoxA cluster are elevated upon RA induction (Fig. 1C, right). Furthermore, E1- and E2-neighboring regions had tighter contacts in RA-treated ESCs (Fig. 1C, right), suggesting a cooperation of these two regions during ESC differentiation. Together with the finding that E1 and E2 had the most prominent increase of H3K27ac upon differentiation, we hypothesized that these elements participate in transcriptional activation of HoxA genes during RA-induced differentiation.

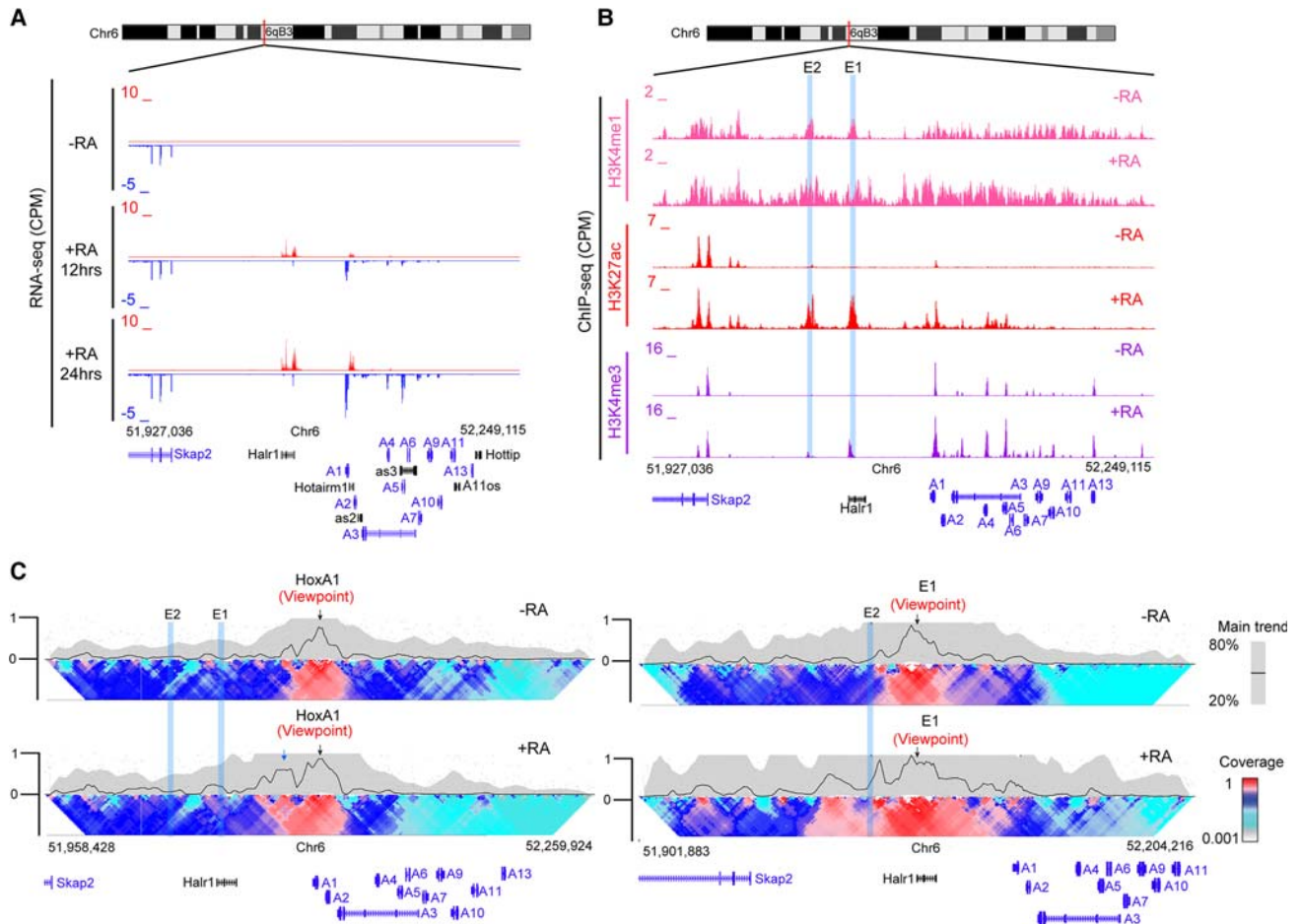


Figure 1. Identification of distal *cis*-regulatory elements at the HoxA cluster. (A) University of California at Santa Cruz (UCSC) genome browser view of RNA-seq signals of undifferentiated and RA-treated ESCs at two different time points. The HoxA gene cluster is shown. Genes on the Watson strand are labeled in black, and those on the Crick strand are marked in blue. (CPM) Counts per million mapped reads. (B) UCSC genome browser view of H3K4me1, H3K27ac, and H3K4me3 ChIP-seq tracks at the HoxA cluster in undifferentiated and RA-treated ESCs. The centers of potential distal regulatory regions are marked with blue stripes. (E1) Putative enhancer 1; (E2) putative enhancer 2. (C) 4C-seq [circularized chromosome conformation capture [4C] combined with sequencing] analysis with the *Hoxa1* promoter as the viewpoint (*left*) and E1 as the viewpoint (*right*) in undifferentiated and RA-treated ESCs. (Blue arrow) Increased contact between HoxA-proximal enhancers and the viewpoint. The median and 20th and 80th percentiles of a sliding 5-kb window determine the main trend line. The color-coded scale represents enrichment relative to the maximum median value at a resolution of 12 kb.

Deletion of enhancers E1 and E2 impairs HoxA cluster activation in response to RA

To test whether the putative E1 and E2 enhancer regions play a role in transcriptional activation of the HoxA cluster, we deleted them using clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9-guided homologous recombination (Supplemental Fig. S1A,B). We generated E1 knockout ESCs by replacing the endogenous 3.2-kb genomic DNA, including most of the first intron and the second exon of *Halr1*, with a fragment of donor DNA containing a green fluorescent protein (GFP) gene and a G418-resistant gene cassette (Supplemental Fig. S1B). After G418 selection, ESC clones were screened with Southern blotting, and a 5-kb band representing the knockout allele could be detected in the E1 knockout clones (Supplemental Fig. S1C). Out of the 240 clones

picked, 18 homozygous E1-deleted clones were identified. We then transfected two of the 18 clones with a Cre recombinase-expressing plasmid to remove the donor cassette to generate E1 knockout cells. Genotyping PCR using primers inside (wild-type primer) and outside (knockout primer) the guide RNA (gRNA) recognition sites, together with Sanger sequencing, confirmed the excision of donor sequences in the E1 knockout clones (Supplemental Fig. S1D,E). E1 knockout cells had morphology similar to that of their parental wild-type ESCs (data not shown) and comparable expression profiles (Supplemental Fig. S2A). Similarly, we deleted the 3.1-kb E2 region with two gRNAs and derived nine homozygous ESC lines out of the 96 picked clones. Deletion of the E2 region was confirmed by PCR genotyping and Sanger sequencing (Supplemental Fig. S1D,E). As DNA looping between E1 and E2 was noticeably increased upon RA-induced

differentiation (Fig. 1C), we generated E1 and E2 double-knockout ESCs by deleting the E1 region in E2 knockout ESCs using CRISPR–Cas9-guided gene editing and confirmed the successful deletion with Sanger sequencing (Supplemental Fig. S1E). As expected, the morphology of undifferentiated E2 and double-knockout cells was indistinguishable from wild-type ESCs (data not shown), and the ESC transcriptome was not significantly perturbed by these genetic manipulations, as indicated by RNA-seq analysis (Supplemental Fig. S2A).

To investigate the functions of the potential regulatory regions identified by ChIP-seq and 4C-seq, we performed RNA-seq in wild-type, E1 knockout, E2 knockout, and double-knockout ESCs treated with RA. Our results demonstrate that although individual deletion of E1 or E2 does not have a significant impact on *Hoxa1* expression upon RA induction, double deletion of these two regions dampens the activation of *Hoxa1* (Fig. 2A). Interestingly, such impairment also applied to many other HoxA genes, such as *Hoxa4*, *Hoxa5*, and *Hoxa7* (Fig. 2B). *Cyp26a1*,

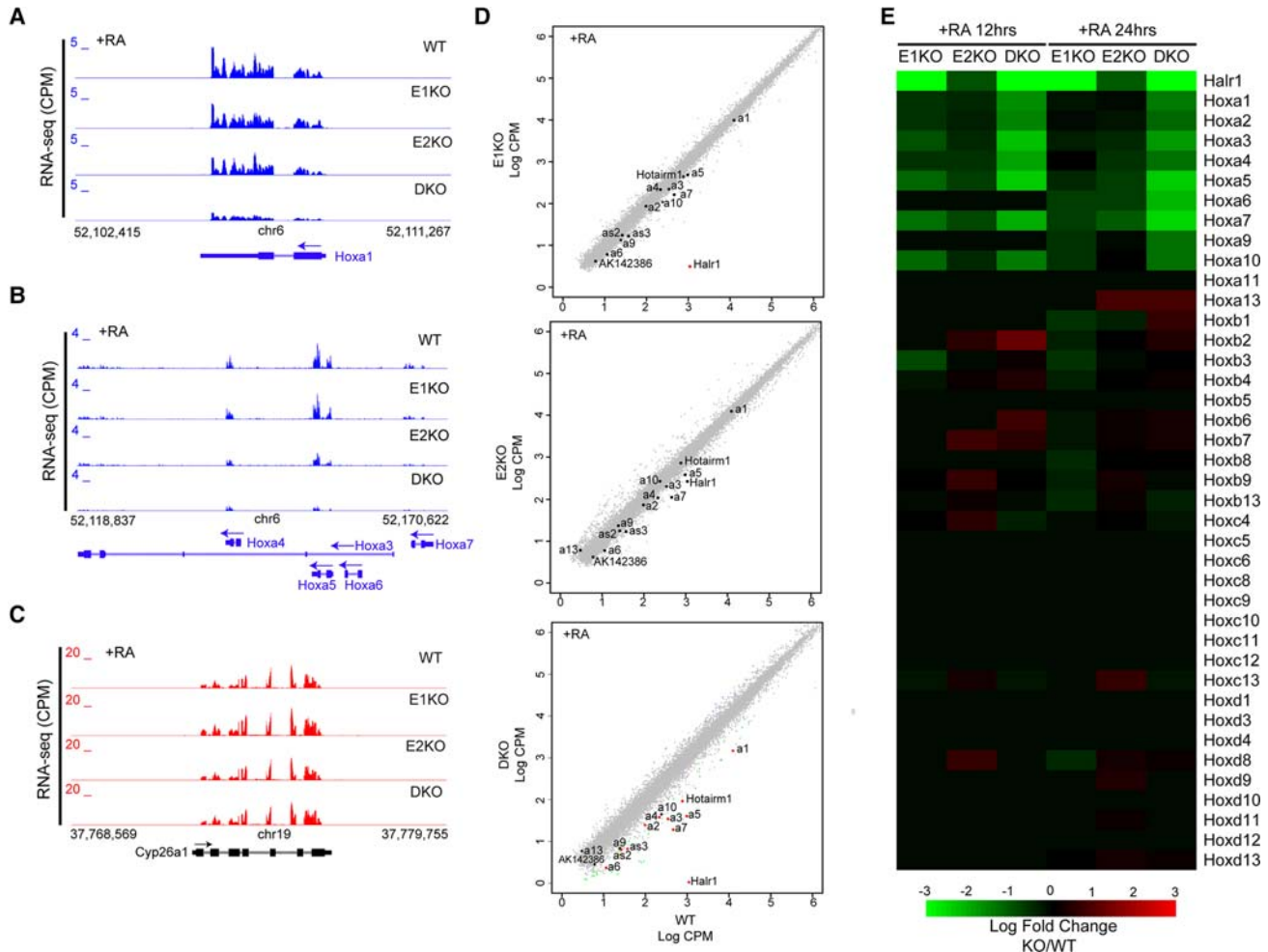


Figure 2. Redundancy of the putative enhancers on the activation of HoxA genes. (A) UCSC genome browser view of *Hoxa1* expression levels in wild-type, E1 knockout, E2 knockout, and double-knockout ESCs treated with RA. The arrow indicates transcription direction. (B) UCSC genome browser view of *Hoxa3–Hoxa7* expression levels in wild-type, E1 knockout, E2 knockout, and double-knockout ESCs treated with RA. (C) UCSC genome browser view of *Cyp26a1* expression levels in wild-type, E1 knockout, E2 knockout, and double-knockout ESCs treated with RA. (D) Correlation analysis of gene expression levels between wild-type and E1 knockout (top), E2 knockout (middle), and double-knockout (bottom) ESCs treated with RA. Plots were generated based on two biological replicates of RNA-seq experiments from two independent cell clones for each genotype. The X-axis represents the expression level (\log_2 normalized CPM) of wild-type cells, and the Y-axis represents the expression level of knockout cells. Significantly down-regulated genes (compared with wild type) are labeled in green, and up-regulated ones are labeled in purple. HoxA genes that were changed significantly (adjusted $P < 0.01$) are marked with red dots, while unchanged ones are marked with black dots. Other unchanged genes are labeled with gray dots. (E) Heat map analysis comparing the expression fold changes of the 39 HoxA genes and *Halr1* in wild-type ESCs with that in E1 knockout, E2 knockout, and double-knockout cells. The heat map was generated based on four biological replicates of RNA-seq experiments from two independent cell clones for each genotype. All values were normalized to wild-type values at the indicated time points to derive the fold changes.

one of the genes most rapidly induced by RA (Lin et al. 2011), had a comparable level across all cell lines upon RA treatment (Fig. 2C), indicating that the failure in transcriptional activation during ESC differentiation is specific to the HoxA gene cluster. RNA-seq analysis demonstrated that the majority of genes located in the HoxA cluster had significantly reduced expression levels in double-knockout ESCs compared with wild-type cells upon RA induction (Fig. 2D, bottom), suggesting that E1 and E2 together contribute to the activation of the entire HoxA gene cluster. Except for the slight and nonsignificant reduction in expression levels of a few genes such as *Hoxa5* and *Hoxa7*, other genes in the HoxA cluster remained unchanged in the E1 or E2 knockout ESCs compared with wild type (Fig. 2D, top and middle), indicating that the two putative enhancers act redundantly to activate the HoxA gene cluster.

We next examined the expression levels of the HoxB genes and found that none of them exhibited significant changes in the single-knockout or double-knockout ESCs (Supplemental Fig. S2B), confirming the specificity of E1 and E2 on activating HoxA cluster. Furthermore, heat map analysis demonstrated that the effects of knockouts were restricted to the HoxA cluster, as genes in the other three Hox gene clusters were largely unaffected (Fig. 2E). We considered that the E1 region was located within the *Halr1* gene, whose transcript has been shown to repress the expression of HoxA genes (Maamar et al. 2013; Yin et al. 2015; Liu et al. 2016). To investigate the potential antagonism between the enhancer activity of E1 and the repressive role of *Halr1*, we depleted *Halr1* by shRNA-guided knockdown to test whether the lncRNA regulates HoxA gene expression in our system. *Halr1* depletion did not affect the activation of HoxA genes despite effective knockdown of the lncRNA (Supplemental Fig. S2C). Taken together, our data suggest redundant functions of the two putative enhancers E1 and E2 on the activation of the HoxA gene cluster during ESC differentiation.

Deletion of enhancers E1 and E2 disrupts the chromatin structure of the HoxA cluster

Changes in promoter–enhancer interactions are associated with implementation of epigenetic marks (Soshnikova and Duboule 2009; Noordermeer et al. 2011). To determine the roles of E1 and E2 on chromatin dynamics at the HoxA cluster during differentiation, we performed ChIP-seq of H3K4me1 and H3K27ac in wild-type and double-knockout cells before and after RA induction. Unexpectedly, double deletion of E1 and E2 led to the loss of the broad H3K4me1 domain formed between *Skap2* and *Hoxa1* upon RA induction (Fig. 3A). In contrast, H3K4me1 occupancy was comparable across the HoxB cluster in wild-type and double-knockout cells (Supplemental Fig. S3A), indicating that the H3K4me1 loss at distal regions is specific to the HoxA cluster. These effects require cooperative activities of E1 and E2, since removal of a single enhancer does not fully disrupt the H3K4me1 domain in RA-treated ESCs (Supplemental Fig. S3B).

Although H3K4me1 levels at HoxA genes did not exhibit significant changes when comparing wild-type and double-knockout cells, H3K27ac peaks located within the cluster and in proximity to *Hoxa1* were diminished (Fig. 3B), consistent with the reduction in transcription of the entire HoxA cluster. In contrast, a decrease in H3K27ac levels was not observed at the HoxB cluster (Supplemental Fig. S3A), and H3K27ac was not strongly reduced at the HoxA cluster in the single enhancer knockout cells upon differentiation (Supplemental Fig. S3B). As seen with H3K4me1, H3K4me2 levels within the HoxA gene cluster did not exhibit significant changes in double-knockout cells (Supplemental Fig. S3C). These data indicate that enhancer regions E1 and E2 are required for establishing specific chromatin features at the HoxA cluster during RA-induced differentiation.

Three-dimensional (3D) chromatin structure is critical for the spatiotemporal activation of Hox genes during development (Montavon et al. 2011; Noordermeer et al. 2011; Andrey et al. 2013). To investigate whether the deletion of E1 and E2 affects the higher-order chromatin structure at the HoxA cluster, we generated interaction maps for the *Hoxa1* promoter by high-resolution 4C-seq in double-knockout ESCs. The interaction between *Hoxa1* and its 3' distal regions in the undifferentiated state was already diminished drastically compared with wild-type cells (Fig. 3C, cf. the two top panels). Upon RA induction, interactions between *Hoxa1* and its regulatory regions increased in wild-type cells, while such increases in double-knockout cells could be detected only at more proximal regions. Interestingly, deletion of E1 and E2 caused a reduction in interactions between *Hoxa1* and its proximal enhancers (Fig. 3C, bottom panels). Moreover, *Hoxa1* was in closer proximity to posterior HoxA genes in double-knockout cells in contrast to wild-type cells, suggesting that *Hoxa1* is under a more compact chromatin environment in double-knockout versus wild-type cells. Hi-C (chromosome capture followed by high-throughput sequencing) data (Dixon et al. 2012) showed that the HoxA cluster was located on the border of two neighboring topologically associated domains (TADs) in ESCs (Supplemental Fig. S3D). Thus, our data suggest that deletion of the distal regulatory regions of the HoxA cluster disrupts the boundary of these neighboring TADs in double-knockout cells, leading to topological compartment switching of the anterior HoxA genes. Such alterations in higher-order chromatin structure could play a pivotal role in establishing the expression pattern of the HoxA genes.

E1 and E2 are bound by multiple chromatin-modifying proteins and TFs

We next sought to identify factors that function through E1 and E2 to regulate the activation of HoxA genes in our system. Multiple lines of evidence have shown that cohesin connects enhancers and promoters and plays an important role in the formation of 3D chromatin architecture (Kagey et al. 2010; Downen et al. 2014). We thus examined whether cohesin binds E1 and E2 enhancers by browsing previously published ChIP-seq data sets (Kagey

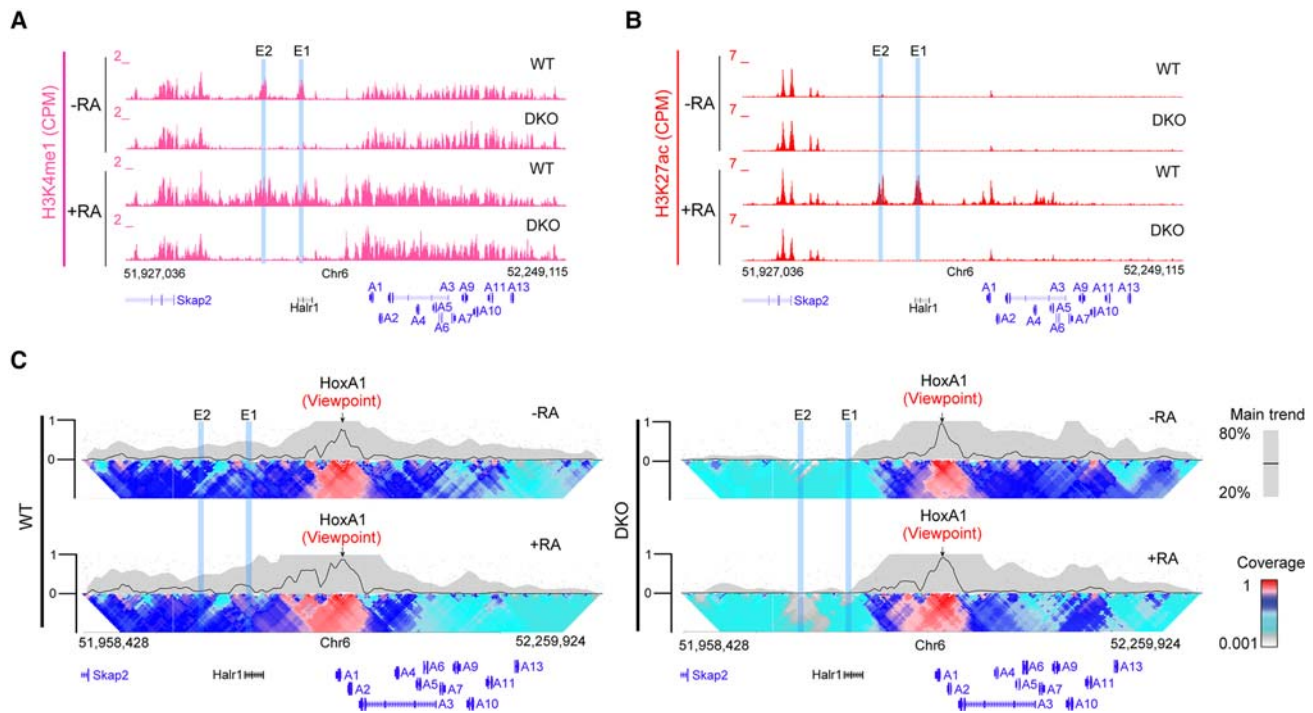


Figure 3. Effects of enhancer deletion on epigenetic marks and the three-dimensional (3D) chromatin architecture. (A) UCSC genome browser view of H3K4me1 ChIP-seq tracks at the *HoxA* gene cluster in wild-type and double-knockout ESCs. E1 and E2 are marked with blue stripes. (B) UCSC genome browser view of H3K27ac ChIP-seq tracks at the *HoxA* gene cluster in wild-type and double-knockout ESCs. E1 and E2 are marked with blue stripes. (C) 4C-seq analysis with the *Hoxa1* promoter as the viewpoint in wild-type (left) and double-knockout (right) ESCs without (top) or with (bottom) RA treatment. The median and 20th and 80th percentiles of a sliding 5-kb window determine the main trend line. The color-coded scale represents enrichment relative to the maximum median value at a resolution of 12 kb.

et al. 2010). Indeed, cohesin core component SMC1A was recruited to both E1 and E2 in undifferentiated ESCs (Supplemental Fig. S4A). To test whether cohesin contributes to the transcriptional regulation of *HoxA* genes, we performed RNA-seq in RA-treated and SMC1A-depleted ESCs (Supplemental Fig. S4B,C). Unexpectedly, the expression levels of *HoxA* genes were not perturbed by SMC1A depletion, suggesting that cohesin is not required for their activation in response to RA treatment.

To search for TFs that bind E1 and E2, we performed motif analysis of the enhancers and a control region that is located downstream from E1 and has distinct epigenetic features (Supplemental Fig. S4D) using the MatInspector software developed by Genomatix (Cartharius et al. 2005). We identified 121, 136, and 148 TF matrices for E1, E2, and the control region, respectively, under stringent conditions. Overlapping these analyses, we observed that E1 and E2 shared 56 TF matrices, within which 36 matrices could be found on the control sequence (Supplemental Fig. S4E). Due to the interaction and redundancy of E1 and E2, we focused on the 20 matrices found for E1 and E2 but not for the control sequence. A careful examination revealed that the 20 matrices contained motifs of 17 TFs, six of which have moderate to high expression levels in both undifferentiated and RA-treated ESCs (Supplemental Fig. S4F).

Within these six TFs, HIF1A protein is known to be degraded rapidly under normoxic conditions (Hu et al. 2006),

excluding the possibility that it participates in the rapid induction of *Hox* genes in our system. On the other hand, NFE2L1 knockout mouse embryos die at late gestation due to an impaired liver erythropoiesis (Chan et al. 1998), and ESCs null for NFE2L1 are able to contribute to the majority of chimeric tissues (Chen et al. 2003), while HSF2 is not essential for normal development (McMillan et al. 2002), suggesting that these two TFs do not contribute to early development and differentiation. It was reported that the chromatin remodeling protein BPTF knockout ESCs display up-regulated expression of *Hox* genes (Landry et al. 2008), indicating that BPTF acts as a repressor of *Hox* genes. ZIC2 functions with Mbd3/NuRD and participates in the regulation of certain posterior *HoxC* genes in undifferentiated ESCs (Luo et al. 2015). Interestingly, *Zic2* expression levels had a twofold increase upon RA treatment (Supplemental Fig. S4G), suggesting its importance in RA-induced differentiation, whereas YY1 has been reported to bind to both active promoters and enhancers (Sigova et al. 2015), with its dual activity on gene transcription likely resulting from its interactions with both PRC2 and the chromatin remodeler INO80 (Satijn et al. 2001; Vella et al. 2012).

Consistent with our motif analysis, ZIC2 and YY1 were recruited to E1 and E2 in ESCs (Supplemental Fig. S4H). It is also noteworthy that the H3K27 acetyltransferase P300 binds to E1 and E2 in ESCs (Supplemental Fig. S4H) prior to the activation of these two enhancers.

Such observations prompted us to deplete these known E1- and E2-binding factors to explore their functions in Hox gene regulation. Although ZIC2 depletion did not cause significant changes in *Hoxa1* and *Hoxa5* expression, both YY1 and P300 knockdown led to an increase in transcription of these two Hox genes (Supplemental Fig. S4I). It has been reported that differentiation-related genes were up-regulated in respective P300- and YY1-depleted ESCs (Zhong and Jin 2009; Vella et al. 2012), suggesting that the observed increase in Hox gene expression is at least partially due to skewed differentiation programs in ESCs.

MLL1–4 are dispensable for Hox cluster activation during RA-induced differentiation

Despite extensive studies linking COMPASS activity to transcriptional regulation of the Hox cluster, the roles of these histone methyltransferase complexes in the early activation of Hox genes were unknown. To study whether any of the COMPASS family members are recruited to Hox clusters and their regulatory sequences, we performed CHIP-seq of SET1A, MLL2, and MLL4 in RA-induced and undifferentiated ESCs. Upon RA induction, each of the three COMPASS proteins was observed to have increased occupancy at multiple regulatory sequences in the HoxA and HoxB clusters, including E1 and E2 (Supplemental Fig. S5). Many of the sites that gain COMPASS proteins also had increased active histone modifications during differentiation (Supplemental Fig. S5), suggesting an instructive role of COMPASS on the activation of Hox genes.

The loss of broad H3K4me1-enriched domains at the intergenic regions in double-knockout ESCs (Fig. 3A) suggests that the implementation of this histone mark may have an impact on the activation of HoxA genes. Our laboratory (Hu et al. 2013a) has shown that the MLL3/MLL4 branch of the mammalian COMPASS family is the major regulator of H3K4me1 at enhancers; therefore, we generated respective MLL3 knockout ESCs and MLL4 SET [Su(var)3-9, Enhancer of zeste, and Trithorax] domain truncated ESCs (referred to here as “MLL4ΔSET”) by CRISPR–Cas9-guided gene editing to determine whether these histone methyltransferases are involved in the early activation of Hox genes.

To generate MLL3 knockout ESCs, we deleted exons 8 and 9 of the *Kmt2c* gene to produce a frameshift mutant of MLL3 (Fig. 4A). For MLL4, we deleted exons 52–54 of *Kmt2d*, which encode the SET and post-SET domains, eliminating the catalytic activity of MLL4 (Fig. 4B). It is noteworthy that the mRNA levels of mutant *Kmt2c* are lower than its wild-type counterpart and that Western blotting demonstrates lower levels of mutant MLL3 protein than wild-type MLL3 in ESCs (Fig. 4A,C). Moreover, mutant MLL3 was not able to bind RBBP5, one of the core subunits of COMPASS, suggesting a loss of function of the mutant (Fig. 4C). In contrast, the level of the SET domain truncated MLL4 protein was comparable with wild-type MLL4 (Fig. 4D), indicating the stability of mutant MLL4. Both MLL3 knockout and MLL4ΔSET ESCs had normal ESC morphology (data not shown). MLL4ΔSET

cells have lower levels of H3K4me1 (Fig. 4E), suggesting that MLL4 is the major H3K4me1-catalyzing enzyme in ESCs. By treating the mutant cells and their wild-type counterparts with RA and performing RNA-seq, we found that the expression levels of Hox genes upon RA induction were not affected by individual mutants of MLL3 or MLL4 (Fig. 4F). CHIP-seq analysis indicated that H3K4me1 levels at HoxA enhancer regions were slightly decreased in the intergenic region between *Hoxa1* and *Skap2* in MLL4ΔSET cells in both pluripotent and differentiated states (Fig. 4G). These results demonstrate that single inactivation of the MLL3/MLL4 branch of the COMPASS family was not enough to impair Hox gene activation by RA induction in ESCs.

As the individual mutation of MLL3 or MLL4 did not have a significant impact on H3K4me1 levels at HoxA enhancer regions, we depleted MLL3 in MLL4ΔSET ESCs by shRNA to determine whether MLL3 and MLL4 redundantly regulate the transcription and enhancer H3K4me1 levels of Hox genes during differentiation. Although MLL3 knockdown did not cause global changes in H3K4me1 levels in wild-type ESCs (Hu et al. 2013b), H3K4me1 levels were reduced in MLL3-depleted MLL4ΔSET (MLL3 and MLL4 loss of function, referred to here as “MLL3/MLL4 LOF”) ESCs compared with control knockdown cells, while H3K4me3 levels were unchanged (Supplemental Fig. S6A), indicating the redundancy of MLL3 and MLL4 in H3K4me1 implementation. Unexpectedly, several HoxA genes, including *Hoxa1* and *Hoxa5*, were slightly up-regulated in MLL3/MLL4 LOF cells compared with control shRNA-infected cells upon RA-induced differentiation (Supplemental Fig. S6B,D). Gene ontology (GO) analysis indicated that genes involved in neuron development were enriched in up-regulated genes in MLL3/MLL4 LOF cells, while genes functioning in embryo development and stem cell maintenance were down-regulated (Supplemental Fig. S6C). These data suggest that the moderate increase in expression levels of HoxA genes in MLL3/MLL4 LOF ESCs is probably due to the accelerated differentiation state of these cells. Indeed, pluripotency genes such as *Pou5f1* and *Nanog* were down-regulated in MLL3/MLL4 LOF cells compared with MLL4ΔSET cells upon differentiation (Supplemental Fig. S6E). H3K4me1 levels at several enhancers near the HoxA cluster were reduced by MLL3/MLL4 LOF in undifferentiated cells, and H3K4me1 at multiple enhancers failed to increase or spread upon differentiation of MLL3/MLL4 LOF cells (Supplemental Fig. S6F). These results indicate that MLL3 and MLL4 coregulate enhancer H3K4me1 at the HoxA gene cluster, but the loss of H3K4me1 levels in MLL3/MLL4 LOF cells does not lead to reduction in the expression of HoxA genes, suggesting that the context of MLL3/MLL4 COMPASS could play a central role over their methyltransferase function.

TrxG and PcG proteins have been proposed to fine-tune the expression of Hox genes (Ernst et al. 2004; Ringrose and Paro 2004; Boyer et al. 2006; Piunti and Shilatifard 2016; Rickels et al. 2016). To further explore the roles of TrxG proteins in Hox gene activation, we generated MLL1 knockout ESCs by deleting exons 11–15 of the

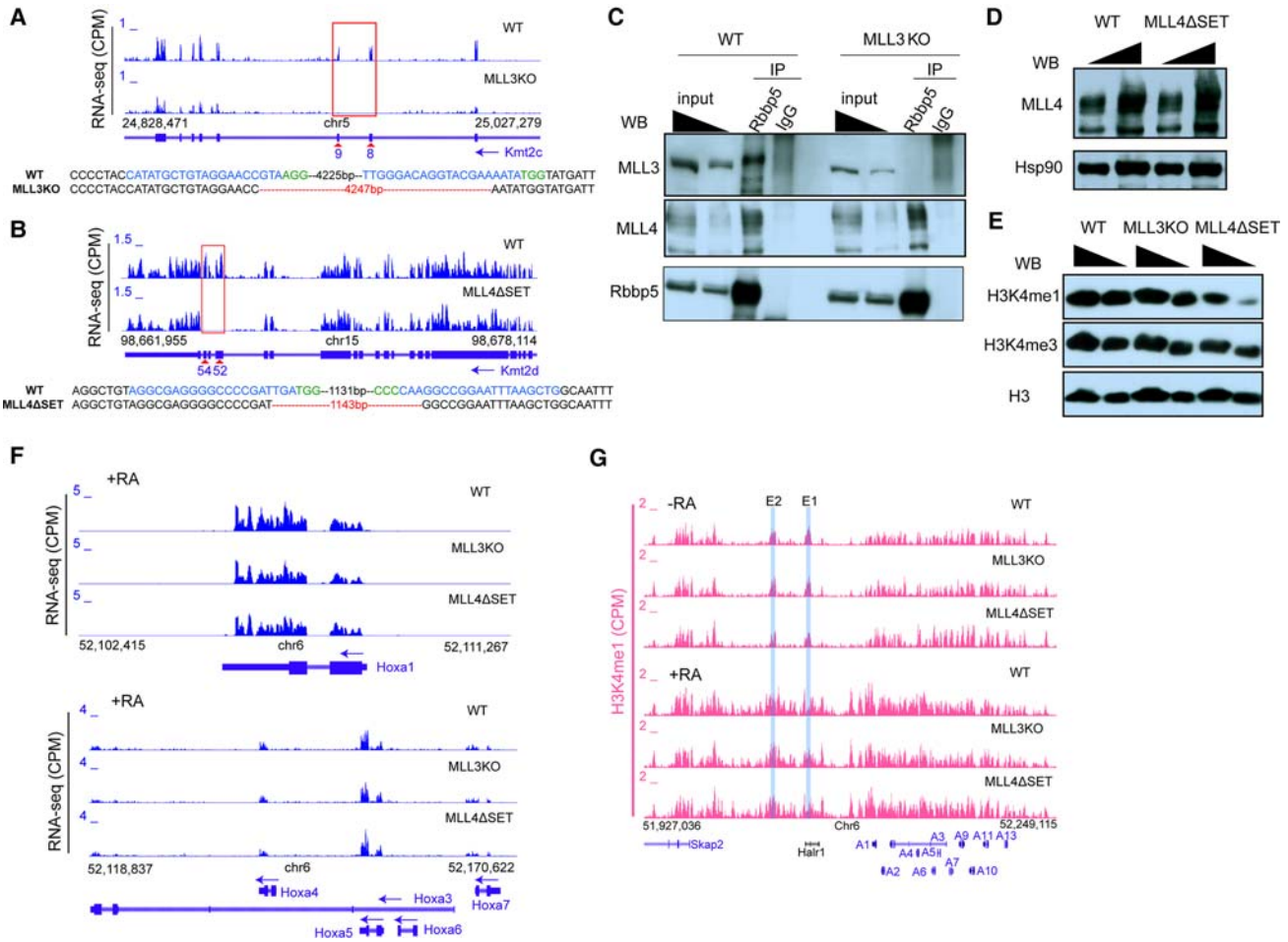


Figure 4. Inactivation of MLL3 and MLL4 in ESCs has little effect on Hox gene activation. (A,B) UCSC genome browser view of RNA-seq signals and Sanger sequencing results showing successful deletion of the indicated exons (red arrows and rectangles) at *Kmt2c* (A) and *Kmt2d* (B) genes, respectively. Genome locations of gRNAs are labeled in blue, and PAM sequences are labeled in green. Red dashed lines represent deleted sequences. (C) Coimmunoprecipitation assay using Rbbp5 antibody and IgG with nuclear extract from wild-type and MLL3 knockout ESCs. Antibodies used for Western blotting are labeled beside each blot. (D) Western blotting assay of wild-type and MLL4ΔSET ESCs with antibodies against MLL4 and Hsp90. (E) Western blotting assay of wild-type, MLL3 knockout, and MLL4ΔSET cell lysates with H3K4me1, H3K4me3, and H3 antibodies. (F) UCSC genome browser view of RNA-seq signals at the *Hoxa1* (top) and *Hoxa3–a7* (bottom) genes of wild-type, MLL3 knockout, and MLL4ΔSET ESCs upon RA induction. (G) UCSC genome browser view of H3K4me1 ChIP-seq at the *HoxA* cluster in undifferentiated (top) and RA-treated (bottom) wild-type, MLL3 knockout, and MLL4ΔSET ESCs, respectively.

Kmt2a gene. RNA-seq analysis and Sanger sequencing confirmed the deletion at *Kmt2a* in MLL1 knockout ESCs (Supplemental Fig. S7A), and Western blotting demonstrated that MLL1 protein was undetectable in these cells (Supplemental Fig. S7B). Bulk H3K4me levels were not perturbed by MLL1 depletion, as indicated by Western blotting (Supplemental Fig. S7C). We did not observe changes in HoxA expression in MLL1 knockout or previously generated MLL2 knockout ESCs (Hu et al. 2017) before or after RA treatment (Supplemental Fig. S7D), supporting the notion that MLL1 and MLL2 are dispensable for early activation of Hox genes (Hu et al. 2013b; Denisov et al. 2014). Furthermore, MLL1 or MLL2 depletion had little effect on the levels of H3K4me1 at Hox gene clusters (Supplemental Fig. S7E). Overall, our results indi-

cate that the inactivation of MLL1–4 methyltransferases has little impact on transcriptional activation of Hox genes by RA induction in ESCs.

SET1A/COMPASS regulates transcriptional activation of genes within the HoxA cluster independently of the shadow enhancers

Four of the six methyltransferases of the COMPASS family are dispensable for RA-induced Hox gene activation when individually deleted or inactivated in ESCs (Fig. 4; Supplemental Figs. S6, S7). In the remaining two branches of the COMPASS family, SET1B is lowly expressed in ESCs (data not shown), while SET1A is required for ESC pluripotency and mouse blastocyst development in vivo

(Bledau et al. 2014). Furthermore, SET1A binding at many Hox regulatory sequences was elevated upon differentiation (Supplemental Fig. S5). Based on this evidence, we hypothesized that SET1A may participate in regulating Hox gene expression during ESC differentiation. To test this hypothesis, we attempted to generate SET1A-null ESCs by CRISPR–Cas9-guided gene editing but were not able to derive homozygous SET1A knockout ESC clones (T Sun, K Cao, and A Shilatifard, unpubl.), indicating the importance of SET1A in ESC self-renewal. We therefore depleted SET1A by shRNA knockdown in ESCs and induced differentiation by RA treatment. Knockdown of SET1A led to a substantial reduction in SET1A protein

levels (Fig. 5A) and decreased H3K4me3 levels, consistent with previous observations (Wu et al. 2008; Fang et al. 2016). Unlike other COMPASS family members, SET1A depletion led to a drastic reduction in the expression levels of *Hoxa4–Hoxa7* genes in RA-induced ESCs despite *Hoxa1* levels remaining unaffected, as shown by RNA-seq (Fig. 5B).

To test whether regulation of HoxA genes by SET1A/COMPASS relies on E1 and E2 enhancers, we depleted SET1A in double-knockout ESCs and differentiated the cells by RA treatment. We observed that *Hoxa4–Hoxa7* expression levels were further diminished in double-knockout ESCs by SET1A/COMPASS depletion (Fig.

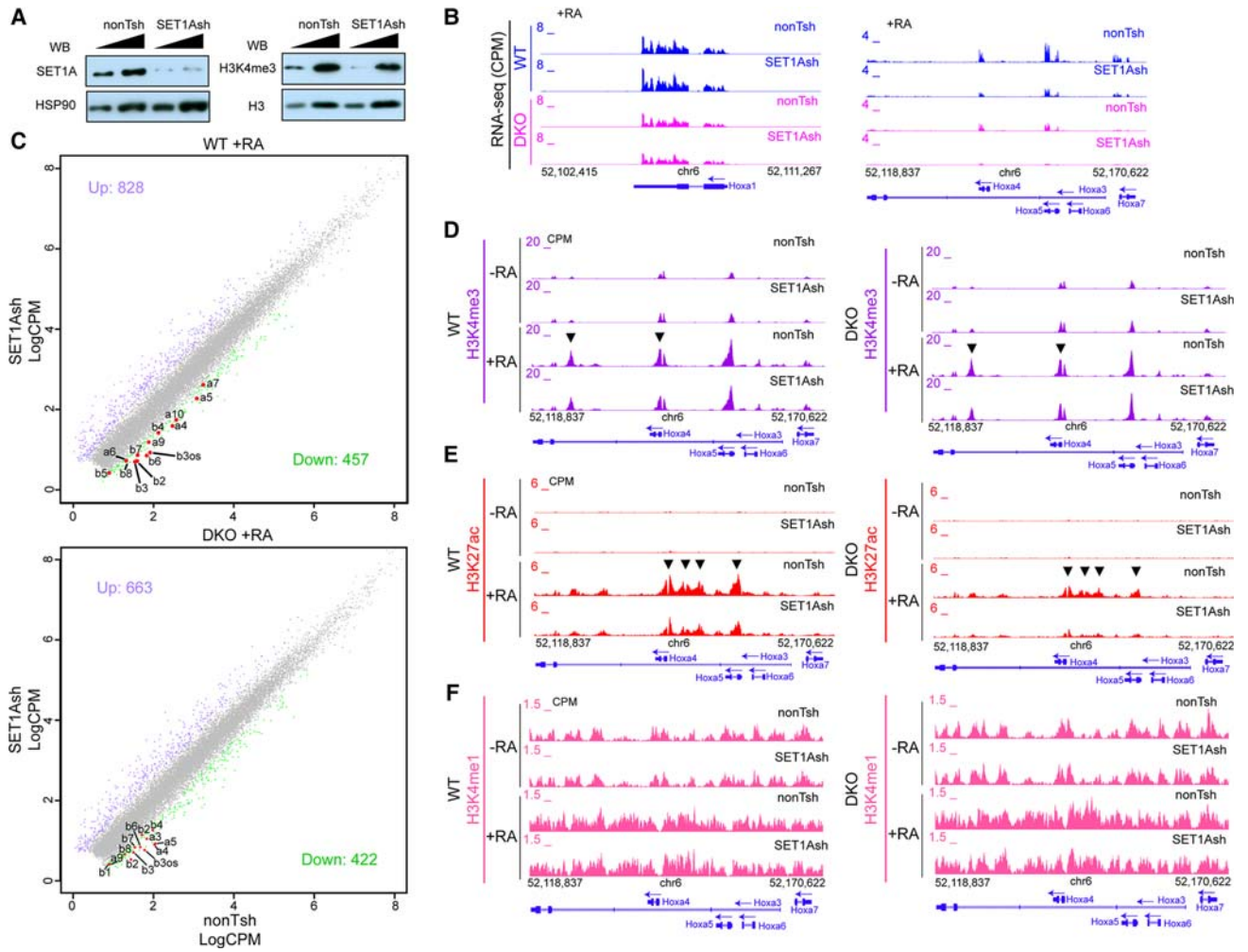


Figure 5. Depletion of SET1A leads to failures in Hox gene activation. (A) Western blotting assay of SET1A (left) and H3K4me3 (right) with control and SET1A-depleted cell lysates, respectively. HSP90 and H3 Western blots served as loading controls, respectively. (B) UCSC genome browser view of RNA-seq signals at the *Hoxa1* (left) and *Hoxa3–a7* (right) genes of RA-treated wild-type (blue) and double-knockout (pink) ESCs infected with respective nontargeting shRNA (nonTsh) and SET1A shRNA (SET1Ash). (C) Correlation analysis of gene expression levels in SET1Ash- and nonTsh-infected wild-type (top) and double-knockout (bottom) ESCs treated with RA. The X-axis represents the expression level (log₂ normalized CPM) of nonTsh-infected cells, and the Y-axis represents the expression level of SET1Ash-infected cells. Significantly changed Hox genes are labeled with red dots. Significantly down-regulated genes (compared with nonTsh-infected cells) are labeled in green, and up-regulated ones are labeled in purple. The numbers of significantly changed genes are shown in each plot. (D–F) UCSC genome browser view of H3K4me3 (D), H3K27ac (E), and H3K4me1 (F) ChIP-seq at *Hoxa3–a7* genes in undifferentiated or RA-treated wild-type (left) and double-knockout (right) ESCs infected with respective nonTsh and SET1Ash. Black arrowheads indicate peaks of histone marks altered in SET1A-depleted cells.

5B), suggesting E1- and E2-independent regulatory mechanisms of SET1A on these HoxA genes. On the other hand, *Hoxa1* expression levels in double-knockout cells remained unchanged by SET1A knockdown, further indicating that multiple pathways and factors are potentially involved in Hox gene activation. The effects of SET1A/COMPASS on gene expression were not restricted to the HoxA cluster, as HoxB genes were also down-regulated in SET1A-depleted cells upon RA induction (Fig. 5C). GO analysis of misregulated genes in SET1A-depleted RA-treated ESCs showed that P53 pathway genes were significantly up-regulated, suggesting activation of apoptotic pathways (Supplemental Fig. S8A), whereas pattern specification process-related genes, including Hox genes, were significantly down-regulated (Supplemental Fig. S8B). Similar groups of genes were found through GO analysis on SET1A-regulated genes in double-knockout ESCs (data not shown). To explore whether Hox gene down-regulation by SET1A/COMPASS depletion reflected a RA induction-specific role for SET1A/COMPASS activity, we identified a gene set significantly up-regulated by RA treatment in wild-type cells and examined the behavior of these targets upon SET1A depletion. Of the 1212 RA-activated genes, only 164 genes in this set were altered upon SET1A depletion; moreover, of this small subset, half displayed down-regulation, whereas the other half were up-regulated (Supplemental Fig. S8C). Collectively, this analysis strongly suggests that SET1A does not have a general role in regulating RA-responsive genes but rather has a locus-specific function at Hox gene clusters. It is noteworthy that SET1A depletion did not decelerate RA-induced differentiation, as reflected by the expression levels of pluripotency genes and differentiation markers (Supplemental Fig. S8D). Thus, we conclude that SET1A regulates RA-triggered activation of Hox genes in ESCs independently of the E1 and E2 enhancers.

To further explore the mechanisms through which SET1A modulates the transcriptional levels of Hox cluster genes, we performed ChIP-seq of H3K4me1, H3K27ac, and H3K4me3 in control and SET1A-depleted ESCs. The distribution of H3K4me1 was similar between control and SET1A-depleted cells at the HoxA cluster (Fig. 5F; Supplemental Fig. S8E), while H3K4me3 levels at *Hoxa3* and *Hoxa4* loci were reduced upon SET1A depletion in RA-treated cells (Fig. 5D; Supplemental Fig. S8E). Similar patterns could be observed at the HoxB cluster, which had unchanged H3K4me1 and local reductions in H3K4me3 signals upon SET1A depletion in differentiated cells (Supplemental Fig. S8E, right). Interestingly, H3K27ac peaks were reduced at *Hoxa3–a7*, *Hoxb2*, and a previously reported enhancer region between *Hoxb4* and *Hoxb5* (Sharpe et al. 1998) in SET1A-depleted cells (Fig. 5E; Supplemental Fig. S8D). Knockdown of SET1A in double-knockout cells also led to further reductions in H3K27ac levels in RA-treated cells (Fig. 5E), confirming the additive effects of enhancers and SET1A on HoxA gene regulation.

Histone H3K4me2 has been shown previously to be catalyzed by different COMPASS family members in different organisms (Wang et al. 2009; Herz et al. 2012; Hu et al. 2013a; Rickels et al. 2016). To test which form of

COMPASS is responsible for H3K4me2 levels at Hox clusters in ESCs, we performed H3K4me2 ChIP-seq in undifferentiated and RA-treated wild-type and COMPASS-inactivated cells. H3K4me2 levels within Hox clusters were drastically reduced in MLL2 knockout ESCs, while inactivation of other COMPASS proteins had little impact (Supplemental Fig. S9). RA treatment led to a significant increase of H3K4me2 levels at enhancers and promoters in HoxA and HoxB clusters, while MLL2 deletion only led to decreased H3K4me2 at promoter-proximal regulatory sequences. These data indicate that MLL2 is the major methyltransferase responsible for H3K4me2 at Hox genes in ESCs, while H4K4me2 at distal Hox enhancers such as E1 and E2 may be implemented redundantly by several COMPASS family members.

Discussion

TFs within the Hox family are responsible for establishing animal body patterning, and their misregulation has been linked to oncogenesis. Transcriptional regulation of the genes encoding these TFs includes *cis*-regulation through enhancer–promoter looping and *trans*-regulation through protein complexes such as PcG and TrxG. Unveiling how Hox genes are turned on during development and malignant transformation would pave the way for understanding cell fate specification and cancer pathogenesis as well as developing novel cancer therapies.

In this study, we identified two *cis*-regulatory elements located upstream of the HoxA cluster that govern activation of the entire cluster in response to RA. Analysis of active enhancer marks in undifferentiated and RA-treated ESCs indicated that E1 and E2 are the most prominent putative enhancers in the gene desert region between *Hoxa1* and *Skap2* (Fig. 1B). However, we note that there are other potential enhancer-like regions that harbor H3K27ac and interact with *Hoxa1* upon differentiation, including previously reported RARE-containing regions (Frasch et al. 1995; De Kumar et al. 2015) located near or within the HoxA cluster. It is noteworthy that the HoxA cluster interacts with broad domains covering regions with localized H3K27ac upon RA treatment (Fig. 1C). We surmise that the H3K27ac-decorated regions such as E1 and E2 are enhancer “cores,” while the anchors of chromatin loops are away from the “cores,” and differentiation-triggered chromatin looping brings enhancers into proximity to Hox genes to drive gene expression. These enhancer elements likely act cooperatively as a regulatory archipelago for the transcription of anterior HoxA genes, similar to what was reported for the HoxD cluster (Montavon et al. 2011).

Interestingly, the E1 region identified in our study is located within the lncRNA gene *Halr1*, whose expression levels are highly elevated upon RA induction. Several studies have shown that Halr1 RNA dampens the activation of HoxA genes through a PRC2 recruitment mechanism (Maamar et al. 2013; Yin et al. 2015; Liu et al. 2016). However, our data argue that Halr1 depletion does not alter HoxA gene expression levels in ESCs

(Supplemental Fig. 2C). ESCs grown in “2i+LIF” (two inhibitors [MEK inhibitor and GSK3 inhibitor] and leukemia inhibitory factor) conditions used in the current study have transcriptome and epigenome profiles different from traditional “serum+LIF”-grown ESC such that H3K27me3 levels and bivalent domains are significantly reduced in “2i+LIF”-grown cells (Marks et al. 2012). The differential cell states established by growth conditions may have an impact on the downstream effects of *Halr1* depletion. It is also worth noting that mice deleted of *Halr1* gene from exon 2 to the end of exon 3, a region not overlapping with E1, are viable (Sauvageau et al. 2013), demonstrating that the *Halr1* transcript is dispensable for embryonic development. It would be interesting to determine the expression levels of HoxA genes in *Halr1*-deleted embryos at different developmental stages to further dissect the potential antagonisms between *Halr1* and HoxA activation. During the preparation of our report, the loss of an ~7-kb region—from the promoter to ~2 kb after the second exon of *Halr1*—has been shown to cause a minor defect in the transcriptional activation of several HoxA genes (Yin et al. 2015). Unlike this truncation, our E1 deletion does not include the H3K4me3-enriched *Halr1* promoter or the CTCF-binding site after the second exon of *Halr1* (Supplemental Fig. S1A,B). The subtle differences in our experimental design may provide further molecular insight into HoxA cluster regulation.

The existence of two enhancers apparently acting redundantly for the early activation of Hox genes is reminiscent of the “shadow enhancers” that regulate transcription in *Drosophila* embryos (Hong et al. 2008; Smith and Shilatifard 2014). Individual knockout of E1 or E2 has little impact on HoxA gene expression upon RA-induced differentiation, whereas compound deletion of the two regions causes significant and specific loss of transcription at the HoxA gene cluster (Fig. 2). The redundancy observed for E1 and E2 is different from α -globin enhancers, which act in an additive manner (Hay et al. 2016), suggesting the locus specificity of enhancer commissioning (Smith and Shilatifard 2014). Although the previously generated 58-kb knockout ESCs (Yin et al. 2015) showed a reduction of HoxA gene induction upon RA treatment, the 58-kb deleted region not only contains E1 and E2 but also covers a CTCF-binding site and a broad H3K4me1-decorated domain harboring several RAREs (Fig. 1B; De Kumar et al. 2015). The loss of such a large region may not reflect the effects of deleting individual *cis*-regulatory elements but rather the outcome of losing numerous positive and negative regulatory elements. Similarly, a recent study reported that the deletion of a 40-kb region proximal to *Hoxa1* containing the RARE enhancers led to a delay of *Hoxa1* induction through Wnt pathway activation (Neijts et al. 2016). There are at least three potential *cis*-regulatory elements in this 40-kb region based on increased H3K27ac as observed in our study (Fig. 1B). Whether these elements have additive, synergistic, or redundant effects on *Hoxa1* expression could not be assessed by deleting a broad domain containing multiple *cis*-regulatory regions. On the other hand, our study focuses on specific putative enhancer regions (~3 kb each) and

reveals cooperation between the enhancer regions without disrupting other potential regulatory elements.

The E1- and E2-dependent spreading of H3K4me1 over the majority of the ~150-kb gene desert between *Skap2* and *Hoxa1* upon RA-induced differentiation (Fig. 3A), concomitant with increased interactions between E1, E2, and the surrounding DNA regions (Fig. 1C), suggests that E1 and E2 are likely anchoring points or “hubs” for events triggered by RA. DNA-binding motifs for 17 TFs could be found on both E1 and E2, and a number of chromatin proteins and TFs are recruited to E1 and E2 in ESCs, including cohesin, ZIC2, YY1, and P300 (Supplemental Fig. S4). However, depletion of these factors did not dampen the transcription levels of *Hoxa1* and *Hoxa5* in our system. It is possible that unknown factors or cooperation of multiple chromatin regulators contribute to the activity of enhancer elements; thus, depletion of a single factor would not be able to undermine the transcriptional outputs. The interaction between E1 and E2 during RA-induced differentiation adds to the complexity and possible mechanisms of *cis*-regulatory elements on transcription regulation. Such an intricate regulation network may be necessary to assure the proper regulation of critical developmental gene clusters. Future studies, such as unbiased screening with shRNA or CRISPR/Cas9 libraries, may shed further light on additional factors required for enhancer function and gene activation at Hox clusters.

Our ChIP-seq analysis demonstrated that COMPASS family members are recruited to HoxA and HoxB clusters (Supplemental Fig. S5). However, SET1A, MLL2, and MLL4 preferentially occupy distinct regulatory regions. MLL2 and SET1A have higher occupancy at proximal regulatory sequences, while MLL4 binds more distal sites, suggesting different roles for COMPASS proteins in modulating histone modifications and transcription on Hox clusters. Inactivation of MLL3 or MLL4 did not alter Hox gene transcription. Our MLL3 mutation generated a frameshift mutated form of MLL3 that does not bind RBBP5, one of the common components of the COMPASS family of methyltransferase complexes, while the MLL4 Δ SET mutant lacks the catalytic domain (Fig. 4). We did not observe a drastic loss of H3K4me1 at the Hox clusters in either mutant cell line, although H3K4me1 was decreased globally in the MLL4 Δ SET cells. The MLL3 and MLL4 branches of COMPASS are responsible for H3K4me1 deposition at enhancers (Herz et al. 2012; Hu et al. 2013a). Consistently, we observed redundancy of MLL3 and MLL4 in maintaining H3K4me1 levels in ESCs and found that MLL3 and MLL4 are responsible for spreading of H3K4me1 in the HoxA cluster during differentiation (Supplemental Fig. S6). Nevertheless, transcription levels of Hox genes were not reduced in MLL3/MLL4 LOF ESCs upon differentiation, suggesting that the reduction of H3K4me1 levels at enhancers is insufficient to impair early Hox gene activation in ESCs.

The residual H3K4me1 that we observed at E1 and E2 in MLL3/MLL4 LOF ESCs suggests additional redundancy within the COMPASS family. Indeed, MLL2 and SET1A are recruited to these enhancers and may contribute to the deposition of H3K4me1. It is also noteworthy that

the potential acceleration of differentiation in MLL3/MLL4 LOF ESCs may overcome the local activating functions of MLL3 and MLL4 on Hox genes. Therefore, dissecting the functions of other components of MLL3 and MLL4 complexes may reveal roles on Hox gene activation, as the loss of function of certain components may not impair ESC pluripotency as severely as MLL3/MLL4 LOF. UTX, a H3K27 demethylase and common component of MLL3 and MLL4 COMPASS, has been shown to regulate RA induction of transcription and H3K4me3/H3K27me3 levels at bivalent genes, including Hox genes (Lee et al. 2007; Dhar et al. 2016). Depletion of TRR, the *Drosophila* homolog of MLL3/MLL4, leads to the degradation of UTX (Herz et al. 2012). However, whether the functions of UTX on Hox gene regulation are dependent on MLL3/MLL4 COMPASS is unknown. Unlike MLL3 and MLL4, which are required for mouse embryogenesis, UTX-null male mice develop to term without defects in early development (Shpargel et al. 2012). MLL3/MLL4 conditional knockout mice would be useful tools to dissect the roles of these methyltransferases on Hox gene activation in a stage-specific and tissue type-specific manner during development.

SET1A and SET1B are mammalian homologs of *Drosophila* SET1. SET1A is highly expressed in ESCs, while SET1B is lowly expressed. SET1A knockout mouse embryos die before embryonic day 7.5 (E7.5) with failures in gastrulation (Bledau et al. 2014). Moreover, ESCs could not be derived from outgrowth of SET1A-null blastocysts (Bledau et al. 2014), suggesting the requirement of SET1A for ESC self-renewal. We found that SET1A depletion triggers apoptosis and impairs the activation of Hox genes induced by RA (Fig. 5). Such results are consistent with the reported developmental failure of SET1A-null inner cell mass (Bledau et al. 2014; Fang et al. 2016). SET1A acts independently from the enhancers E1 and E2, as depletion of SET1A in the double-knockout ESCs caused further reduction of HoxA gene activation (Fig. 5). Besides E1 and E2, there are at least six putative enhancer regions that gain H3K27ac upon RA induction in the gene desert between *Skap2* and *Hoxa1* (Fig. 1). SET1A recruitment increases at one of these putative enhancers and at multiple gene-proximal elements in response to RA treatment, and SET1A depletion leads to reduced H3K27ac at many of these sites (Supplemental Fig. S8), suggesting that SET1A directly regulates the activity of these *cis*-regulatory elements. H3K4me3 peaks located near *Hoxa3* and *Hoxa4* were also reduced in SET1A-depleted cells, supporting a promoter-proximal role of SET1A in regulating Hox gene expression. Similar activities of SET1A are observed at previously characterized HoxB enhancers. We also noted that, unlike E1's and E2's effects on *Hoxa1* transcription, SET1A depletion did not alter *Hoxa1* expression (Fig. 5B). It is possible that SET1A is dispensable for the activity of strong proximal enhancers such as the RARE elements, which require the distal enhancers E1 and E2 to maintain optimal *Hoxa1* expression. Indeed, we observed robust interactions between *Hoxa1*-proximal enhancers and E1 (Fig. 1C), suggesting cooperative effects of the distal and proximal enhancers. Furthermore, as shown by

Neijts et al. (2016), HoxA-proximal enhancers have a stronger impact on *Hoxa1* than posterior HoxA genes (e.g., *Hoxa5–7*), which may rely on both distal enhancers and SET1A to get activated. Thus, SET1A depletion has a more profound effect on the activation of posterior HoxA genes.

In summary, we identified two distal *cis*-regulatory elements governing the early activation of the HoxA gene cluster and have shown that these enhancers act in a redundant manner on gene transcription. Deletion of the enhancers leads to the removal of active histone modifications and reorganization of 3D chromatin structure at HoxA genes, while SET1A/COMPASS regulates Hox gene transcription through activating proximal enhancers and promoters. Deciphering the enigmatic mechanisms of Hox gene regulation could provide critical insight for clinical applications in diseases driven by misregulation of Hox gene expression.

Materials and methods

Antibodies

The following antibodies were used in this study: anti-H3K4me1 (generated in-house), anti-H3K4me2 (generated in-house), anti-H3K4me3 (generated in-house), anti-H3K27ac (Cell Signaling, 8173), anti-H3 (generated in-house), anti-SMC1A (Bethyl Laboratories, A300-055A), anti-SET1A (generated in-house), anti-MLL1 (Cell Signaling, 14689), anti-MLL2 (generated in-house), anti-MLL3 (generated in-house), anti-MLL4 (generated in-house), anti-RBBP5 (Bethyl Laboratories, A300-109A), anti-HSP90 (Santa Cruz Biotechnology, 7947), and anti-tubulin (Developmental Studies Hybridoma Bank, E7).

ESC culture, shRNA knockdown, and CRISPR-Cas9-guided knockout

V6.5 ESCs were grown in N2B27 medium supplemented with 2i/LIF as described previously (Ying et al. 2008). For RA-induced differentiation, ESCs were grown in plain N2B27 medium with 1 μ M all-*trans* RA (Sigma) without addition of 2i or LIF. All RA treatments were performed for 24 h unless specifically mentioned. Knockdown was performed as described in Hu et al. (2013b). The sequences of the shRNA against *Kmt2c* and *Set1a* transcripts were published previously (Hu et al. 2013b; Fang et al. 2016).

Oligos encoding the desired gRNA sequences were annealed and cloned into pX459 plasmid according to a published protocol (Ran et al. 2013). Donor plasmids containing homology arms flanking CMV-GFP and PGK-Neo were constructed in pBlue-Script SK(+) vector. ESCs were transfected with pX459 containing the desired gRNA-coding sequences and donor plasmid using Nucleofector 2b (Lonza) according to the manufacturer's instructions. Transfected ESCs were selected with 2 μ g/mL puromycin (Life Technologies) for 48 h and 200 μ g/mL G418 (Life technologies) for 10 d until cell clones were pickable. Cell clones were screened with Southern blotting assay, and the genotypes were confirmed by PCR. Desired cell clones were transfected with pCAG-Cre (Addgene, 26646) to remove the GFP/Neo cassette. The following oligo sequences encoding sgRNAs were used in this study: E1 knockout (left, AGCACACGTGCTTTTAAGTC; right, CTCAGCTTCTCTGGAAGAGC), E2 knockout (left, GCTCCATTCCATTAAGAACA; right, ACAGTAAATATGCGCGACAC), MLL1 knockout (left, TCACTTGCTGCGCTACCG

TC; right, TCTAAGCAAACTTGTGGAA), MLL3 knockout (left, CATATGCTGTAGGAACCGTA; right, TTGGGACAGG TACGAAAATA), and MLL4 Δ SET (left, AGGCGAGGGG CCCCATTGA; right, CAGCTTAAATTCCGGCCTTG).

ChIP-seq assay

ChIP was performed as described previously (Lee et al. 2006) with modifications. Briefly, ESCs were fixed with 1% formaldehyde for 10 min. After quenching and cell lysis, chromatin were sonicated using an E220 focused ultrasonicator (Covaris). One-hundred micrograms of sheared chromatin, 5 μ g of antibody, and 50 μ L of protein A/G beads (Santa Cruz Biotechnology) were used for each immunoprecipitation. Immunoprecipitated DNA were purified after washing, eluting, and reverse cross-linking and submitted for library preparation. ChIP-seq libraries were generated with KAPA HTP library preparation kit (KAPA Biosystems) following the manufacturer's instruction and loaded onto NextSeq 500 sequencer (Illumina) for sequencing. At least two biological replicates were performed for ChIP-seq of H3K4me1, H3K27ac, and H3K4me3 under each experimental condition.

4C-seq

4C-seq was performed following the protocol published in van de Werken et al. (2012a) with DpnII (New England Biolabs) and NlaIII (New England Biolabs) as the first and second restriction enzymes, respectively. The primers below were used to amplify the final 4C template for generating 4C-seq libraries, and index sequences are underlined. Reads were sorted into individual sample FastQ files according to the index sequences, and then 4C-seq-pipe (van de Werken et al. 2012b) was used to align the reads and analyze the DNA interactions around viewpoints. The primers used were as follows: HoxA1 (forward, AATGATACGGCG ACCACCGAGATCTACACTCTTCCCTACACGACGCTCTT CCGATCTNNNNNNATGCCACTGAAACGGTGATC; reverse, CAAGCAGAAGACGGCATAACGAGATCAGAGGATTG ACTGGGAGGA) and E1 (forward, AATGATACGGCGACC ACCGAGATCTACACTCTTCCCTACACGACGCTCTCCG ATCTNNNNNNNTAACCTGCCCTCAGGAGATC; reverse, CAAGCAGAAGACGGCATAACGAGATTGGAGAGACGGTCC AGAGTT).

Data availability

Next-generation sequencing data sets have been deposited at Gene Expression Omnibus with accession number GSE98140.

Acknowledgments

We thank the Shilatifard laboratory members for helpful suggestions and discussions. We thank Dr. Edwin Smith for critical reading and editing of this manuscript, and Dr. Robb Krumlauf for discussion. This study was supported in part by grant R01GM069905 from the National Institute of Health to A.S.

References

Alexander T, Nolte C, Krumlauf R. 2009. Hox genes and segmentation of the hindbrain and axial skeleton. *Annu Rev Cell Dev Biol* **25**: 431–456.
 Andrey G, Montavon T, Mascres B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. 2013. A switch be-

tween topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**: 1234–1237.
 Bledau AS, Schmidt K, Neumann K, Hill U, Ciotta G, Gupta A, Torres DC, Fu J, Kranz A, Stewart AF, et al. 2014. The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development* **141**: 1022–1035.
 Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, et al. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**: 349–353.
 Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, Frisch M, Bayerlein M, Werner T. 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**: 2933–2942.
 Chambeyron S, Bickmore WA. 2004. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev* **18**: 1119–1130.
 Chambeyron S, Da Silva NR, Lawson KA, Bickmore WA. 2005. Nuclear re-organisation of the Hoxb complex during mouse embryonic development. *Development* **132**: 2215–2223.
 Chan JY, Kwong M, Lu R, Chang J, Wang B, Yen TS, Kan YW. 1998. Targeted disruption of the ubiquitous CNC-bZIP transcription factor, Nrf-1, results in anemia and embryonic lethality in mice. *EMBO J* **17**: 1779–1787.
 Chen L, Kwong M, Lu R, Ginzinger D, Lee C, Leung L, Chan JY. 2003. Nrf1 is critical for redox balance and survival of liver cells during development. *Mol Cell Biol* **23**: 4673–4686.
 De Kumar B, Parrish ME, Slaughter BD, Unruh JR, Gogol M, Seidel C, Paulson A, Li H, Gaudenz K, Peak A, et al. 2015. Analysis of dynamic changes in retinoid-induced transcription and epigenetic profiles of murine Hox clusters in ES cells. *Genome Res* **25**: 1229–1243.
 Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA. 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**: 1233–1244.
 Deng W, Rupon JW, Krivega I, Breda L, Motta I, Jahn KS, Reik A, Gregory PD, Rivella S, Dean A, et al. 2014. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158**: 849–860.
 Denissov S, Hofemeister H, Marks H, Kranz A, Ciotta G, Singh S, Anastassiadis K, Stunnenberg HG, Stewart AF. 2014. Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141**: 526–537.
 Dhar SS, Lee SH, Chen K, Zhu G, Oh W, Allton K, Gafni O, Kim YZ, Tomoiga AS, Barton MC, et al. 2016. An essential role for UTX in resolution and activation of bivalent promoters. *Nucleic Acids Res* **44**: 3659–3674.
 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376–380.
 Down JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schuijers J, Lee TI, Zhao K, et al. 2014. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**: 374–387.
 Dupe V, Davenne M, Brocard J, Dolle P, Mark M, Dierich A, Chambon P, Rijli FM. 1997. In vivo functional analysis of the Hoxa-1 3' retinoic acid response element (3'RARE). *Development* **124**: 399–410.
 Ernst P, Mamon M, Davidson AJ, Zon LI, Korsmeyer SJ. 2004. An Mll-dependent Hox program drives hematopoietic progenitor expansion. *Curr Biol* **14**: 2063–2069.

- Fang L, Zhang J, Zhang H, Yang X, Jin X, Zhang L, Skalnik DG, Jin Y, Zhang Y, Huang X, et al. 2016. H3K4 methyltransferase Set1a is a key Oct4 coactivator essential for generation of Oct4 positive inner cell mass. *Stem Cells* **34**: 565–580.
- Frasch M, Chen X, Lufkin T. 1995. Evolutionary-conserved enhancers direct region-specific expression of the murine Hoxa-1 and Hoxa-2 loci in both mice and *Drosophila*. *Development* **121**: 957–974.
- Hay D, Hughes JR, Babbs C, Davies JO, Graham BJ, Hanssen LL, Kassouf MT, Oudelaar AM, Sharpe JA, Suciuc MC, et al. 2016. Genetic dissection of the α -globin super-enhancer in vivo. *Nat Genet* **48**: 895–903.
- Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, Voets O, Verrijzer CP, Shilatifard A. 2012. Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian Mll3/Mll4. *Genes Dev* **26**: 2604–2620.
- Hong JW, Hendrix DA, Levine MS. 2008. Shadow enhancers as a source of evolutionary novelty. *Science* **321**: 1314.
- Hu CJ, Iyer S, Sataur A, Covello KL, Chodosh LA, Simon MC. 2006. Differential regulation of the transcriptional activities of hypoxia-inducible factor 1 α (HIF-1 α) and HIF-2 α in stem cells. *Mol Cell Biol* **26**: 3514–3526.
- Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. 2013a. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol Cell Biol* **33**: 4745–4754.
- Hu D, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, Shilatifard A. 2013b. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat Struct Mol Biol* **20**: 1093–1097.
- Hu D, Gao X, Cao K, Morgan MA, Mas G, Smith ER, Volk AG, Bartom ET, Crispino JD, Di Croce L, et al. 2017. Not all H3K4 methylations are created equal: Mll2/COMPASS dependency in primordial germ cell specification. *Mol Cell* **65**: 460–475 e466.
- Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**: 430–435.
- Kmita M, Duboule D. 2003. Organizing axes in time and space: 25 years of colinear tinkering. *Science* **301**: 331–333.
- Landry J, Sharov AA, Piao Y, Sharova LV, Xiao H, Southon E, Matta J, Tessarollo L, Zhang YE, Ko MS, et al. 2008. Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. *PLoS Genet* **4**: e1000241.
- Lee TI, Johnstone SE, Young RA. 2006. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* **1**: 729–748.
- Lee MG, Villa R, Trojer P, Norman J, Yan KP, Reinberg D, Di Croce L, Shiekhattar R. 2007. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318**: 447–450.
- Lin C, Garrett AS, De Kumar B, Smith ER, Gogol M, Seidel C, Krumlauf R, Shilatifard A. 2011. Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). *Genes Dev* **25**: 1486–1498.
- Liu GY, Zhao GN, Chen XF, Hao DL, Zhao X, Lv X, Liu DP. 2016. The long noncoding RNA Gm15055 represses Hoxa gene expression by recruiting PRC2 to the gene cluster. *Nucleic Acids Res* **44**: 2613–2627.
- Lonfat N, Montavon T, Darbellay F, Gitto S, Duboule D. 2014. Convergent evolution of complex regulatory landscapes and pleiotropy at Hox loci. *Science* **346**: 1004–1006.
- Luo Z, Gao X, Lin C, Smith ER, Marshall SA, Swanson SK, Florens L, Washburn MP, Shilatifard A. 2015. Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Mol Cell* **57**: 685–694.
- Maamar H, Cabili MN, Rinn J, Raj A. 2013. linc-HOXA1 is a non-coding RNA that represses Hoxa1 transcription in cis. *Genes Dev* **27**: 1260–1271.
- Mallo M, Wellik DM, Deschamps J. 2010. Hox genes and regional patterning of the vertebrate body plan. *Dev Biol* **344**: 7–15.
- Marks H, Kalkan T, Menafrà R, Denissov S, Jones K, Hofemeister H, Nichols J, Kranz A, Stewart AF, Smith A, et al. 2012. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**: 590–604.
- McMillan DR, Christians E, Forster M, Xiao X, Connell P, Plummer JC, Zuo X, Richardson J, Morgan S, Benjamin IJ. 2002. Heat shock transcription factor 2 is not essential for embryonic development, fertility, or adult cognitive and psychomotor function in mice. *Mol Cell Biol* **22**: 8005–8014.
- Montavon T, Soshnikova N, Mascréz B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F, Duboule D. 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**: 1132–1145.
- Morey C, Da Silva NR, Perry P, Bickmore WA. 2007. Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation. *Development* **134**: 909–919.
- Morrison A, Moroni MC, Ariza-McNaughton L, Krumlauf R, Mavilio F. 1996. In vitro and transgenic analysis of a human HOXD4 retinoid-responsive enhancer. *Development* **122**: 1895–1907.
- Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. 2015. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**: 1017–1021.
- Neijts R, Amin S, van Rooijen C, Tan S, Creyghton MP, de Laat W, Deschamps J. 2016. Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos. *Genes Dev* **30**: 1937–1942.
- Noordermeer D, Leleu M, Splinter E, Rougemont J, De Laat W, Duboule D. 2011. The dynamic architecture of Hox gene clusters. *Science* **334**: 222–225.
- Noordermeer D, Leleu M, Schorderet P, Joye E, Chabaud F, Duboule D. 2014. Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. *Elife* **3**: e02557.
- Piunti A, Shilatifard A. 2016. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* **352**: aad9780.
- Popperl H, Bienz M, Studer M, Chan SK, Aparicio S, Brenner S, Mann RS, Krumlauf R. 1995. Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. *Cell* **81**: 1031–1042.
- Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR–Cas9 system. *Nat Protoc* **8**: 2281–2308.
- Rickels R, Hu D, Collings CK, Woodfin AR, Piunti A, Mohan M, Herz HM, Kvon E, Shilatifard A. 2016. An evolutionary conserved epigenetic mark of polycomb response elements implemented by Trx/MLL/COMPASS. *Mol Cell* **63**: 318–328.
- Ringrose L, Paro R. 2004. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu Rev Genet* **38**: 413–443.
- Satijn DP, Hamer KM, den Blaauwen J, Otte AP. 2001. The polycomb group protein EED interacts with YY1, and both

- proteins induce neural tissue in *Xenopus* embryos. *Mol Cell Biol* **21**: 1360–1369.
- Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, Hacisuleyman E, Li E, Spence M, et al. 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *Elife* **2**: e01749.
- Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. 2007. Genome regulation by polycomb and trithorax proteins. *Cell* **128**: 735–745.
- Sharpe J, Nonchev S, Gould A, Whiting J, Krumlauf R. 1998. Selectivity, sharing and competitive interactions in the regulation of Hoxb genes. *EMBO J* **17**: 1788–1798.
- Shpargel KB, Sengoku T, Yokoyama S, Magnuson T. 2012. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. *PLoS Genet* **8**: e1002964.
- Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, Jangi M, Giallourakis CC, Sharp PA, Young RA. 2015. Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**: 978–981.
- Smith E, Shilatifard A. 2014. Enhancer biology and enhanceropathies. *Nat Struct Mol Biol* **21**: 210–219.
- Soshnikova N, Duboule D. 2009. Epigenetic temporal control of mouse Hox genes in vivo. *Science* **324**: 1320–1323.
- Studer M, Popperl H, Marshall H, Kuroiwa A, Krumlauf R. 1994. Role of a conserved retinoic acid response element in rhombomere restriction of Hoxb-1. *Science* **265**: 1728–1732.
- van de Werken HJ, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E, de Laat W. 2012a. 4C technology: protocols and data analysis. *Methods Enzymol* **513**: 89–112.
- van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, et al. 2012b. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods* **9**: 969–972.
- Vella P, Barozzi I, Cuomo A, Bonaldi T, Pasini D. 2012. Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Res* **40**: 3403–3418.
- Wang P, Lin C, Smith ER, Guo H, Sanderson BW, Wu M, Gogol M, Alexander T, Seidel C, Wiedemann LM, et al. 2009. Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol Cell Biol* **29**: 6074–6085.
- Wellik DM. 2007. Hox patterning of the vertebrate axial skeleton. *Dev Dyn* **236**: 2454–2463.
- Wu M, Wang PF, Lee JS, Martin-Brown S, Florens L, Washburn M, Shilatifard A. 2008. Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol Cell Biol* **28**: 7337–7344.
- Yin Y, Yan P, Lu J, Song G, Zhu Y, Li Z, Zhao Y, Shen B, Huang X, Zhu H, et al. 2015. Opposing roles for the lincRNA Haunt and its genomic locus in regulating HOXA gene activation during embryonic stem cell differentiation. *Cell Stem Cell* **16**: 504–516.
- Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A. 2008. The ground state of embryonic stem cell self-renewal. *Nature* **453**: 519–523.
- Zakany J, Duboule D. 2007. The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* **17**: 359–366.
- Zhong X, Jin Y. 2009. Critical roles of coactivator p300 in mouse embryonic stem cell differentiation and Nanog expression. *J Biol Chem* **284**: 9168–9175.