JBC RESEARCH ARTICLE

CTCF-binding element regulates ESC differentiation *via* orchestrating long-range chromatin interaction between enhancers and HoxA

Received for publication, November 10, 2020, and in revised form, January 19, 2021 Published, Papers in Press, February 11, 2021, https://doi.org/10.1016/j.jbc.2021.100413

Guangsong Su, Wenbin Wang, Jun Chen, Man Liu, Jian Zheng, Dianhao Guo, Jinfang Bi, Zhongfang Zhao, Jiandang Shi, Lei Zhang*, and Wange Lu*

From the College of Life Sciences, Nankai University, Tianjin, China

Edited by Ronald Wek

Proper expression of Homeobox A cluster genes (*HoxA*) is essential for embryonic stem cell (ESC) differentiation and individual development. However, mechanisms controlling precise spatiotemporal expression of *HoxA* during early ESC differentiation remain poorly understood. Herein, we identified a functional CTCF-binding element (CBE^{+47}) closest to the 3′-end of *HoxA* within the same topologically associated domain (TAD) in ESC. CRISPR-Cas9-mediated deletion of CBE^{+47} significantly upregulated *HoxA* expression and enhanced early ESC differentiation induced by retinoic acid (RA) relative to wild-type cells. Mechanistic analysis by chromosome conformation capture assay (Capture-C) revealed that CBE^{+47} deletion decreased interactions between adjacent enhancers, enabling formation of a relatively loose enhancer–enhancer interaction complex (EEIC), which overall increased interactions between that EEIC and central regions of *HoxA* chromatin. These findings indicate that CBE^{+47} organizes chromatin interactions between its adjacent enhancers and *HoxA*. Furthermore, deletion of those adjacent enhancers synergistically inhibited *HoxA* activation, suggesting that these enhancers serve as an EEIC required for RA-induced *HoxA* activation. Collectively, these results provide new insight into RA-induced *HoxA* expression during early ESC differentiation, also highlight precise regulatory roles of the CTCF-binding element in orchestrating high-order chromatin structure.

Embryonic stem cell (ESC), which are derived from the inner cell mass (ICM) of blastocysts, exhibit unlimited self-renewal and differentiation into the three embryonic germ layers (1–3). Therefore, ESC are of great utility to the fields of regenerative medicine and tissue engineering (4, 5). ESC pluripotency requires expression of master control transcription factors such as *Oct4*/*Nanog*/*Sox2* (6, 7), while expression of *HoxA*/*Gata4*/*Pax6* genes is essential for ESC differentiation (8–10). Recently, studies indicate that high-order chromatin structures play a critical role in regulating gene expression and maintaining ESC status (11, 12). Numerous studies report that topologically associated domains (TADs) are the basic unit of nuclear chromatin (13, 14). Functional DNA elements, such as enhancers, within the same TAD are more likely to promote target gene expression *via* direct long-range chromatin interactions (15, 16). Long-range chromatin interactions require proteins to mediate those interactions, among them CTCF. Using auxin-induced degradation of CTCF protein, Nora *et al.* (17) found that CTCF loss promoted disappearance of TADs, revealing CTCT as essential for maintaining TAD formation. CTCF forms a chromatin loop through directly binding CTCF-binding elements (CBEs) enriched at TAD boundaries and relies on binding polarity to maintain TAD formation (18–20). Although many report that CBEs at TAD boundaries play key regulatory roles in biological development (21), analysis of CBE activity inside a TAD is lacking.

Retinoic acid (RA) rapidly induces ESC differentiation (22, 23) and activates *HoxA* expression (24, 25). Recent studies show that CBE activity functions in tumorigenesis (26, 27), immunogenesis (28, 29), and stem cell development (21, 30) by controlling target gene expression through organizing high-order chromatin structures. We and others have found that multiple enhancers are required for RA-induced *HoxA* expression and early ESC differentiation *via* long-range chromatin interactions with *HoxA* genes (31–34). Previous studies have also reported many regulatory patterns for *HoxA* activation during ESC differentiation; for example, several CBEs (also known as CBS or CTCF-binding sites) within the *HoxA* locus can organize *HoxA* chromatin structures and play significant regulatory roles in ESC differentiation and leukemogenesis (26, 35, 36). However, it is unknown whether CBEs function to regulate *HoxA* expression by orchestrating enhancer chromatin structures during RA-induced early ESC differentiation.

Here, we identified a novel functional CBE^{+47} required for RA to activate *HoxA* expression and promote early ESC differentiation. Relevant to mechanism, we report that CBE^{+47} regulates *HoxA* expression by organizing precise chromatin interactions between its adjacent enhancers and *HoxA*. Our study increases understanding of mechanisms underlying RA-induced *HoxA* activation and early ESC differentiation and highlights the unique and precise regulatory roles of CBE in high-order chromatin structure.

* For correspondence: Lei Zhang, joyleizhang@nankai.edu.cn; Wange Lu, wangelv@gmail.com.

Results

Identification of a CTCF-binding element (CBE^{+47}) closest to the 3′-end of HoxA within the same TAD in ESC

TADs are basic structural units in high-order chromatin, and DNA functional elements such as enhancers or insulators in the same TAD likely regulate expression of neighboring genes (15, 16). To identify functional elements that regulate *HoxA* expression, we first analyzed Hi-C data in ESC. We found that the *HoxA* 5′-end was located at a TAD boundary region, while the 3′-end was within the TAD (Fig. 1*A*). Others had shown that CBEs located at the *HoxA* locus regulate *HoxA* expression and ESC differentiation (36, 37). Interestingly, we identified many significant CBEs in the same TAD at the *HoxA* 3′-end (Fig. 1*B*), but their function was not previously characterized.

Here, we focused on a significant CBE closest to the *HoxA* 3′-end, ∼47 kb from *HoxA* and located in the second intron of *Halr1*; we designate that element CBE^{+47} (Fig. 1*B* and Fig. S1). We observed significant binding of MED1, MED12, and YY1 upstream of CBE^{+47} (Fig. 1*B*) as well as enrichment for the epigenetic modification-related markers H3K27ac and H3K4me1/2/3 (Fig. 1*C*). We also observed significant chromatin accessibility (based on DNase and ATAC-seq analysis) and transcriptional activity marked by PolIIenrichment in these regions (Fig. 1*D*). These data indicate that CBE^{+47} is located between *HoxA* and potential regulatory elements. Others had shown that CBEs regulate target gene expression

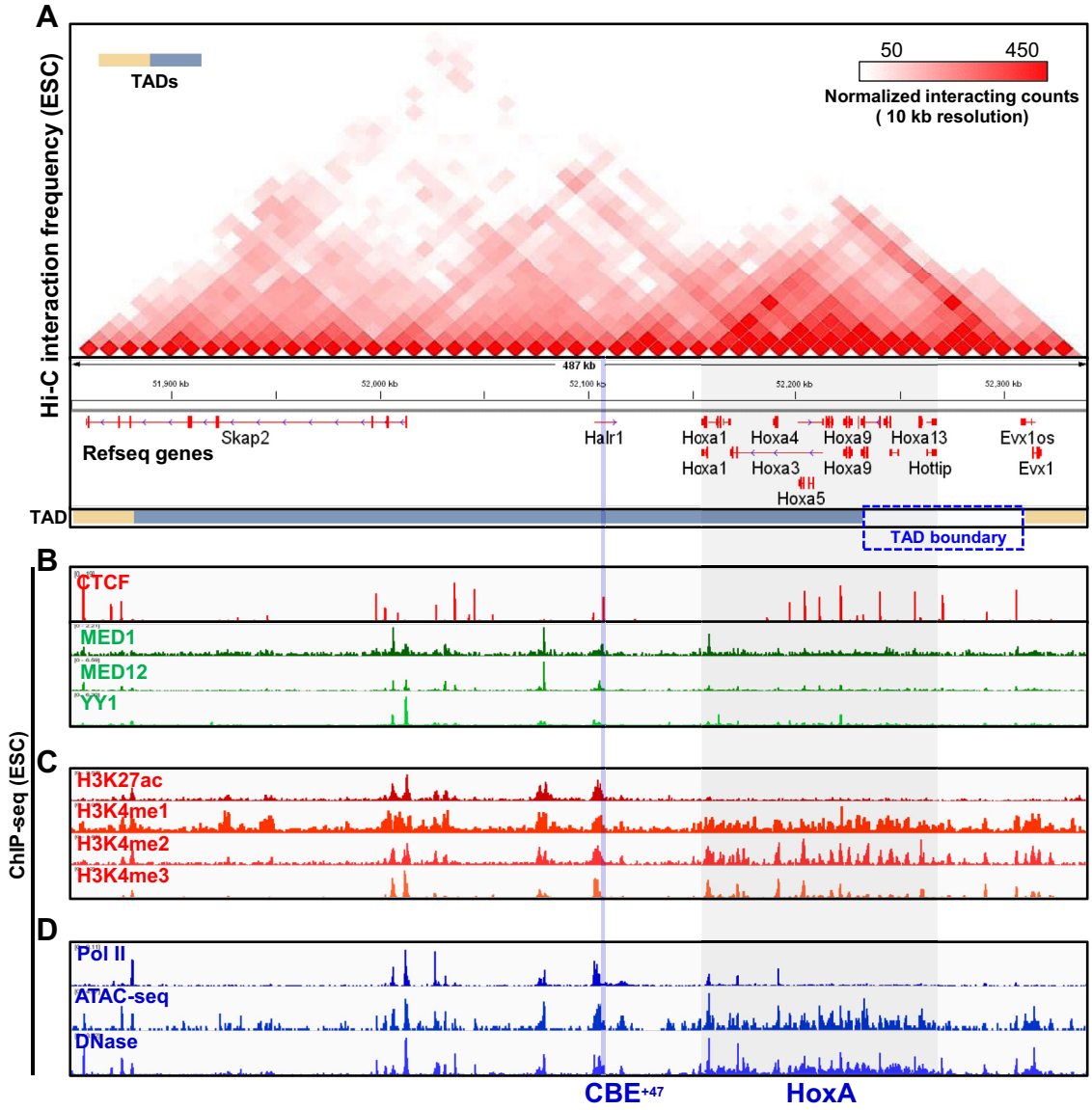

Figure 1. Identification of a CBE (CBE^{+47}) closest to the *HoxA* 3′-end within the same TAD in ESC. *A*, Hi-C interaction map of ∼0.5 Mb region surrounding the *HoxA* locus in undifferentiated ESC. Data were extracted from Bonev *et al.*, 2017. *B*, IGV (Integrative Genomics Viewer) screenshots showing gene tracks of CTCF, MED1, MED12, and YY1 ChIP-seq signal occupancy at *Skap2* and *HoxA* loci in ESC. CTCF ChIP-seq shows CTCF binding throughout the locus. Multiple CTCF sites are located downstream of the *HoxA* locus. *C* and *D*, IGV view of selected ChIP-seq tracks at *Skap2* and *HoxA* loci in ESC. Shown are H3K27ac, H3K4me1, H3K4me2, H3K4me3 (*C*), and Pol II, ATAC-seq and DNase (*D*). *Vertical blue line* indicates the CBE^{+47} region. *Gray shadowing* indicates *HoxA* locus. In (*A*), *blue dotted box area* shows the TAD boundary region. Other results relevant to these findings are shown in Fig. S1.

by organizing chromatin interactions between a functional element and a target gene (38, 39). Therefore, we postulated that CBE^{+47} may play a regulatory role.

Gene expression analysis in cultured WT and CBE^{+47}-deleted cells grown in self-renewal conditions

To assess CBE^{+47} function, we employed CRISPR-Cas9 knockout methodology (31). Specifically, we designed two sgRNAs, upstream and downstream of CBE^{+47} (Fig. 2*A*) and after Cas9 cleavage and DNA recombination, obtained two homozygous CBE^{+47} knockout lines (CBE^{+47}-KO). PCR of genomic DNA using specific primers and Sanger sequencing confirmed CBE^{+47} deletion (Fig. 2*B*). AP staining of CBE^{+47} knockout lines revealed no significant differences in cell morphology relative to WT cells in self-renewing culture conditions (Fig. 2*C*). Moreover, expression of *Halr1* and *Sakp2*

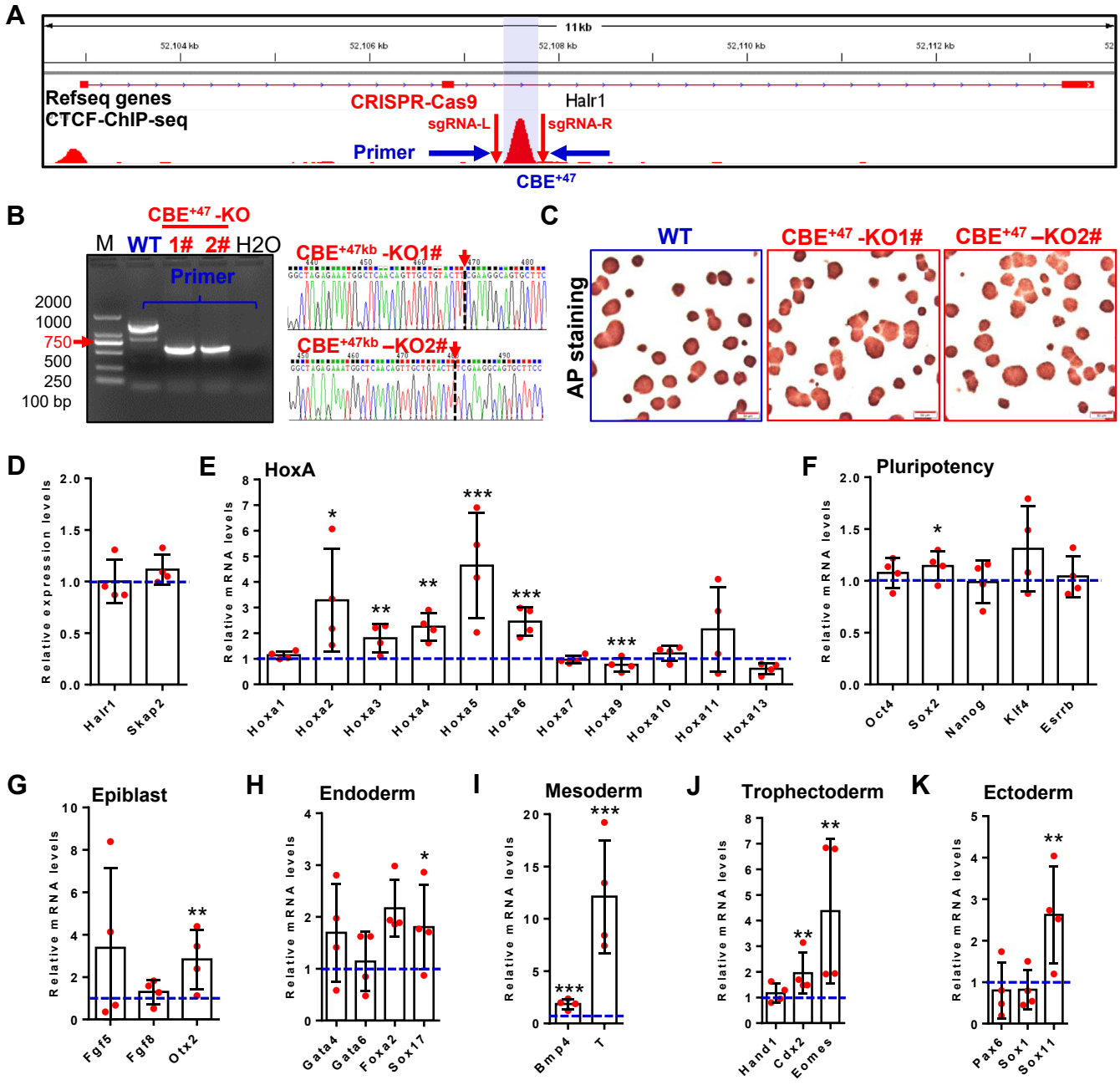

Figure 2. Gene expression analysis in WT and CBE^{+47}-deleted cells grown in self-renewal culture conditions. *A*, schematic showing CRISPR/Cas9-mediated deletion of CBE^{+47} (*blue shadowing*) using two sgRNAs. Indicated primers shown in *blue* were used to distinguish CBE^{+47}-KO (CBE^{+47} knockout) from wild-type (WT) clones. *B*, validation of knockout lines by genomic DNA PCR and sequencing. *Left*, images of agarose gel showing PCR analysis of representative clones. *Right*, DNA sequencing of CBE^{+47}-KO cell clones #1 and #2 using the indicated primer. *C*, appearance of WT and CBE^{+47}-KO cells from lines 1 and #2 stained with alkaline phosphatase (AP). *D–K*, data derived from qRT-PCR analysis of WT and CBE^{+47}-KO cells showing indicated transcripts in undifferentiated ESC. Expression levels in WT ESC were set to 1 and shown as the blue dotted line. Data are represented as means ± SD. Significance was based on Student's *t*-test (two-tailed; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). In all analyses, n = 4 (two CBE^{+47}-KO lines (#1 and #2) and two biological replicates per line). In (*B*): M, DNA marker. In (*C*): scale bar, 50 μm.

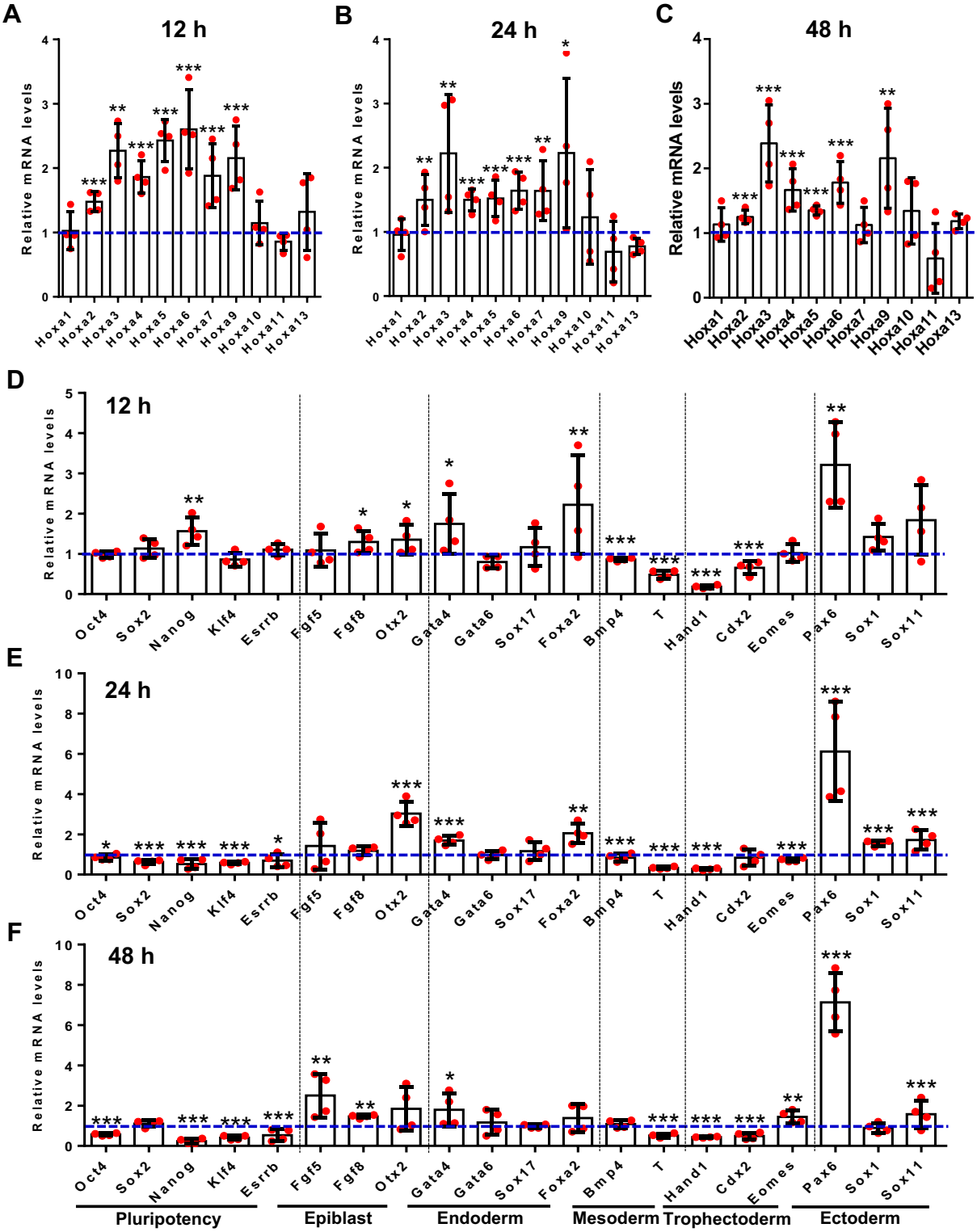

Figure 3. CBE^{+47} deletion potentiates RA-induced *HoxA* expression and perturbs ESC differentiation. *A–C*, qRT-PCR of WT and CBE^{+47}-KO cells showing *HoxA* gene expression in ESC after RA treatment for 12 h (*A*), 24 h (*B*), and 48 h (*C*). *D–F*, qRT-PCR of WT and CBE^{+47}-KO ESC showing transcripts of indicated genes in ESC after RA treatment for 12 h (*D*), 24 h (*E*), and 48 h (*F*). In *A–F*, expression levels in WT ESC were set to 1 and shown as the *blue dotted line*. Data are represented as means ± SD. Significance is based on Student's *t*-test (two-tailed; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). In all analyses, n = 4 (two CBE^{+47}-KO lines (#1 and #2) and two biological replicates per line). Other results relevant to this figure are shown in Figs. S2 and S3.

mRNAs was comparable in WT and KO lines, based on qRT-PCR analysis (Fig. 2*D*). However, most of the 3′-end genes of *HoxA*, including *Hoxa2-a6*, were significantly upregulated, while the TAD boundary gene *Hoxa9* was significantly downregulated in CBE^{+47}-KO relative to WT cells (Fig. 2*E*). These data show that CBE^{+47} deletion significantly changes *HoxA* expression and indicate that even in an undifferentiated ESC state, CBE^{+47} is required to maintain proper *HoxA* expression.

Given the critical function of *HoxA* in ESC differentiation (8, 32, 40) and the observation that CBE^{+47} deletion significantly increased expression of a subset of *HoxA* genes, we asked whether ESC pluripotency was perturbed by CBE^{+47} deletion by evaluating expression of pluripotency- and differentiation-related genes by qRT-PCR (Fig. 2, *F–K*). We observed no significant differences in expression of pluripotency-regulated genes in CBE^{+47}-KO *versus* WT cells, except for slight upregulation of *Sox2*. However, mesoderm and trophectoderm genes were significantly upregulated in CBE^{+47}-KO cells. Expression of other germ layer genes such as *Otx2* (epiblast), *Sox17* (endoderm), and *Sox11* (ectoderm) also significantly increased in CBE^{+47}-KO cells. These results indicate that CBE^{+47} deletion does not regulate ESC pluripotency but may alter its differentiation capacity.

CBE^{+47} deletion significantly potentiates RA-induced HoxA expression and promotes abnormal differentiation of early ESC

During early RA-induced differentiation of ESC, *HoxA* genes become significantly activated (25, 32) (Fig. S2). Therefore, we asked whether CBE^{+47} regulates those activities. To do so, we treated CBE^{+47}-KO and WT ESC with RA for 12, 24, and 48 h (Fig. S3*A*) and then compared *HoxA* expression levels by qRT-PCR. *HoxA* 3′-end genes (*Hoxa2-a6*) showed significantly higher expression in CBE^{+47}-KO cells, as did the TAD boundary genes *Hoxa7* and *Hoxa9*. However, expression of *HoxA* 5′-end genes (*Hoxa10*, *Hoxa11*, and *Hoxa13*) was comparable in both genotypes (Fig. 3, *A–C*). Also, antisense long noncoding RNAs (such as *Hoxaas3*) at the *HoxA* locus were significantly upregulated in CBE^{+47}-KO relative to WT cells (Fig. S3, *B–D*). These data suggest overall that CBE^{+47} functions to restrict RA-induced overexpression of *HoxA*.

To assess whether CBE^{+47} knockout alters ESC differentiation, we used qRT-PCR to analyze potential changes in pluripotency- or differentiation-related master control genes in WT and CBE^{+47}-KO ESC under RA induction conditions. Expression of pluripotency genes (*Nanog* and *Klf4*) did not change significantly at 12 h after ESC differentiation but significantly decreased in CBE^{+47}-KO relative to WT cells by 24 and 48 h. The epiblast master control gene *Otx2*, endoderm master control gene *Gata4*, and ectoderm master control gene *Pax6* showed significantly higher expression levels in CBE^{+47}-KO *versus* WT cells, while the mesoderm master gene *T* and the trophectoderm master gene *Hand1* were significantly downregulated (Fig. 3, *D* and *E*). These findings suggest overall that CBE^{+47} is essential to maintain RA-induced *HoxA* expression and proper early ESC differentiation.

Transcriptome analysis of WT and CBE^{+47kb}-deleted cells during RA-induced early ESC differentiation

To further assess effects of CBE^{+47} loss on early (24 h) RA-induced ESC differentiation, we performed transcriptome analysis using RNA-seq after RA-induced differentiation of WT and CBE^{+47}-KO cells. Relative to WT cells, statistical analysis of CBE^{+47}-KO cells revealed 869 genes upregulated and 360 downregulated (fold-change ≥ 2, $p < 0.05$) (Fig. 4*A*). Among them, the most significantly upregulated genes were at the 3′-end of *HoxA* (Fig. 4*B*). Because these genes located in the same TAD with CBE^{+47}, we hypothesized that CBE+47 deletion significantly promotes gene expression restricted to that TAD. These findings reveal that CBE^{+47} serves as a *cis*-regulatory element playing a local regulatory role (Fig. S4). Moreover, relative to WT cells, CBE^{+47} KO cells expressed significantly higher levels of *HoxB/C/D* cluster genes (Fig. S5*A*). Genes relevant to RA signaling, such as *Crabp2*, *Cyp26a1*, and *Stra6*, were also significantly overexpressed in CBE^{+47} KO cells, as were the differentiation-related genes *Cbx4* and *Foxa1* (Fig. S5*B*). By contrast, pluripotency genes such as *Myc*, *Lin28a*, and *Etv4* were significantly suppressed in CBE^{+47} KO relative to WT cells (Fig. S5*C*). These data suggest that in normal cells CBE^{+47} functions to enable proper differentiation.

To assess mechanisms underlying these activities, we performed bioinformatic analysis of up- and downregulated genes in CBE^{+47} KO cells (Fig. S6). GO analysis indicated that biological processes (BP) relevant to upregulated genes include gland development and positive regulation of nervous system development (Fig. 4*C*). GO-BPs of downregulated genes mainly involved embryonic placenta development, positive regulation of vasoconstriction, and epithelial cell differentiation (Fig. 4*D*). KEGG pathway enrichment analysis revealed upregulated genes to be enriched in pathways including *cGMP-PKG* signaling, neuroactive ligand–receptor interaction, axon guidance, stem cell pluripotency, *Hippo* signaling, *cAMP* signaling, *Wnt* and Hedgehog signaling (Fig. S7*A*). Enrichment results relevant to downregulated genes revealed pathways related to proteoglycans in cancer, transcriptional dysregulation in cancer, arachidonic acid metabolism, HTLV-I infection, *Hippo* signaling, and thyroid hormone signaling (Fig. S7*A*). These findings, which were further validated by Gene Set Enrichment Analysis (GSEA) (Fig. 4*E* and Fig. S7*B*), suggest that CBE^{+47} deletion promotes abnormal gene expression and perturbs proper RA-induced early ESC differentiation.

CBE^{+47} is required for proper chromatin interactions between adjacent enhancers and HoxA

Previous studies report that CTCF regulates target gene expression by controlling chromatin interaction between enhancers and target genes and that this activity occurs within a TAD (41). Studies by us and others have identified three enhancers (E1, E2, and E3) at the *HoxA* 3′-end (Fig. 5*A*) required

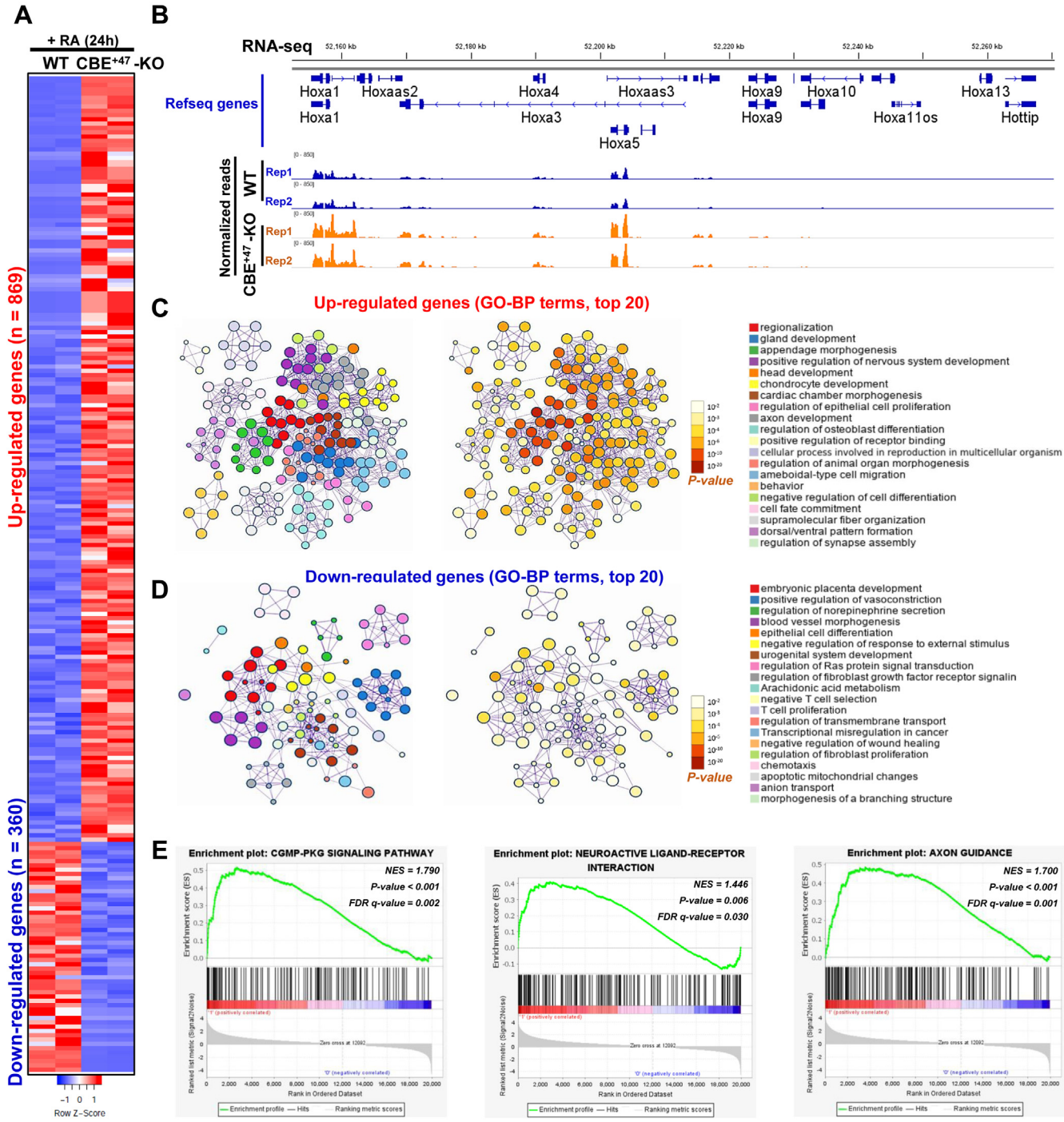

Figure 4. Transcriptome analysis in WT and CBE^{+47}-deleted cells following RA-induced early ESC differentiation. *A*, heatmap depicting gene expression changes in WT and CBE^{+47}-KO cells treated 24 h with RA (Fold-change ≥ 2; $p < 0.05$, as determined by DESeq2). *B*, RNA-seq results shown heatmap of *HoxA*. *C* and *D*, GO-BP analyses indicating differentially expressed genes. *E*, GSEA plots for the top three KEGG signaling pathways showing log2 fold change for the entire transcriptome. NES, Normalized Enrichment Score. Other results related to this figure are shown in Figs. S4–S7.

for *HoxA* expression following RA-induced ESC differentiation (31–34). Among them, E1 and E3 reportedly mediate significant long-range interactions with *HoxA* chromatin. Remarkably, CBE^{+47} is located between those three enhancers and *HoxA* (Fig. 5*A*). Therefore, we speculate that the CBE^{+47} may be involved in organizing the interactions between the three enhancers and *HoxA* chromatin.

Next we used Capture-C methodology, which is widely used to define chromatin interactions between enhancers and target genes (31, 42–46), to define interactions between E1, E2, and E3 and *HoxA* chromatin in WT and CBE^{+47}-KO cells. For this analysis we used cells induced 24 h with RA, with enhancers E1, E2, and E3 as respective anchor regions (Fig. 5, *B–D*). In WT cells, we observed

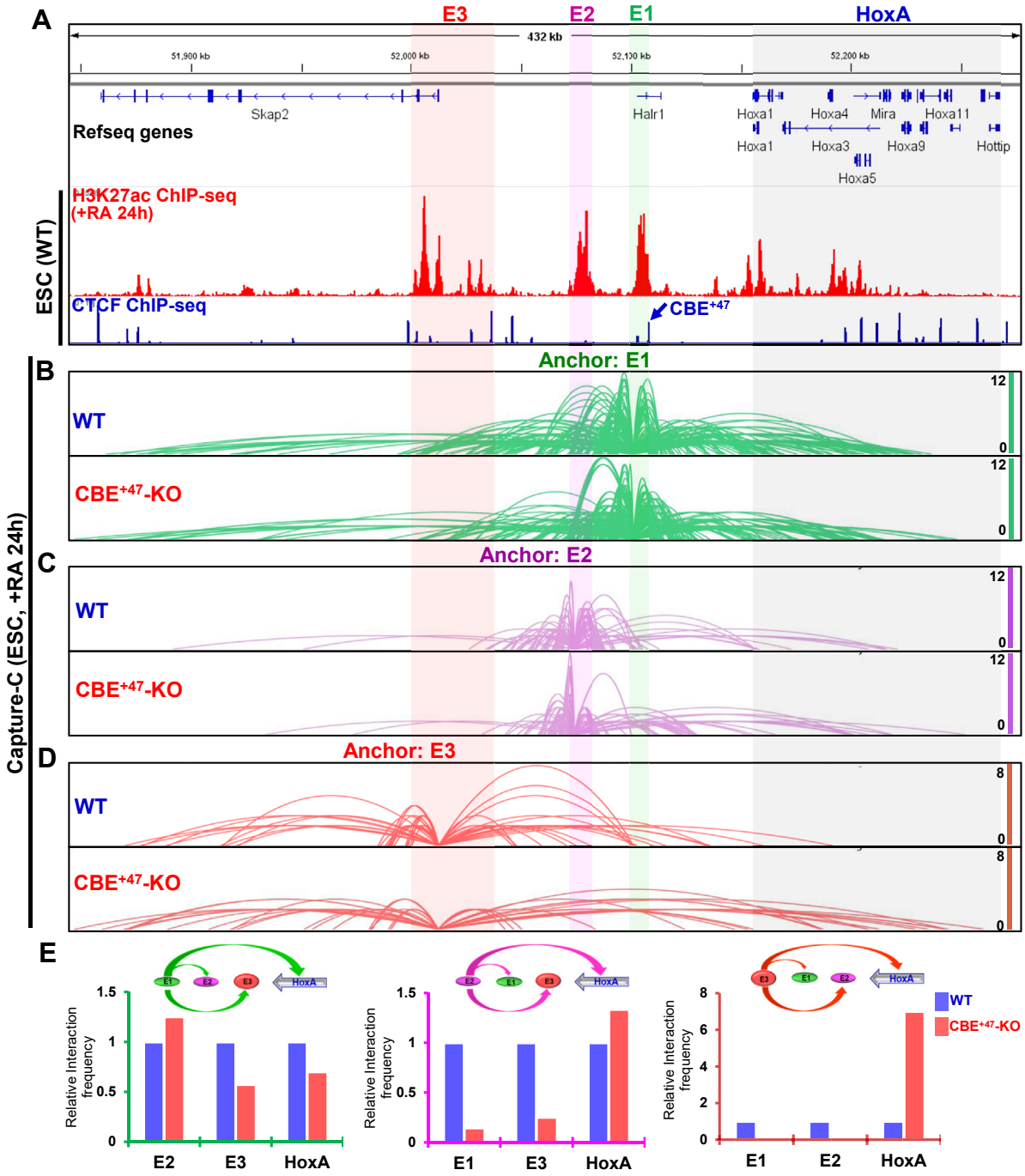

Figure 5. CBE^{+47} is required for proper chromatin interactions between *HoxA* and adjacent enhancers. *A*, IGV (Integrative Genomics Viewer) screenshots showing gene tracks of H3K27ac (+RA, 24 h) and CTCF ChIP-seq signal occupancy at *Skap2* and *HoxA*loci in WT ESC. *B–D*, chromatin interaction profiles using indicated enhancers as anchors in WT or CBE^{+47}-KO cells after 24 h of RA treatment. Numbers on Y-axis denote interaction frequency. *Light green*, *light purple*, *light red*, and *light gray shadows* mark E1, E2, E3, and *HoxA* regions, respectively. *E*, quantitative results showing interaction differences seen in WT and CBE^{+47}-KO cells.

significant interactions between all three enhancers and *HoxA* chromatin, with E1 interacting most strongly (Fig. 5*B*). However, in CBE^{+47} KO cells, interactions between the E1 and *HoxA* chromatin were significantly reduced relative to WT cells. However, E2 and E3 interaction with *HoxA* chromatin increased significantly in CBE^{+47} KO relative to WT cells, and the intensity of chromatin interactions shifted from the 3'-end to the 5'-end of *HoxA* (Fig. 5, *C* and *D*). In addition, the results also show that the interaction between enhancers decreases significantly except for the slight increase of E1 and E2 (Fig. 5*E*). These results indicate overall that CBE^{+47} is essential to maintain these proper chromatin interactions during RA-induced early ESC differentiation.

Multiple enhancer deletions show synergistic effect on RA-induced HoxA expression

Next, we used CRISPR-Cas9 to delete enhancers individually or in groups to establish multiple homozygous enhancer knockout lines (namely, ΔE1, ΔE2, ΔE3, ΔE1/2, ΔE1/2/3) (Fig. S8). We then treated each line 24 h with RA and performed qRT-PCR to determine *HoxA* expression (Fig. 6, *A–E*). ΔE1 and ΔE2 cells showed similar phenotypes in terms of *HoxA* regulation, that is, both 3'-end (*Hoxa1*) and central (*Hoxa5-a10*) genes were significantly suppressed relative to WT cells (Fig. 6, *A* and *B*). Also relative to WT cells, in ΔE3 cells, *Hoxa1-a7* were significantly inhibited, indicating that these enhancers have specific and unique regulatory effects (Fig. 6*C*). In ΔE1/2 and ΔE1/2/3 cells, *HoxA* expression was significantly reduced compared with ΔE1, ΔE2, or ΔE3 (Fig. 6, *D* and *E* and Fig. S9, *A* and *B*). In addition, considering the closest distance between E1 and CBE^{+47}, we also compared the expression differences of *HoxA* cluster gene between them. The results showed that expressions of *HoxA* 3'-end gene ware significantly different (Fig. S9*C*), which indicated that the regulation of *HoxA* by CBE^{+47} and E1 was opposite. Over all, these findings indicate that interactions among E1, E2, and E3 enhancers with *HoxA* chromatin synergize to promote RA-induced *HoxA* activation.

Discussion

Precise expression of *HoxA* genes is critical for ESC differentiation. Although others have shown that CBEs play an important role in ESC differentiation by regulating high-order chromatin structure (36, 47), identification of a functional CBE and mechanisms regulating *HoxA* expression remain largely unknown. Here, we identified a new functional CBE^{+47} and propose a novel model based on our results. Specifically, in WT cells (Fig. 7, *left*), three enhancers near CBE^{+47} act as an EEIC to interact with the 3'-end of *HoxA* chromatin to maintain normal *HoxA* expression and promote correct early

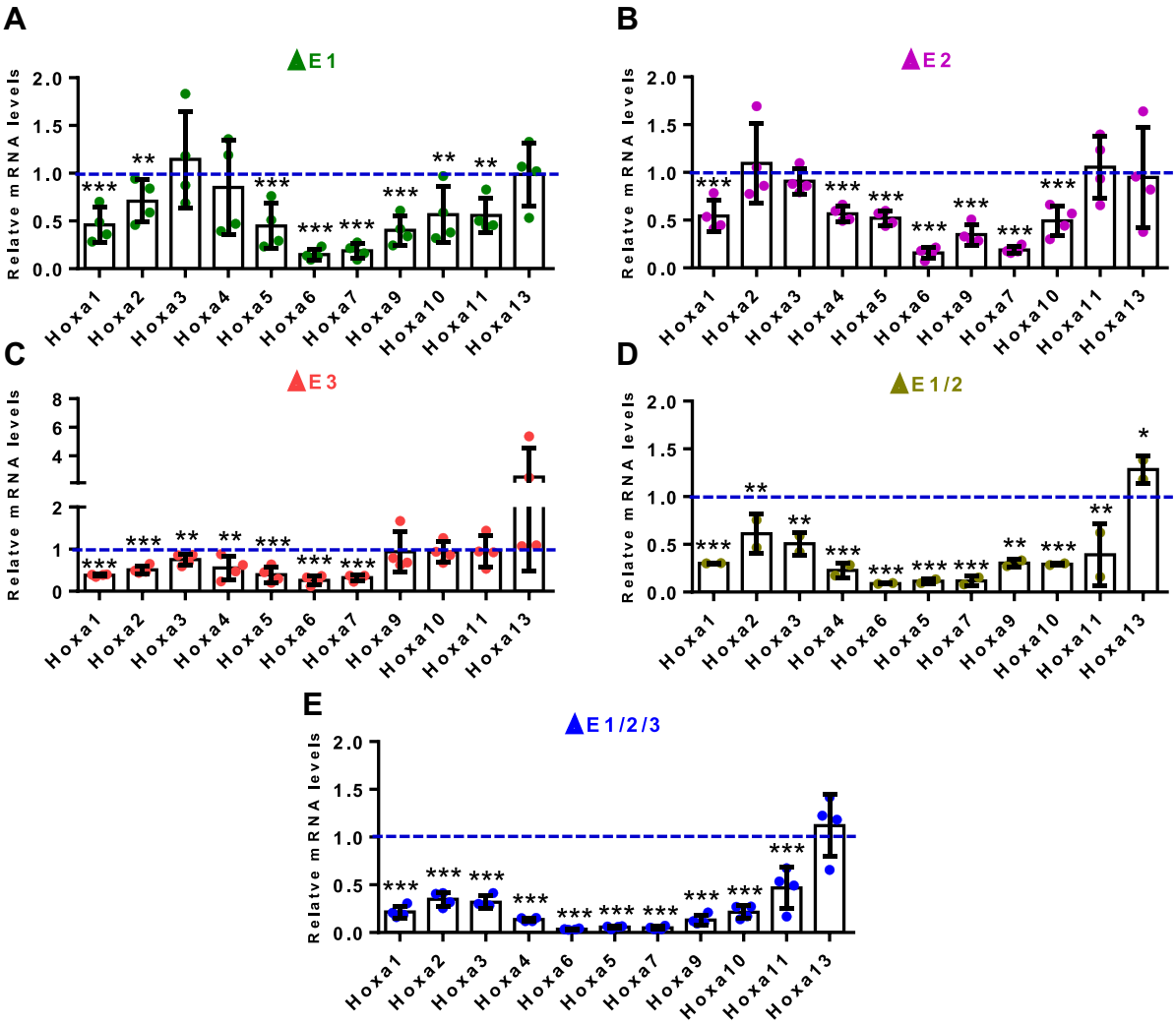

Figure 6. Deletions of multiple enhancers synergize to alter RA-induced *HoxA* gene expression. *A–E*, transcripts of *HoxA* genes were measured by qRT–PCR and normalized to WT levels in ΔE1 (*A*), ΔE2 (*B*), ΔE3 (*C*), ΔE1/2 (*D*), andΔE1/2/3 (*E*) cells following 24 h of RA treatment. Expression levels in WT ESC were set to 1 and shown as the *blue dotted line*. Data are represented as means ± SD. Significance is based on Student's *t*-test (two-tailed; **p* < 0.05, ***p* < 0.01, ****p* < 0.001.) In *A–C* and *E*, n = 4, including two enhancer knockout lines and two biological replicates per line. In *D*, n = 2, including one enhancer knockout line and two biological replicates per line. Other results relevant to these are shown in Figs. S8 and S9.

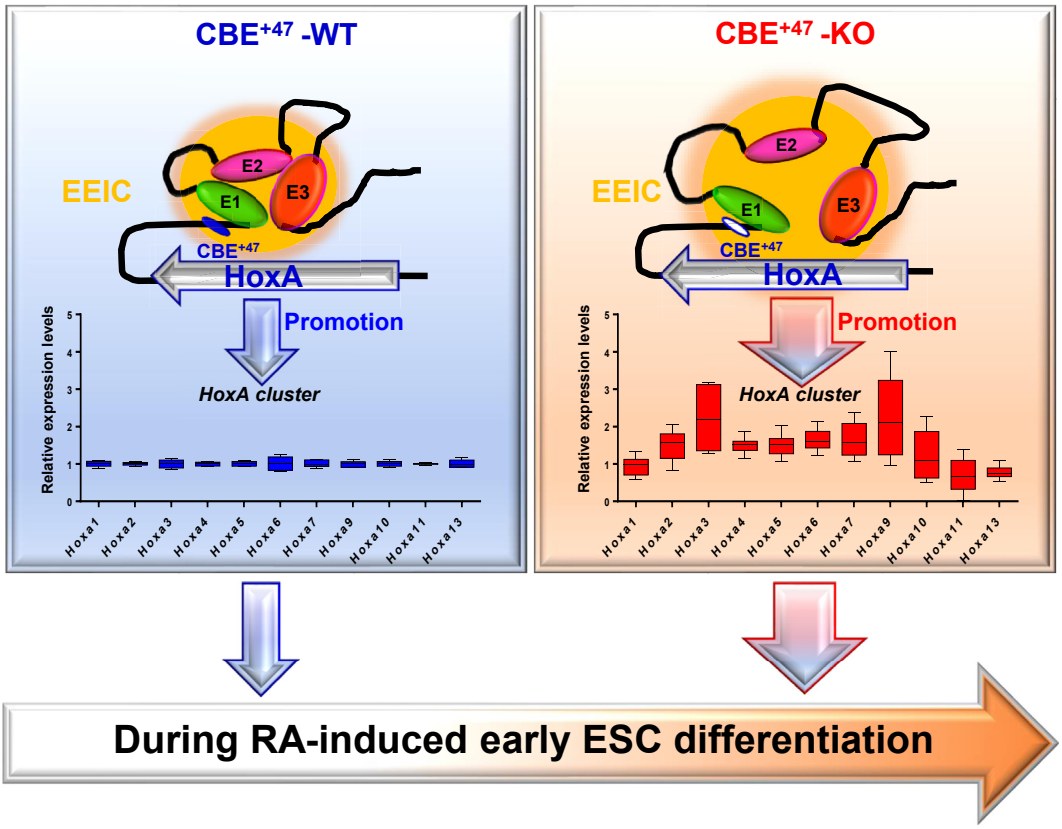

Figure 7. Schematic showing how CBE^{+47} regulates RA-induced *HoxA* expression and early ESC differentiation by orchestrating long-range chromatin interactions between *HoxA* and adjacent enhancers. *Left*, model showing CBE^{+47} activity in WT cells. Three enhancers near CBE^{+47} act as an EEIC and interact with the 3′-end of *HoxA* chromatin to maintain normal *HoxA* normal expression, allowing proper ESC differentiation after RA treatment. *Right*, in CBE^{+47}-KO cells interactions between enhancers decrease, resulting in a relatively loose EEIC and allowing increased interactions between that EEIC and the central region of *HoxA* chromatin. As a result, *HoxA* is overexpressed and early ESC differentiation proceeds abnormally.

differentiation of ESC induced by RA. In CBE^{+47}-KO cells (Fig. 7, *right*), enhancer interactions decrease, resulting in a relatively loose EEIC. Furthermore, EEIC interactions with central *HoxA* chromatin regions increase following CBE^{+47} loss, promoting *HoxA* overexpression. These results indicate that a functional CBE^{+47} maintains normal *HoxA* expression during ESC differentiation by organizing precise interactions between adjacent enhancers and *HoxA* chromatin. Our findings highlight indispensable fine-regulatory roles of CBEs as functional elements in high-order chromatin structure and reveal a direct regulatory effect of the CBE^{+47} on RA-induced *HoxA* expression and early ESC differentiation.

Previously, Ferraiuolo *et al.* (48) showed that *HoxA* exhibits a special chromatin structure required for its expression and that CTCF regulates formation of that structure and thus gene expression. Recent studies focusing on CBEs within the *HoxA* locus, including the TAD boundary, have found that CBEs organize *HoxA* chromatin structure, maintain normal *HoxA* expression, and regulate ESC differentiation (36). Narendra *et al.* (36) showed that CBE5/6 knockout significantly promoted *HoxA* expression, altered *HoxA* chromatin structure, and enhanced ESC neural differentiation. Studies in other cells, such as AML cells, also reveal that CBEs located at the *HoxA* locus significantly regulate *HoxA* expression and affect tumorigenesis (26, 35). Here, we find that besides several CBEs in the *HoxA* locus, there are several CBEs downstream of *HoxA* within the same TAD (Fig. 1*B*). Among them, CBE^{+47} KO significantly enhanced RA-induced *HoxA* expression and promoted early ESC differentiation, strongly suggesting that CBE^{+47} is a functional CBE. However, other CBEs have not yet been evaluated and their potential activity requires future analysis. Recently, others developed a CRISPR-Cas9-based genetic screen method to detect functional CBEs in breast cancer and AML (49, 50). Comparable approaches may be taken in the future to screen for functional CBEs that regulate ESC differentiation.

Our transcriptome analysis showed that after CBE^{+47} deletion, genes in the same TAD with CBE^{+47} are significantly upregulated relative to genes in nearby TADs (Fig. S4), suggesting that CBE^{+47} is a *cis*-regulatory element that plays a local regulatory role. We also observed upregulation of RA signaling-related genes (*Crabp2*, *Cyp26a1*, *Stra6*) when CBE^{+47} was deleted, while pluripotency-related genes (*Myc* and *Lin28a*) were significantly downregulated (Fig. S5), supporting the idea that CBE^{+47} loss promotes RA-induced early ESC differentiation. Recent studies show that *Lin28a* inhibits *HoxA* expression to maintain limb development (51), while De Kumar *et al.* (40) found that HOXA1 binds to the *Myc* gene near enhancers, suggesting positive feedback between *HoxA* and *Myc* or *Lin28a* in early ESC

differentiation induced by RA. These possibilities require further analysis.

RA significantly induced *HoxA* expression during early ESC differentiation (Fig. S2). Others have shown that at least three enhancers (E1, E2, and E3) are needed for *HoxA* expression in this context (31–34). CBE^{+47} lies between these three enhancers and *HoxA* (Fig. 5*A*). As an insulator, CBE may block interaction between enhancers and target genes, exerting an inhibitory effect (52–54). Our analysis first reveals that the three enhancers interact significantly as an EEIC. Second, we find that the EEIC interacts significantly with *HoxA*. Third, the interaction of these enhancers with *HoxA* also has its own pattern. When CBE^{+47} is deleted, enhancer interactions significantly decrease, resulting in formation of a relatively loose EEIC. We conclude that CBE^{+47} is required to organize enhancer interaction. Following CBE^{+47} loss, interaction of E1 becomes more concentrated in the middle of the *HoxA* locus, and interaction of E2 and E3 with *HoxA* moves from the 3'-end to middle regions. Furthermore, CBE^{+47} deletion significantly promotes higher expression of *Hoxa2-a9*, which corresponds to changes in chromatin interactions between enhancers and *HoxA*. Thus CBE^{+47} not only organizes enhancer interactions but orchestrates interactions between enhancers and *HoxA*, ensuring correct *HoxA* expression. Although the CBE^{+47} organizes high-order chromatin structures and previous studies show that the binding direction of CTCF alters formation of a CTCF-mediated chromatin loop (19), other CBEs may pair with CBE^{+47} to form chromatin loops, a possibility that requires further analysis.

Multiple enhancers reportedly maintain RA-induced *HoxA* expression through long-range chromatin interactions with *HoxA* (31–34). Here, we show that knockout of multiple enhancers synergizes to alter *HoxA* expression (Fig. 6), a finding partially consistent with a previous report (33). Although E1/E2/E3 enhancers are known to have regulatory effects, we also observed other enhancers located in the *Skap2* region in the same TAD with *HoxA* (Fig. 5*A*), although their functional effects are unknown. Furthermore, although CBE^{+47} organizes interactions between E1, E2, and E3 enhancers, we did not observe significant CTCF-binding peaks in the E2 region, suggesting that other factors may regulate enhancer interactions. A recent study using the ChIA-PET method (Chromatin Interaction Analysis by Paired-End Tag sequencing) from Wang *et al.* (55) shows that RARα (Retinoic Acid Receptor alpha) mediates chromatin interactions in AML cells. Activation of target genes by RA signaling requires RAR/RXR (retinoic acid receptor/retinoid X receptor) signaling (56, 57). Remarkably, we also found significant RAR/RXR binding peaks in E1, E2, and *HoxA* regions over the course of RA-induced ESC differentiation (data not shown) (58), suggesting that RAR/RXR regulates these interactions, a possibility that needs further verification. We also found that chromatin interactions between a single enhancer and *HoxA* are different. Thus, mechanisms underlying maintenance of these different interactions require further analysis.

In summary, we identified a new functional CBE^{+47} that regulates RA-induced *HoxA* expression and early ESC differentiation *via* orchestrating long-range chromatin interactions between its adjacent enhancers and *HoxA*. These findings further our understanding of intrinsic mechanisms governing RA-induced early ESC differentiation and highlight specific regulatory roles of CBE in high-order chromatin structure.

Experimental procedures

Embryonic stem cell culture

Mouse ESC E14 were cultured as previously described with some modifications (31). ESC were grown in culture dishes coated with 0.1% gelatin (Sigma, Lot # SLBQ9498V) in Dulbecco's Modified Eagle's Medium (DMEM, Gibco, Lot # 8119213) supplemented with 15% fetal bovine serum (FBS, AusGeneX, Lot # FBS00717-1), 1x nonessential amino acids (100x, Gibco, Lot # 2027433), 1x L-Glutamate (100x, Gibco, Lot # 2085472), 1x Penicillin-streptomycin (100x, Gibco, Lot # 2029632), 50 μM β-mercaptoethanol (Sigma, Cas # 60-24-2), 10 ng/ml LIF (ESGRO, Lot # ESG1107), 1 μM PD0325901 (a MEK inhibitor, MedChemExpress, Lot # HY-10254), and 3 μM CHIR99021 (a GSK inhibitor, MedChemExpress, Lot # HY-10182). The medium was replaced every 1 to 2 days. All cells were maintained at 37 °C in a 5% CO_2 incubator.

Retinoic acid (RA)-induced early ESC differentiation

Cells were gently washed with 1x phosphate buffer saline solution (PBS), dissociated, and plated at an appropriate density on gelatin-coated plates in LIF/2i withdrawal medium supplemented with 2 μm RA (Solarbio, Lot # 1108F031). The culture medium was replaced at the given point in time.

RNA extraction, reverse transcription, and quantitative real-time PCR (qRT-PCR)

Total RNA was extracted from differentiated or undifferentiated ESC using TRIzol Reagent (Life Technologies, Lot # 213504). cDNA synthesis was performed using a PrimerScriptTM RT reagent Kit with gDNA Eraser (TaKaRa, Lot # AJ51485A) according to the manufacturer's instructions. PCR reactions were performed using HieffTMqPCR SYBR Green Master Mix (YEASEN, Lot # H28360) and a BioRad CFX Connect Real-Time system. PCR cycling conditions were as previously reported: 95 °C for 5 min, 40 cycles of 95 °C for 15 s, 60 °C for 15 s, and 72 °C for 30 s. We then constructed a melting curve of amplified DNA (31). Target gene values were normalized to *Gapdh* expression and the experimental control using ΔΔCt methods (59). Primer sequences used in this study are shown in Table S1.

CRISPR/Cas9-mediated CBE^{+47} and enhancers deletion in ESC

The CRISPR/Cas9 system was used following published protocols (31, 60, 61). Briefly, target-specific guide RNAs (sgRNAs) were designed using an online tool (http://chopchop.cbu.uib.no/). sgRNAs of the appropriate site and score were selected. sgRNA sequences are shown in Table S2. For CBE^{+47} and enhancer knockout, sgRNAs were cloned into

a Cas9-puro vector using the Bsmb1 site. ESC were transfected with two sgRNA plasmids using Lipofectamine 3000 (Life Technologies, Lot # 2125386), and 24 h later, cells were treated with 5 μM puromycin (MCE, Lot # 64358) for 24 h and then cultured in medium without puromycin for another 5 ∼ 7 days. Individual colonies were picked and validated by genomic DNA PCR and subsequent Sanger DNA sequencing. PCR primers used for genotyping are listed in Table S3.

Alkaline phosphatase (AP) staining of WT and CBE^{+47}-deleted cells

ESC were plated at low density in 12-well plates coated with gelatin for 4 days and then washed twice with PBS and incubated with reagents from the AP Staining kit (Cat # AP100R-1, System Biosciences) following the manufacturer's instructions. Digital images were taken using an Olympus Inverted Fluorescence Microscope.

Enhancer capture-C and data analysis

Capture-C probes were designed using an online tool within three enhancer regions (http://apps.molbiol.ox.ac.uk/CaptureC/cgi-bin/CapSequm.cgi): the E1 bait (Mouse, mm10, chr6: 52012160–52015360), the E2 bait (Mouse, mm10, chr6:52075358–52076639), and the E3 bait (Mouse, mm10, chr6: 52101884–52104217). Probe sequences are listed in Table S4.

Capture-C libraries were prepared as previously described with minor modifications (31, 42, 43). Briefly, RA-induced differentiated ESC (WT and CBE^{+47}-KO) were fixed with 1% (vol/vol) formaldehyde for 10 min at room temperature, quenched with 125 mM glycine in PBS, and then lysed in cold lysis buffer [10 mM Tris-HCl, pH7.5, 10 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EGTA, 0.2% NP-40, 1× complete protease inhibitor cocktail (Roche, Lot # 3024150)]. Chromatin was digested with DpnII (New England Biolabs, Lot # 10014860) at 37 °C overnight. Fragments were then diluted and ligated with T4 DNA ligase (Takara, Lot # 1211707) at 16 °C overnight. Cross-linking was reversed by overnight incubation at 60 °C with proteinase K (Bioline, Lot # BIO-37037). Then 3C libraries were purified by phenol-chloroform followed by chloroform extraction and ethanol-precipitated at −80 °C overnight. Sequencing libraries were prepared from 10 μg of the 3C library by sonication to an average size of 200 ∼ 300 bp and indexed using NEBnext reagents (New England Biolabs, Lot # 0031607), according to the manufacturer's protocol. Enrichment of 2 μg of an indexed library incubated with 3 μM of a pool of biotinylated oligonucleotides was performed using the SeqCap EZ Hybridization reagent kit (Roche/NimbleGen, Lot # 05634261001), following the manufacturer's instructions. Two rounds of capture employing 48∼72 and 24 h hybridizations, respectively, were used. Correct library size was confirmed by agarose gel electrophoresis, and DNA concentration was determined using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Lot # 0000248352). All sequencing was performed on Hi-Seq 2500 platforms using paired 150 bp protocols (Illumina).

Capture-C data were analyzed using previously described methods (31, 42, 43). Briefly, clean paired-end reads were reconstructed into single reads using FLASH (62). After *in silico* DpnII digestion using the DpnII2E.pl script, reads were mapped back to the mm10 mouse genome using Bowtie1. Finally, chimeric reads containing captured reads and Capture-Reporter reads were analyzed using CCanalyser3.pl. Results were visualized using the Integrated Genome Browser (IGV) (63) and online tool (https://epgg-test.wustl.edu/browser/) (64).

RNA-seq and bioinformatics analysis

ESC were lysed with Trizol reagent (Life Technologies, Lot # 265709) and RNA was extracted based on the manufacturer's instructions. RNA was then sequenced by a company (Novogene). Clean reads were mapped to the Ensemble mm10 mouse genome using Hisat2 with default parameters. Gene reads were counted by Htseq (65). Fold changes were computed as a log2 ratio of normalized reads per gene using the DEseq2 R package (66). Gene expression with|log2 (fold change)|≥ 1 ($p < 0.05$) was considered significantly altered. Heatmaps were drawn using the heatmap.2 function. Two biological replicates were analyzed for each experimental condition.

Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway analyses

GO and KEGG pathway analyses were performed using the following online tools: Metascape (http://metascape.org/gp/index.html#/main/step1) (67) and DAVID Functional Annotation Bioinformatics Microarray Analysis tool (http://david.abcc.ncifcrf.gov/) (68).

Gene set enrichment analysis (GSEA)

GSEA was carried out using the online tool (https://www.gsea-msigdb.org/gsea/index.jsp), as previously reported (69, 70). KEGG pathway-related gene sets were obtained from the KEGG database (https://www.genome.jp/kegg/) (71).

Statistical analyses

Data were analyzed by Student's *t*-test (two-tailed) unless otherwise specified. Statistically significant *p*-values are indicated in Figures as follows: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Data availability

Data supporting this study are available from the corresponding author upon request. Raw data reported here (DNA-seq and RNA-seq) have been deposited in the NCBI Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE154495.

Published data sets were downloaded using online tools (http://cistrome.org/db/#/ and http://promoter.bx.psu.edu/hi-c/) and analyzed in this study (72, 73). These include: GSM2588420 for H3K27ac ChIP-seq analyses and GSE124306 for E3 enhancer Capture-C in mouse ESC (WT) 24 h after RA

induction (31, 33); GSE96107 for Hi-C (74); GSM859491 for H3K27ac (75); GSM2630487 for H3K4me1 (76); GSM881353 for H3K4me2 (75); GSM1258237 for H3K4me3 (77); GSM699165 for CTCF (78); GSM2645432 for YY1 (79); GSM1439567 for *Med1* (80); GSM560345 for *Med12* (81); GSM1276711 for RNA PolII (77); GSM1014154 for DNase (82); and GSM2651154 for ATAC-seq data (83) in mouse undifferentiated ESC.

Supporting information—This article contains supporting information.

Acknowledgments—We thank all members of our laboratory for many helpful discussions. We also thank Dr Elise Lamar for editing our article. This work was supported by the National Key R&D Program of China (NO.2017YFA0102600), Chinese National Natural Science Foundation of China (NSFC31530027).

Author contributions—W. L., G. S., and L. Z. conceived and supervised the study and designed the experiments. G. S., J. C., M. L., J. Z., J. B., Z. Z., and J. S. performed the experiments and analyzed the data. J. C., D. G., and W. W. analyzed the sequencing data. G. S. and W. W. performed bioinformatic analysis. G. S., W. L., and L. Z. wrote the article. All the authors have read and approved the final article.

Conflict of interests—The authors declare that they have no conflict of interest.

Abbreviations—The abbreviations used are: AP, alkaline phosphatase; CBE, CTCF-binding element; EEIC, enhancer–enhancer interaction complex; ESC, embryonic stem cell; *HoxA*, Homeobox A; ICM, inner cell mass; RA, retinoic acid; TAD, topologically associated domain.

References

1. Martin, G. R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 7634–7638
2. 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
3. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872
4. Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A. H. (2004) Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat. Med.* **10**, 55–63
5. Metcalfe, A. D., and Ferguson, M. W. (2007) Tissue engineering of replacement skin: The crossroads of biomaterials, wound healing, embryonic development, stem cells and regeneration. *J. R. Soc. Interface* **4**, 413–437
6. Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., Wong, K. Y., Sung, K. W., Lee, C. W., Zhao, X. D., Chiu, K. P., *et al.* (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431–440
7. Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., and Young, R. A. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956
8. Martinez-Ceballos, E., Chambon, P., and Gudas, L. J. (2005) Differences in gene expression between wild type and Hoxa1 knockout embryonic stem cells after retinoic acid treatment or leukemia inhibitory factor (LIF) removal. *J. Biol. Chem.* **280**, 16484–16498
9. Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., Miyazaki Ji, J., and Niwa, H. (2002) Differentiation of embryonic stem cells is induced by GATA factors. *Genes Dev.* **16**, 784–789
10. Gajovic, S., St-Onge, L., Yokota, Y., and Gruss, P. (1997) Retinoic acid mediates Pax6 expression during *in vitro* differentiation of embryonic stem cells. *Differ. Res. Biol. Divers.* **62**, 187–192
11. Levasseur, D. N., Wang, J., Dorschner, M. O., Stamatoyannopoulos, J. A., and Orkin, S. H. (2008) Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes Dev.* **22**, 575–580
12. Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., Wang, K., and Lu, W. (2013) Klf4 organizes long-range chromosomal interactions with the Oct4 locus in reprogramming and pluripotency. *Cell Stem Cell* **13**, 36–47
13. Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D. L., Wang, Y., Hansen, R. S., Canfield, T. K., Thurman, R. E., Cheng, Y., Gulsoy, G., Dennis, J. H., Snyder, M. P., *et al.* (2014) Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–405
14. Luppino, J. M., Park, D. S., Nguyen, S. C., Lan, Y., Xu, Z., Yunker, R., and Joyce, E. F. (2020) Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nat. Genet.* **52**, 840–848
15. Lupianez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., *et al.* (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025
16. Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., Shachar, S., Weintraub, A. S., Hnisz, D., Pegoraro, G., Lee, T. I., Misteli, T., Jaenisch, R., and Young, R. A. (2016) 3D chromosome regulatory landscape of human pluripotent cells. *Cell Stem Cell* **18**, 262–275
17. Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A., and Bruneau, B. G. (2017) Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e922
18. Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., Jung, I., Wu, H., Zhai, Y., Tang, Y., Lu, Y., Wu, Y., Jia, Z., Li, W., Zhang, M. Q., *et al.* (2015) CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162**, 900–910
19. de Wit, E., Vos, E. S., Holwerda, S. J., Valdes-Quezada, C., Verstegen, M. J., Teunissen, H., Splinter, E., Wijchers, P. J., Krijger, P. H., and de Laat, W. (2015) CTCF binding polarity determines chromatin looping. *Mol. Cell* **60**, 676–684
20. Nanni, L., Ceri, S., and Logie, C. (2020) Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. *Genome Biol.* **21**, 197
21. Zheng, H., and Xie, W. (2019) The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550
22. Tay, Y., Zhang, J., Thomson, A. M., Lim, B., and Rigoutsos, I. (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**, 1124–1128
23. Bain, G., Kitchens, D., Yao, M., Huettner, J. E., and Gottlieb, D. I. (1995) Embryonic stem cells express neuronal properties *in vitro*. *Dev. Biol.* **168**, 342–357
24. Kashyap, V., Gudas, L. J., Brenet, F., Funk, P., Viale, A., and Scandura, J. M. (2011) Epigenomic reorganization of the clustered Hox genes in embryonic stem cells induced by retinoic acid. *J. Biol. Chem.* **286**, 3250–3260
25. De Kumar, B., Parrish, M. E., Slaughter, B. D., Unruh, J. R., Gogol, M., Seidel, C., Paulson, A., Li, H., Gaudenz, K., Peak, A., McDowell, W., Fleharty, B., Ahn, Y., Lin, C., Smith, E., *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
26. Luo, H., Wang, F., Zha, J., Li, H., Yan, B., Du, Q., Yang, F., Sobh, A., Vulpe, C., Drusbosky, L., Cogle, C., Chepelev, I., Xu, B., Nimer, S. D.,

Licht, J., *et al.* (2018) CTCF boundary remodels chromatin domain and drives aberrant HOX gene transcription in acute myeloid leukemia. *Blood* **132**, 837–848
27. Li, Y., Liao, Z., Luo, H., Benyoucef, A., Kang, Y., Lai, Q., Dovat, S., Miller, B., Chepelev, I., Li, Y., Zhao, K., Brand, M., and Huang, S. (2020) Alteration of CTCF-associated chromatin neighborhood inhibits TAL1-driven oncogenic transcription program and leukemogenesis. *Nucleic Acids Res.* **48**, 3119–3133
28. Jain, S., Ba, Z., Zhang, Y., Dai, H. Q., and Alt, F. W. (2018) CTCF-binding elements mediate accessibility of RAG substrates during chromatin scanning. *Cell* **174**, 102–116.e114
29. Ba, Z., Lou, J., Ye, A. Y., Dai, H.-Q., Dring, E. W., Lin, S. G., Jain, S., Kyritsis, N., Kieffer-Kwon, K.-R., Casellas, R., and Alt, F. W. (2020) CTCF orchestrates long-range cohesin-driven V(D)J recombinational scanning. *Nature* **586**, 305–310
30. Arzate-Mejia, R. G., Recillas-Targa, F., and Corces, V. G. (2018) Developing in 3D: The role of CTCF in cell differentiation. *Development* **145**, dev137729
31. Su, G., Guo, D., Chen, J., Liu, M., Zheng, J., Wang, W., Zhao, X., Yin, Q., Zhang, L., Zhao, Z., Shi, J., and Lu, W. (2019) A distal enhancer maintaining Hoxa1 expression orchestrates retinoic acid-induced early ESCs differentiation. *Nucleic Acids Res.* **47**, 6737–6752
32. Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., Zhao, Y., Shen, B., Huang, X., Zhu, H., Orkin, S. H., and Shen, X. (2015) Opposing roles for the lncRNA haunt and its genomic locus in regulating HOXA gene activation during embryonic stem cell differentiation. *Cell Stem Cell* **16**, 504–516
33. Cao, K., Collings, C. K., Marshall, S. A., Morgan, M. A., Rendleman, E. J., Wang, L., Sze, C. C., Sun, T., Bartom, E. T., and Shilatifard, A. (2017) SET1A/COMPASS and shadow enhancers in the regulation of homeotic gene expression. *Genes Dev.* **31**, 787–801
34. Liu, G. Y., Zhao, G. N., Chen, X. F., Hao, D. L., Zhao, X., Lv, X., and Liu, D. P. (2016) The long noncoding RNA Gm15055 represses Hoxa gene expression by recruiting PRC2 to the gene cluster. *Nucleic Acids Res.* **44**, 2613–2627
35. Ghasemi, R., Struthers, H., Wilson, E. R., and Spencer, D. H. (2021) Contribution of CTCF binding to transcriptional activity at the HOXA locus in NPM1-mutant AML cells. *Leukemia* **35**, 404–416
36. Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O., and Reinberg, D. (2015) CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–1021
37. Rousseau, M., Crutchley, J. L., Miura, H., Suderman, M., Blanchette, M., and Dostie, J. (2014) Hox in motion: Tracking HoxA cluster conformation during differentiation. *Nucleic Acids Res.* **42**, 1524–1540
38. Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., and Tilghman, S. M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**, 486–489
39. Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., and Ohlsson, R. (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10684–10689
40. De Kumar, B., Parker, H. J., Parrish, M. E., Lange, J. J., Slaughter, B. D., Unruh, J. R., Paulson, A., and Krumlauf, R. (2017) Dynamic regulation of Nanog and stem cell-signaling pathways by Hoxa1 during early neuroectodermal differentiation of ES cells. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5838–5845
41. Zhao, H., Li, Z., Zhu, Y., Bian, S., Zhang, Y., Qin, L., Naik, A. K., He, J., Zhang, Z., Krangel, M. S., and Hao, B. (2020) A role of the CTCF binding site at enhancer Eα in the dynamic chromatin organization of the Tcra–Tcrd locus. *Nucleic Acids Res.* **48**, 9621–9636
42. Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D. R. (2014) Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**, 205–212
43. Davies, J. O., Telenius, J. M., McGowan, S. J., Roberts, N. A., Taylor, S., Higgs, D. R., and Hughes, J. R. (2016) Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* **13**, 74–80
44. Liang, Y. C., Wu, P., Lin, G. W., Chen, C. K., Yeh, C. Y., Tsai, S., Yan, J., Jiang, T. X., Lai, Y. C., Huang, D., Cai, M., Choi, R., Widelitz, R. B., Lu, W., and Chuong, C. M. (2020) Folding keratin gene clusters during skin regional specification. *Dev. Cell* **53**, 561–576.e569
45. Zhang, L., He, A., Chen, B., Bi, J., Chen, J., Guo, D., Qian, Y., Wang, W., Shi, T., Zhao, Z., Shi, J., An, W., Attenello, F., and Lu, W. (2020) A HOTAIR regulatory element modulates glioma cell sensitivity to temozolomide through long-range regulation of multiple target genes. *Genome Res.* **30**, 155–163
46. Qian, Y., Zhang, L., Cai, M., Li, H., Xu, H., Yang, H., Zhao, Z., Rhie, S. K., Farnham, P. J., Shi, J., and Lu, W. (2019) The prostate cancer risk variant rs55958994 regulates multiple gene expression through extreme long-range chromatin interaction to control tumor progression. *Sci. Adv.* **5**, eaaw6710
47. Beagan, J. A., Duong, M. T., Titus, K. R., Zhou, L., Cao, Z., Ma, J., Lachanski, C. V., Gillis, D. R., and Phillips-Cremins, J. E. (2017) YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* **27**, 1139–1152
48. Ferraiuolo, M. A., Rousseau, M., Miyamoto, C., Shenker, S., Wang, X. Q., Nadler, M., Blanchette, M., and Dostie, J. (2010) The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res.* **38**, 7472–7484
49. Luo, H., Sobh, A., Vulpe, C. D., Brewer, E., Dovat, S., Qiu, Y., and Huang, S. (2019) HOX loci focused CRISPR/sgRNA library screening identifying critical CTCF boundaries. *J. Vis. Exp.* https://doi.org/10.3791/59382
50. Korkmaz, G., Manber, Z., Lopes, R., Prekovic, S., Schuurman, K., Kim, Y., Teunissen, H., Flach, K., Wit, E., Galli, G. G., Zwart, W., Elkon, R., and Agami, R. (2019) A CRISPR-Cas9 screen identifies essential CTCF anchor sites for estrogen receptor-driven breast cancer cell proliferation. *Nucleic Acids Res.* **47**, 9557–9572
51. Sato, T., Kataoka, K., Ito, Y., Yokoyama, S., Inui, M., Mori, M., Takahashi, S., Akita, K., Takada, S., Ueno-Kudoh, H., and Asahara, H. (2020) Lin28a/let-7 pathway modulates the Hox code via polycomb regulation during axial patterning in vertebrates. *Elife* **9**, e53608
52. Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenkov, V. V., and Ren, B. (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–1245
53. Gaszner, M., and Felsenfeld, G. (2006) Insulators: Exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* **7**, 703–713
54. Dowen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., Weintraub, A. S., Schujiers, J., Lee, T. I., Zhao, K., and Young, R. A. (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387
55. Wang, P., Tang, Z., Lee, B., Zhu, J. J., Cai, L., Szalaj, P., Tian, S. Z., Zheng, M., Plewczynski, D., Ruan, X., Liu, E. T., Wei, C. L., and Ruan, Y. (2020) Chromatin topology reorganization and transcription repression by PML-RARalpha in acute promyeloid leukemia. *Genome Biol.* **21**, 110
56. Valcarcel, R., Holz, H., Jimenez, C. G., Barettino, D., and Stunnenberg, H. G. (1994) Retinoid-dependent *in vitro* transcription mediated by the RXR/RAR heterodimer. *Genes Dev.* **8**, 3068–3079
57. Rhinn, M., and Dolle, P. (2012) Retinoic acid signalling during development. *Development* **139**, 843–858
58. Simandi, Z., Horvath, A., Wright, L. C., Cuaranta-Monroy, I., De Luca, I., Karolyi, K., Sauer, S., Deleuze, J. F., Gudas, L. J., Cowley, S. M., and Nagy, L. (2016) OCT4 acts as an integrator of pluripotency and signal-induced differentiation. *Mol. Cell* **63**, 647–661
59. Schmittgen, T. D., and Livak, K. J. (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* **3**, 1101–1108
60. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823
61. Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., McDonel, P. E., Guttman, M., and Lander, E. S. (2016) Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455

62. Magoc, T., and Salzberg, S. L. (2011) FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963
63. Freese, N. H., Norris, D. C., and Loraine, A. E. (2016) Integrated genome browser: Visual analytics platform for genomics. *Bioinformatics* **32**, 2089–2095
64. Zhou, X., Lowdon, R. F., Li, D., Lawson, H. A., Madden, P. A., Costello, J. F., and Wang, T. (2013) Exploring long-range genome interactions using the WashU epigenome browser. *Nat. Methods* **10**, 375–376
65. Anders, S., Pyl, P. T., and Huber, W. (2015) HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169
66. 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
67. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., and Chanda, S. K. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523
68. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57
69. 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. U. S. A.* **102**, 15545–15550
70. Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., Houstis, N., Daly, M. J., Patterson, N., Mesirov, J. P., Golub, T. R., *et al.* (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273
71. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–462
72. Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M. N. K., Li, Y., Hu, M., Hardison, R., Wang, T., and Yue, F. (2018) The 3D genome browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* **19**, 151
73. Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C. H., Brown, M., Zhang, X., Meyer, C. A., and Liu, X. S. (2019) Cistrome data browser: Expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* **47**, D729–D735
74. Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., Xu, X., Lv, X., Hugnot, J. P., Tanay, A., and Cavalli, G. (2017) Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e524
75. Xiao, S., Xie, D., Cao, X., Yu, P., Xing, X., Chen, C. C., Musselman, M., Xie, M., West, F. D., Lewin, H. A., Wang, T., and Zhong, S. (2012) Comparative epigenomic annotation of regulatory DNA. *Cell* **149**, 1381–1392
76. Cao, K., Collings, C. K., Morgan, M. A., Marshall, S. A., Rendleman, E. J., Ozark, P. A., Smith, E. R., and Shilatifard, A. (2018) An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition in embryonic stem cells. *Sci. Adv.* **4**, eaap8747
77. Denissov, S., Hofemeister, H., Marks, H., Kranz, A., Ciotta, G., Singh, S., Anastassiadis, K., Stunnenberg, H. G., and Stewart, A. F. (2014) Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141**, 526–537
78. Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., Lee, C. W., Ye, C., Ping, J. L., Mulawadi, F., Wong, E., Sheng, J., Zhang, Y., Poh, T., Chan, C. S., *et al.* (2011) CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat. Genet.* **43**, 630–638
79. Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., Abraham, B. J., Cohen, M. A., Nabet, B., Buckley, D. L., Guo, Y. E., Hnisz, D., Jaenisch, R., Bradner, J. E., Gray, N. S., *et al.* (2017) YY1 is a structural regulator of enhancer-promoter loops. *Cell* **171**, 1573–1588.e1528
80. Buganim, Y., Markoulaki, S., van Wietmarschen, N., Hoke, H., Wu, T., Ganz, K., Akhtar-Zaidi, B., He, Y., Abraham, B. J., Porubsky, D., Kulenkampff, E., Faddah, D. A., Shi, L., Gao, Q., Sarkar, S., *et al.* (2014) The developmental potential of iPSCs is greatly influenced by reprogramming factor selection. *Cell Stem Cell* **15**, 295–309
81. Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J., and Young, R. A. (2010) Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435
82. Vierstra, J., Rynes, E., Sandstrom, R., Zhang, M., Canfield, T., Hansen, R. S., Stehling-Sun, S., Sabo, P. J., Byron, R., Humbert, R., Thurman, R. E., Johnson, A. K., Vong, S., Lee, K., Bates, D., *et al.* (2014) Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. *Science* **346**, 1007–1012
83. Tastemel, M., Gogate, A. A., Malladi, V. S., Nguyen, K., Mitchell, C., Banaszynski, L. A., and Bai, X. (2017) Transcription pausing regulates mouse embryonic stem cell differentiation. *Stem Cell Res.* **25**, 250–255