1    **CTCF-binding element regulates ESC differentiation via orchestrating long-range chromatin**

2    **interaction between enhancers and HoxA**

3

4    **AUTHORS**

5    Guangsong Su[1], Wenbin Wang[1], Jun Chen[1], Man Liu[1], Jian Zheng[1], Dianhao Guo[1], Jinfang Bi[1],

6    Zhongfang Zhao[1], Jiandang Shi[1], Lei Zhang[1, 2], Wange Lu[1, 2]

7

8    **AFFILIATIONS**

9    [1] College of Life Sciences, Nankai University, 94 Weijin Road, 300071, Tianjin, China.

10    [2]Correspondence: wangelv@gmail.com or joyleizhang@nankai.edu.cn

11    Phone: +86-22-2350-5967, Fax: +86-22-2350-5967

12

13    **Key Words:** CTCF; Embryonic stem cells; Chromatin interaction; Enhancer; HoxA; Differentiation

14

15    **Manuscript Information:** 146 words abstract; 3,943 words main text (excluding materials and

16    methods); 7 Figures (8 supplemented); 70 references; 4 Supplementary Files

17

18 **ABSTRACT**

19 Proper expressions of Homeobox A cluster genes (*HoxA*) are essential for embryonic stem cells

20 (ESC) differentiation and individual development. However, the mechanisms underlying

21 controlling precise spatiotemporal expressions of *HoxA* during early ESC differentiation remain

22 poorly understood. Herein, we identified a functional CTCF-binding element (CBE$^{+47}$) closest to

23 the 3'-end of *HoxA* within the same topologically associated domain (TAD) in ESC. CRISPR-Cas9

24 mediated the CBE$^{+47}$ deletion significantly promotes retinoic acid (RA)-induced *HoxA* expression

25 and early ESC abnormal differentiation. In delineating its underlying mechanisms, we find that

26 CBE$^{+47}$ can precisely organize chromatin interactions between its adjacent enhancers and *HoxA* by

27 using chromosome conformation capture assay (Capture-C). Furthermore, we also find that its

28 adjacent enhancers as enhancer-enhancer interaction complex (EEIC) are required for RA-induced

29 *HoxA* activation. Collectively, these results provide a new insight for RA-induced *HoxA* expression,

30 also highlight the unique and precise regulatory roles of CBE in ESC differentiation.

31

32

33 **INTRODUCTION**

34      Embryonic stem cells (ESC) are cells derived from the inner cell mass (ICM) (*Martin, 1981*).

35 ESC have the ability of unlimited self-renewal and differentiation into three embryonic germ layers

36 (*Takahashi et al., 2007; Takahashi & Yamanaka, 2006*). Therefore, ESC have great application

37 potential in regenerative medicine and tissue engineering (*Metcalfe & Ferguson, 2007; N. Sato,*

38 *Meijer, Skaltsounis, Greengard, & Brivanlou, 2004*). Different state of ESC requires proper

39 expression of specific gene. The pluripotency state of ESC requires expressions of master control

40 transcription factors such as *Oct4*/*Nanog*/*Sox2* (*Boyer et al., 2005; Loh et al., 2006*), and

41 expressions of genes such as *HoxA*/*Gata4*/*Pax6* are essential for ESC differentiation (*Fujikura et*

42 *al., 2002; Gajovic, St-Onge, Yokota, & Gruss, 1997; Martinez-Ceballos, Chambon, & Gudas,*

43 *2005*). In recent years, studies have shown that high-order chromatin structures play a critical role

44 in regulating gene expression and maintaining ESC status (*Levasseur, Wang, Dorschner,*

45 *Stamatoyannopoulos, & Orkin, 2008; Wei et al., 2013*). Numerous studies have revealed that

46 topologically associated domains (TAD) are the basic unit of nuclear chromatin (*Luppino et al.,*

47 *2020; Pope et al., 2014*). Functional DNA elements, such as enhancers, are more likely to promote

48 expression of target gene through directly long-range chromatin interactions within the same TAD

49 (*Ji et al., 2016; Lupianez et al., 2015*). Long-range chromatin interactions depend on the proteins

50 that mediate its interaction, among them CTCF plays the central role. Nora Elphege P et al. used

51 auxin-induced degradation of CTCF protein and found that the absence of CTCF led to

52 disappearance of TAD, revealing that CTCT is essential for maintaining the formation of TAD

53 (*Nora et al., 2017*). CTCF forms chromatin-loop through directly binding the CTCF-binding

54 elements (CBE) enriched at TAD boundary and relies on its binding polarity criteria to maintain the

55 formation of TAD (*de Wit et al., 2015; Guo et al., 2015; Nanni, Ceri, & Logie, 2020*). Although

56 many studies have found that the CBE at TAD boundary plays key regulatory roles in biological

57 development, the research on CBE inside the TAD is relatively lacking.

58      Retinoic acid (RA) can induce ESC differentiation rapidly (*Bain, Kitchens, Yao, Huettner, &*

59 *Gottlieb, 1995; Tay, Zhang, Thomson, Lim, & Rigoutsos, 2008*). In this process, *HoxA* are

60 significantly activated (*De Kumar et al., 2015; Kashyap et al., 2011*). Recent studies have shown

61 that CBE plays the core regulatory roles in tumorigenesis (*Li et al., 2020; Huacheng Luo et al.,*

62 *2018*), immunogenesis (*Ba et al., 2020; Jain, Ba, Zhang, Dai, & Alt, 2018*) and stem cell

63    development (*Arzate-Mejia, Recillas-Targa, & Corces, 2018; H. Zheng & Xie, 2019*) by

64    controlling target genes expression through organizing high-order chromatin structures. We and

65    other research groups have found that multiple enhancers are required for RA-induced *HoxA*

66    expression and early ESC differentiation by long-range chromatin interactions with *HoxA* (*Cao et*

67    *al., 2017; Liu et al., 2016; Su et al., 2019; Yin et al., 2015*). Besides, previous studies also have

68    reported many regulatory patterns for *HoxA* activation during ESC differentiation, among them,

69    several CBE (also known as CBS, CTCF-binding site) within *HoxA* locus can organize *HoxA*

70    chromatin structures, and play significant regulatory roles in ESC differentiation and

71    leukemogenesis (*Ghasemi, Struthers, Wilson, & Spencer, 2020; Huacheng Luo et al., 2018;*

72    *Narendra et al., 2015*). However, it is unknown whether CBE plays functional roles in regulating

73    *HoxA* expression by orchestrating enhancer chromatin structures during RA-induced early ESC

74    differentiation.

75        In this study, we identified a new functional CTCF-binding element (CBE$^{+47}$) required for RA

76    to activate *HoxA* expression and promote early ESC proper differentiation. On the mechanisms, our

77    results show that CBE$^{+47}$ can regulate the interactions between multiple enhancers, as well as the

78    interactions of these enhancers with *HoxA* chromatin. Our study reveals that a new functional

79    CBE$^{+47}$ can regulate the precise expressions of *HoxA* by accurately organizing chromatin

80    interactions between its adjacent enhancers and *HoxA,* thereby facilitating early correct

81    differentiation of RA-induced ESC. These findings not only increase our understanding of

82    mechanisms of RA-induced *HoxA* activation and early ESC differentiation, but also highlight the

83    unique and precise regulatory roles of CBE in high-order chromatin structures.

84

85    **RESULTS**

86    **Identification of a CTCF-binding element (CBE$^{+47}$) closest to the 3'-end of *HoxA* within the**

87    **same TAD in ESC**

88    The basic structural unit in high-order chromatin structure is the TAD, and DNA functional

89    elements such as enhancers or insulators in the same TAD are more likely to regulate the precise

90    expression of neighboring genes (*Ji et al., 2016; Lupianez et al., 2015*). To discover potential DNA

91    functional elements that regulate the expression of *HoxA*, we first analyzed the Hi-C data in ESC.

92    We find that 5'-end of *HoxA* is located at the TAD boundary region, and 3'-end is within the TAD

93    (*Figure 1A*). Previous studies have shown that several significant CTCF-binding elements (CBE)

94    are located at the *HoxA* locus, and play an important regulatory role in *HoxA* expression and ESC

95    differentiation (*Narendra et al., 2015; Rousseau et al., 2014*). Interestingly, many significant CBE

96    are also found in the same TAD at 3'-end of *HoxA* (*Figure 1B*), but there are no reports on whether

97    these CBE have regulatory effects on *HoxA* expression and ESC differentiation.

98    Here, we focus on a significant CBE closest to the 3'-end of *HoxA*, about 47 kb from *HoxA*

99    (CBE$^{+47}$), located in the second intron of *Halr1* (*Figure 1B; Figure 1–figure supplement 1*). We

100   find significant binding of Med1, Med12 and YY1 at upstream of CBE$^{+47}$ (*Figure 1B*). The

101   epigenetic modification-related markers (e.g. H3K27ac, H3K4me1/2/3) are also enriched upstream

102   of CBE$^{+47}$ (*Figure 1C*). Furthermore, we also find significant chromatin accessibility (DNase and

103   ATAC-seq) and RNA transcription activity (PolII) enrichment in these same regions (*Figure 1D*).

104   These data indicate that CBE$^{+47}$ is located between the potential regulatory elements and *HoxA*.

105   Previous studies have shown that CBE can precisely regulate target gene expression through

106   organizing chromatin interactions between functional element and target gene (*Hark et al., 2000;*

107   *Kurukuti et al., 2006*). Therefore, based on the above results, we speculate that CBE$^{+47}$ may have a

108   regulatory role.

109

110   **Genes expression analysis in WT and CBE$^{+47}$ deletion cells under self-renewal culture**

111   **condition**

112   To investigate whether CBE$^{+47}$ has a regulatory effect, CRISPR-Cas9 mediated deletion

113   method was used for further research (*Su et al., 2019*). Two sgRNAs were designed in upstream

114   and downstream of CBE$^{+47}$ (*Figure 2A*), after Cas9 cleavage and DNA recombination, we obtained

115    two homozygous CBE$^{+47}$ knockout cell lines (CBE$^{+47}$-KO). Genomic DNA PCR using specific

116    primer, and Sanger sequencing confirmed the deletion of CBE$^{+47}$ (*Figure 2B*). AP-staining showed

117    that there were no significant differences in cell morphology in the CBE$^{+47}$ knockout cell lines under

118    self-renewing culture conditions (*Figure 2C*). qRT-PCR results showed that mRNA expression

119    levels of *Halr1* and *Sakp2* did not change significantly (*Figure 2D*). However, 3'-end genes of *HoxA*

120    were significantly up-regulated (*Hoxa2-a6*), while TAD boundary gene *Hoxa9* was significantly

121    down-regulated in CBE$^{+47}$-KO cells compared with WT (*Figure 2E*). These data suggest that

122    deletion of CBE$^{+47}$ significantly changes expressions of *HoxA*, indicating that even in

123    undifferentiated ESC state, the CBE$^{+47}$ is also required to maintain proper expressions of *HoxA*.

124        Considering that expression of *HoxA* is critical for ESC differentiation (*De Kumar et al., 2017;*

125    *Martinez-Ceballos et al., 2005; Yin et al., 2015*), and CBE$^{+47}$ deletion significantly increased some

126    *HoxA* expression, we then tested if the ESC's pluripotency status was impacted by examining the

127    expressions of pluripotency- and differentiation-related genes by qRT-PCR (*Figure 2F-K*). No

128    significant differences in the expressions of pluripotency-regulated genes, except for a slight up-

129    regulation of *Sox2*. Mesoderm and trophectoderm genes are significantly up-regulated. Expression

130    of other germ layer genes such as *Otx2* (epiblast), *Sox17* (endoderm) and *Sox11* (ectoderm) also

131    significantly increased. These results indicate that deletion of CBE$^{+47}$ does not regulate the

132    pluripotency of ESC, but may affect its differentiation ability.

133

134    **CBE$^{+47}$ deletion significantly promotes RA-induced *HoxA* expression and early ESC**

135    **abnormal differentiation**

136        During the early differentiation of ESC induced by RA, *HoxA* are significantly activated

137    compared with undifferentiated ESC (*De Kumar et al., 2015; Yin et al., 2015*) (*Figure 3–figure*

138    *supplement 1*). Therefore, we would like to know whether the CBE$^{+47}$ plays a regulatory role in

139    regulating *HoxA* in the early stage of RA-induced ESC differentiation. ESC differentiation was

140    induced by RA to 3 different time points (*Figure 3–figure supplement 2A*). qRT-PCR was used to

141    analyze expression levels of *HoxA*, and the results showed that *HoxA* 3'-end genes had significantly

142    higher expressions (*Hoxa2-a6*), and TAD boundary genes (*Hoxa7* and *Hoxa9*) also show

143    significantly increased expression levels. However, *HoxA* 5'-end genes did not change significantly

144    (*Hoxa10*, *Hoxa11* and *Hoxa13*) (*Figure 3A-C*). In addition, we also find that antisense long non-

145    coding RNAs at *HoxA* locus are also significantly up-regulated, such as *Hoxaas3* (*Figure 3–figure*

146    *supplement 2B-D*). These data indicate that CBE$^{+47}$ knockout significantly promotes RA-induced

147    expression of *HoxA*, revealing that the CBE$^{+47}$ is to restrict RA-induced over-expression of *HoxA*.

148        Furthermore, to investigate whether CBE$^{+47}$ knockout affects RA-induced ESC differentiation,

149    we used the qRT-PCR method to analyze the changes of pluripotency- and differentiation-related

150    master control genes in WT and CBE$^{+47}$-KO ESC under RA condition. The results show that the

151    expressions of the pluripotency master control genes did not change significantly at 12 hours after

152    ESC differentiation but significantly decreased at 24 and 48 hours (such as *Nanog* and *Klf4*).

153    Epiblast master control gene (*Otx2*), endoderm master control gene (*Gata4*) and ectoderm master

154    control gene (*Pax6*) have significantly high expression levels, while mesoderm master gene (*T*) and

155    trophectoderm master gene (*Hand1*) are significantly down-regulated, these data suggesting that

156    deletion of CBE$^{+47}$ significantly leads to ESC abnormal differentiation (*Figure 3D-E*). Combining

157    these results reveals that the CBE$^{+47}$ is essential to maintain RA-induced *HoxA* expression and early

158    ESC proper differentiation.

159

160    **Transcriptome analysis in WT and CBE$^{+47kb}$ deletion cells during RA-induced early ESC**

161    **differentiation**

162        To further investigate the impacts of CBE$^{+47}$ knockout on RA-induced early ESC

163    differentiation (24h), we performed transcriptome analysis by RNA sequencing (RNA-seq) after

164    RA-induced differentiation of WT and CBE$^{+47}$ knockout cells. Compared with WT cells, a total of

165    869 genes were up-regulated and 360 genes down-regulated in CBE$^{+47}$-KO cells after statistical

166    analysis (Fold change $\geq 2$, $P < 0.05$) (*Figure 4A*). Among them, *HoxA* are significantly up-regulated

167    (*Figure 4B*). Because the *HoxA* 3'-end genes located in the same TAD with CBE$^{+47}$, compared with

168    those genes in two adjacent TADs, we find that CBE$^{+47}$ deletion significantly promotes gene

169    expression in the same TAD, while genes in two adjacent TADs have no significant differences,

170    indicating that direct regulation of CBE$^{+47}$ is restricted within the TAD. These findings reveal that

171    CBE$^{+47}$ services as a *cis*-regulatory element playing a local regulatory role (*Figure 4–figure*

172    *supplement 1*). Moreover, compared with WT cells, we also find that *HoxB/C/D* cluster genes also

173    have significantly higher expressions (*Figure 4–figure supplement 2A*). Relevant genes in the RA

174    signaling pathway are also significantly over-expressed, such as *Crabp2*, *Cyp26a1* and *Stra6*.

175   Differentiation-related genes *Cbx4* and *Foxa1* are also significantly up-regulated (*Figure 4–figure*

176   *supplement 2B*). On the contrary, we also find that genes related to pluripotency such as *Myc*,

177   *Lin28a* and *Etv4* are significantly suppressed (*Figure 4–figure supplement 2C*). These data indicate

178   that in early differentiation of RA-induced ESC, CBE$^{+47}$ deletion significantly promotes early ESC

179   differentiation, suggesting that the CBE$^{+47}$ may inhibit over-differentiation of ESC, thereby ensuring

180   its accurate normal differentiation.

181   In addition, to better understand the molecular mechanisms of CBE$^{+47}$ regulatory roles during

182   RA-induced early ESC differentiation, we performed bioinformatic analysis for the up- and down-

183   regulated genes (*Figure 4–figure supplement 3*). GO analysis indicated that biological processes

184   (BP) of up-regulated genes are mainly involved in gland development, positive regulation of

185   nervous system development, etc (*Figure 4C*). Also, GO-BPs of down-regulated genes are mainly

186   involved in embryonic placenta development, positive regulation of vasoconstriction, epithelial cell

187   differentiation, etc (*Figure 4D*). KEGG pathway enrichment analysis reveals that up-regulated

188   genes are mainly enriched in pathways including *cGMP-PKG* signaling pathway, neuroactive

189   ligand-receptor interaction, axon guidance, signaling pathways regulating pluripotency of stem cells,

190   *Hippo* signaling pathway, *cAMP* signaling pathway, *Wnt* signaling pathway and Hedgehog

191   signaling pathway (*Figure 4E*). Enrichment results of down-regulated genes show that most

192   pathways are related to proteoglycans in cancer, Transcriptional misregulation in cancer,

193   Arachidonic acid metabolism, HTLV-I infection, *Hippo* signaling pathway and Thyroid hormone

194   signaling pathway (*Figure 4E*). These results suggest that CBE$^{+47}$ deletion leads to abnormal genes

195   expression, indicating the essential role of CBE$^{+47}$ for RA-induced early ESC proper differentiation.

196

197   **CBE$^{+47}$ is required for proper chromatin interactions between its adjacent enhancers and**

198   ***HoxA***

199   Previous studies have reported that CTCF regulates target gene expression through controlling

200   chromatin interaction between enhancer and target gene, and this pattern occurs within a TAD

201   (*Zhao et al., 2020*). Studies by us and others have found that there are three enhancers at the 3'-end

202   of *HoxA* (*Figure 5A*). The three enhancers (e.g., E1, E2 and E3) are required for *HoxA* expression

203   during RA-induced ESC differentiation (*Cao et al., 2017; Liu et al., 2016; Su et al., 2019; Yin et*

204   *al., 2015*). Among them, it has been reported that E1 and E3 have significant long-range interactions

205    with *HoxA* chromatin. Remarkably, CBE$^{+47}$ is just located between the three enhancers and *HoxA*

206    (*Figure 5A*). Therefore, we speculate that this CBE may be involved in organizing the interactions

207    between the three enhancers and *HoxA* chromatin.

208        Many experimental techniques, such as Capture-C, have been widely used to study the precise

209    chromatin interactions between enhancers and target genes (*Davies et al., 2016; Hughes et al., 2014;*

210    *Liang et al., 2020; Qian et al., 2019; Su et al., 2019; Zhang et al., 2020*). We used the Capture-C

211    method to study different interactions between enhancers and *HoxA* chromatin in WT and CBE$^{+47}$-

212    KO cells. We used cells induced by RA for 24h to carry out Capture-C, E1, E2 and E3 as the anchor

213    regions, respectively (*Figure 5B-D*). The results show that in WT cells, there are significant

214    interactions between enhancers, and these enhancers also interact with *HoxA* chromatin. Among

215    them, E1 interacts with *HoxA* more strongly (*Figure 5B*). Compared with WT cells, the interactions

216    between enhancers are significantly reduced in CBE$^{+47}$ knockout cells. However, E2 and E3 interact

217    with *HoxA* chromatin increased significantly, and the intensity of chromatin interactions shifted

218    from 3'-end to 5'-end of *HoxA* (*Figure 5C, D*). These results indicate that the deletion of CBE$^{+47}$

219    significantly affects interactions between enhancers and *HoxA* chromatin, revealing that the CBE$^{+47}$

220    is essential to maintain these proper chromatin interactions during RA-induced early ESC

221    differentiation.

222

223    **Multiple enhancers deletion shows synergistic effect on RA-induced *HoxA* expression**

224        To gain a deeper understanding of the regulatory effects of these enhancers on *HoxA* in our

225    system, once again, we used the CRISPR-Cas9 method to knock out these enhancers. We obtained

226    multiple homozygous enhancer knockout cell lines (ie ΔE1, ΔE2, ΔE3, ΔE1/2, ΔE1/2/3) (*Figure*

227    *6–figure supplement 1*). After RA treatment for 24 hours, qRT-PCR was used to determine

228    expressions of *HoxA* (*Figure 6A-E*). The results show that in ΔE1 and ΔE2 cells, enhancers

229    knockout have a similar effect on the regulation of *HoxA*, that is, both the 3'-end (*Hoxa1*) and the

230    central genes (*Hoxa5-a10*) are significantly suppressed (*Figure 6A, B*); In ΔE3 cells, *Hoxa1-a7* are

231    significantly inhibited, which is significantly different from ΔE1 and ΔE2, indicating that these

232    enhancers have their specific regulatory effects (*Figure 6C*). While, in ΔE1/2 and ΔE1/2/3 cells,

233    expressions of *HoxA* are significantly reduced compared with ΔE1 or ΔE2 or ΔE3 (*Figure 6D-E;*

234    *Figure 6–figure supplement 2*). The above results demonstrate that these enhancers are required

235     for RA-induced *HoxA* expression, indicating that the interaction among these three enhancers and

236     *HoxA* chromatin showing a synthetic regulatory effect, which in turn promotes RA-induced *HoxA*

237     activation.

238

239 **DISCUSSION**

240     The precise expressions of *HoxA* are extremely important for ESC differentiation. Although

241 previous studies have shown that CBE plays an important role in the differentiation of ESC by

242 regulating high-order chromatin structures (*Beagan et al., 2017; Narendra et al., 2015*), the

243 identification of functional CBE and mechanisms of its regulation in *HoxA* expression remain

244 largely unknown. Here, a new functional $CBE^{+47}$ is identified in this study and a research model is

245 proposed based on our results (*Figure 7*): In WT cells (Left), the three enhancers near $CBE^{+47}$ acted

246 as EEIC to interact with the 3'-end of *HoxA* chromatin to maintain *HoxA* normal expression, and

247 then promote early correct differentiation of ESC induced by RA; in $CBE^{+47}$-KO cells (Right), the

248 absence of $CBE^{+47}$ leads to reduced interactions between the three enhancers, thereby forming a

249 relatively loose EEIC. Furthermore, the interactions between EEIC and central chromatin regions

250 of *HoxA* are increased, thereby promoting over-expressions of *HoxA* and further enhancing RA-

251 induced early ESC differentiation. These results indicate that a newly identified functional $CBE^{+47}$

252 is required for maintaining the normal expressions of *HoxA* by organizing the precise interactions

253 between its adjacent enhancers and *HoxA* chromatin, thereby regulating the early proper

254 differentiation of ESC. Our findings highlight the unique and indispensable fine-regulatory roles of

255 CBE as a functional element in high-order chromatin structures. The results also reveal direct

256 regulatory effect of the $CBE^{+47}$ on RA-induced *HoxA* expression and early ESC differentiation.

257     In an earlier study, Ferraiuolo MA, et al. have shown that *HoxA* has its special chromatin

258 structures, which is significantly related to *HoxA* expression, and CTCF also plays a regulatory role

259 in *HoxA* chromatin structures and gene expression (*Ferraiuolo et al., 2010*). Recent studies have

260 focused on CBE within *HoxA* locus, including the TAD boundary, and found that the CBE could

261 organize *HoxA* chromatin structures, maintain normal expressions of *HoxA*, and then regulate ESC

262 differentiation (*Narendra et al., 2015*). The study of Varun Narendra, et al. showed that CBE5/6

263 knockout significantly promoted *HoxA* expression, changed *HoxA* chromatin structures and

264 enhanced the neural differentiation of ESC (*Narendra et al., 2015*). Besides, studies in other cells,

265 such as AML cells, have also revealed that CBE located at *HoxA* locus can significantly regulate

266 expressions of *HoxA* and affect tumorigenesis (*Ghasemi et al., 2020; Huacheng Luo et al., 2018*).

267 In this study, we find that besides several CBE in *HoxA* locus, there are other several CBE located

268 downstream of *HoxA* within the same TAD (*Figure 1B*). Among them, the $CBE^{+47}$ knockout

269   significantly enhances RA-induced expressions of *HoxA* and promotes early ESC differentiation.

270   These findings suggest that CBE$^{+47}$ is a new functional CBE. However, several other CBE are still

271   not studied, and whether these CBE have regulatory effects requires further research. In the recent

272   studies, different research groups have developed a CRISPR-Cas9 based genetic screen method to

273   detect functional CBE in breast cancer and AML (*Korkmaz et al., 2019; H. Luo et al., 2019*), which

274   may be used to screen functional CBE in regulating ESC differentiation. These need further

275   exploration.

276          In our subsequent transcriptome analysis, the results show that after CBE$^{+47}$ deletion, genes in

277   the same TAD with CBE$^{+47}$ are significantly up-regulated compared with near TADs (*Figure 4–*

278   *figure supplement 1*), which reveals that the CBE$^{+47}$ may serve as a significant *cis*-regulatory

279   element and play a local regulatory role. In addition, we also find that the RA signaling pathway-

280   related genes are significantly up-regulated (Such as *Crabp2*, *Cyp26a1*, *Stra6*), while the

281   pluripotent-related genes such as *Myc* and *Lin28a* are significantly down-regulated (*Figure 4–*

282   *figure supplement 2*). These results suggest that the absence of CBE$^{+47}$ significantly promotes RA-

283   induced early ESC differentiation. Recent studies have shown that *Lin28a* can inhibit the

284   expressions of *HoxA* and maintain limb development (*T. Sato et al., 2020*), while De Kumar B, et

285   al. found that HOXA1 can bind to *Myc* near enhancers (*De Kumar et al., 2017*), which indicates

286   that there may be positive feedback regulation between *HoxA* and *Myc*, *Lin28a*, etc. in early

287   differentiation of ESC induced by RA. These speculations need further exploration, which will be

288   helpful to understand the differentiation mechanisms and gene expression regulation in ESC.

289          RA can significantly induce *HoxA* expression during early ESC differentiation (*Figure 3–*

290   *figure supplement 1*). Several studies have shown that at least three enhancers are needed for *HoxA*

291   expression in early differentiation of ESC induced by RA (*Cao et al., 2017; Liu et al., 2016; Su et*

292   *al., 2019; Yin et al., 2015*). The CBE$^{+47}$ we studied is just between the three enhancers and *HoxA*

293   (*Figure 5A*). As an insulator, CBE can isolate the interaction between enhancers and target genes

294   and exert inhibitory effect (*Dowen et al., 2014; Gaszner & Felsenfeld, 2006; Kim et al., 2007*).

295   Through our studies, first, we found that the three enhancers interact significantly as EEIC; second,

296   the EEIC has significant interactions with *HoxA*; third, single enhancer interaction with *HoxA* also

297   has its unique model. When the CBE$^{+47}$ is deleted, we find that one is that the interactions between

298   enhancers are significantly reduced, resulting in forming a relatively loose EEIC, which reveals that

299    the CBE$^{+47}$ is required for organizing these enhancers interaction; the second is that the interactions

300    between enhancers and *HoxA* changes significantly, specifically, the interaction of E1 becomes more

301    concentrated in the middle, and the interaction of E2 and E3 with *HoxA* also moves from the 3'-end

302    to the middle regions. Furthermore, we also find that the deletion of the CBE$^{+47}$ significantly

303    promotes higher expressions of *Hoxa2-a9*, which is significantly corresponding to the changes of

304    chromatin interactions between enhancers and *HoxA*. These findings reveal that CBE$^{+47}$ can not

305    only accurately organize interactions between enhancers, but also orchestrating interactions between

306    enhancers and *HoxA*, thus ensuring the correct expressions of *HoxA*. Although the CBE$^{+47}$ can

307    organize high-order chromatin structures, and previous studies have shown that the binding

308    direction of CTCF also can affect the formation of CTCF-mediated chromatin-loop (*de Wit et al.,*

309    *2015*), other CBE paired with the CBE$^{+47}$ to form chromatin-loop are unknown, which needs further

310    exploration.

311    Previous studies have shown that multiple enhancers maintain RA-induced expressions of

312    *HoxA* through long-range chromatin interactions with *HoxA* (*Cao et al., 2017; Liu et al., 2016; Su*

313    *et al., 2019; Yin et al., 2015*). In this study, our results show that knockout of multiple enhancers

314    can show significant synthesis regulation effect on *HoxA* expression (*Figure 6*), which is partly

315    consistent with the previous report (*Cao et al., 2017*). Although these enhancers have been shown

316    to have regulatory effects, we also find that several other enhancers located in the *Skap2* region in

317    the same TAD with *HoxA* (*Figure 5A*). Whether these unreported enhancers have functional

318    regulation effects needs to be further studied. Furthermore, our results also show that CBE$^{+47}$ can

319    significantly organize the interactions between enhancers (E1, E2 and E3). However, the datum

320    shows that there are no significant CTCF-binding peaks in the E2 region, suggesting that other

321    factors might be involved in regulating the interactions between enhancers. A recent study from

322    Wang Ping, et al. have shown that RARα (Retinoic Acid Receptor alpha) can mediate chromatin

323    interactions in AML cells by using the ChIA-PET method (Chromatin Interaction Analysis by

324    Paired-End Tag sequencing) (*P. Wang et al., 2020*). Activation of target genes by RA signaling

325    depends on RAR/RXR (retinoic acid receptor and retinoid X receptor) (*Rhinn & Dolle, 2012;*

326    *Valcarcel, Holz, Jimenez, Barettino, & Stunnenberg, 1994*). Remarkably, we also find that

327    RAR/RXR has significant binding peaks in E1, E2 and *HoxA* regions during RA-induced early ESC

328    differentiation (data not shown) (*Simandi et al., 2016*), demonstrating that RAR/RXR may be

329  involved in regulating these chromatin interactions, which needs further verification. Besides, we

330  also find that chromatin interactions between the single enhancer and *HoxA* are different, and

331  mechanisms of maintaining these different interactions need to be further studied, which will

332  contribute to understanding the regulatory mechanisms of high-order chromatin structures on *HoxA*

333  expression and ESC differentiation.

334      In summary, we identified a new functional CBE$^{+47}$ that can regulate RA-induced expressions

335  of *HoxA* and early ESC differentiation through orchestrating long-range chromatin interactions

336  between its adjacent enhancers and *HoxA*. These findings can enhance our understanding of the

337  intrinsic mechanisms of RA-induced early ESC differentiation, and further highlight the specific

338  regulatory roles of CBE in high-order chromatin structures.

339

340

341 **MATERIALS AND METHODS**

342 **Embryonic stem cells culture**

343      Mouse ESC E14 was cultured as the previous scheme with a few modifications (*Su et al.,*

344 *2019*). ESC were grown in culture dishes coated with 0.1% gelatin (Sigma, Lot # SLBQ9498V) in

345 Dulbecco's Modified Eagle's Medium (DMEM, Gibco, Lot # 8119213) supplemented with 15%

346 fetal bovine serum (FBS, AusGeneX, Lot # FBS00717-1), 1x nonessential amino acids (100x, Gibco,

347 Lot # 2027433), 1x L-Glutamate (100x, Gibco, Lot # 2085472), 1x Penicillin-streptomycin (100x,

348 Gibco, Lot # 2029632), 50 μM β-mercaptoethanol (Sigma, Cas # 60-24-2), 10 ng/mL LIF (ESGRO,

349 Lot # ESG1107), 1 μM PD0325901 (a MEK inhibitor, MedChemExpress, Lot # HY-10254) and 3

350 μM CHIR99021 (a GSK inhibitor, MedChemExpress, Lot # HY-10182). The medium was replaced

351 every 1-2 days. All cells are maintained at 37 °C in a 5% $CO_2$ incubator.

352

353 **Retinoic acid (RA)-induced early ESC differentiation**

354      To induce ESC differentiation, cells were gently washed with 1x phosphate buffer saline

355 solution (PBS), dissociated and plated at an appropriate density on gelatin-coated plates in LIF/2i

356 withdrawal medium supplemented with 2μm RA (Solarbio®, Lot # 1108F031). The culture medium

357 was replaced at the given point-in-time.

358

359 **RNA extraction, reverse transcription, and quantitative Real-Time PCR (qRT-PCR)**

360      Total RNA was extracted from differentiated or undifferentiated ESC using TRIzol Reagent

361 (Life Technologies, Lot # 213504). cDNA synthesis was performed using a PrimerScriptTM RT

362 reagent Kit with gDNA Eraser (TaKaRa, Lot # AJ51485A) according to the manufacturer's

363 instruction. PCR reactions were performed using HieffTMqPCR SYBR Green Master Mix

364 (YEASEN, Lot # H28360) and a BioRad CFX Connect Real-Time system. PCR cycling conditions

365 were as the previous protocol: 95 °C for 5 min, 40 cycles of 95 °C for 15 s, 60 °C for 15 s, and 72 °C

366 for 30 s. And then, a melting curve of amplified DNA was subsequently acquired (*Su et al., 2019*).

367 Quantification of target genes was normalized to *Gapdh* expression and experimental control

368 through ΔΔCt methods (*Schmittgen & Livak, 2008*). Primer sequences used in this study are shown

369 in *Supplementary file 1*.

370

**CRISPR/Cas9-mediated CBE[+47] and enhancers deletion in ESC**

The CRISPR/Cas9 system was used following published protocols (*Cong et al., 2013; Engreitz et al., 2016; Su et al., 2019*). Briefly, target-specific guide RNAs (sgRNAs) were designed using an online tool (http://chopchop.cbu.uib.no/). sgRNAs with the appropriate site and score were selected. sgRNA sequences are shown in *Supplementary file 2*. For CBE 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 hours later, cells were treated with 5 μM puromycin (MCE, Lot # 64358) for 24 hours 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. Genotyping PCR primers are listed in *Supplementary file 3*.

**Alkaline Phosphatase (AP) Staining analysis in WT and CBE[+47] deletion cells**

Low-density ESC were planted in 12-well plates coated with gelatin for 4 days. Then, ESC were washed twice with PBS and incubated with AP Staining kit™ (Cat # AP100R-1, System Biosciences) following the manufacturer's instructions. Digital images were taken by Olympus Inverted Fluorescence Microscope.

**Enhancer Capture-C and data analysis**

The Capture-C probes were designed using an online tool within the three enhancers regions: (http://apps.molbiol.ox.ac.uk/CaptureC/cgi-bin/CapSequm.cgi) E1 bait (Mouse, mm10, chr6: 52012160 – 52015360), E2 bait (Mouse, mm10, chr6:52075358 – 52076639), E3 bait (Mouse, mm10, chr6: 52101884 – 52104217). Probe sequences are listed in *Supplementary file 4*.

Capture-C libraries were prepared as previously described with minor modifications (*Davies et al., 2016; Hughes et al., 2014; Su et al., 2019*). 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)]. Chromatins was digested with DpnII (New England Biolabs, Lot # 10014860) at

400   37°C overnight. Fragments were then diluted and ligated with T4 DNA ligase (Takara, Lot #

401   1211707) at 16°C overnight. Cross-linking was reversed by overnight incubation at 60°C with

402   proteinase K (Bioline, Lot # BIO-37037). Then 3C libraries were purified by phenol-chloroform

403   followed by chloroform extraction and ethanol-precipitated at −80°C overnight. Sequencing

404   libraries were prepared from 10 μg of the 3C library by sonication to an average size of 200 ~ 300

405   bp and indexed using NEBnext reagents (New England Biolabs, Lot # 0031607), according to the

406   manufacturer's protocol. Enrichment of 2 μg of an indexed library incubated with 3 μM of a pool of

407   biotinylated oligonucleotides was performed using the SeqCap EZ Hybridization reagent kit

408   (Roche/NimbleGen, Lot # 05634261001), following the manufacturer's instructions. Two rounds of

409   capture employing 48~72 and 24h hybridizations, respectively, were used. Correct library size was

410   confirmed by agarose gel electrophoresis, and DNA concentrations were determined using a Qubit

411   2.0 Fluorometer (Thermo Fisher Scientific, Lot # 0000248352). All sequencing was performed on

412   Hi-Seq 2500 platforms using paired 150 bp protocols (Illumina).

413   Capture-C data were analyzed using the previously described methods (*Davies et al., 2016;*

414   *Hughes et al., 2014; Su et al., 2019*). Briefly, the clean paired-

415   end reads were reconstructed into single reads using FLASH (*Magoc & Salzberg, 2011*). After in

416   silico DpnII digestion using the DpnII2E.pl script, the reads were mapped back to the mm10 mouse

417   genome using Bowtie1. At last, chimeric reads containing captured reads and Capture-Reporter

418   reads were analyzed using CCanalyser3.pl. The results can be visualized using the Integrated

419   Genome Browser (IGV) (*Freese, Norris, & Loraine, 2016*) and online tool (https://epgg-

420   test.wustl.edu/browser/) (*X. Zhou et al., 2013*).

421

422   **RNA-seq and bioinformatics analysis**

423   ESC were lysed with Trizol reagent (Life Technologies, Lot # 265709) and RNA was extracted

424   according to the manufacturer's instructions. RNA was then sent to a sequencing company

425   (Novogene, Tianjin, China) for sequencing. Clean reads were mapped to Ensemble mm10 mouse

426   genome using Hisat2 with default parameters. Gene reads were counted by Htseq (*Anders, Pyl, &*

427   *Huber, 2015*). Fold changes were computed as the log2 ratio of normalized reads per gene using

428   the DEseq2 R package (*Love, Huber, & Anders, 2014*). Genes expression with $|$ log2 (fold change)

429   $| \geqslant 1$ ($P < 0.05$) were considered as significantly altered. Heatmaps were drawn with heatmap2.

430   Two biological replicates were analyzed for each experimental condition.

431

432   **Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway**

433   **analyses**

434   GO and KEGG pathway analyses were performed using the online tools: Metascape

435   (http://metascape.org/gp/index.html#/main/step1) (*Y. Zhou et al., 2019*) and DAVID Functional

436   Annotation Bioinformatics Microarray Analysis tool (http://david.abcc.ncifcrf.gov/) (*Huang,*

437   *Sherman, & Lempicki, 2009*).

438

439   **Statistical analyses**

440   Data were analyzed by Student's $t$-test (Two-tailed) unless otherwise specified. Statistically

441   significant $P$-values are indicated in Figures as follows: * indicates $P < 0.05$, ** indicates $P < 0.01$,

442   *** indicates $P < 0.001$.

443

444   **Data availability**

445   The data that support this study are available from the corresponding author upon request. The

446   raw data reported in this manuscript for the DNA-seq and RNA-seq data have been deposited in

447   NCBI Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) under the accession

448   number **GSE154495.** The published datasets were download using online tools

449   (http://cistrome.org/db/#/ and http://promoter.bx.psu.edu/hi-c/) and analyzed in this study (*Y. Wang*

450   *et al., 2018; R. Zheng et al., 2019*).

451

452   **The following previously published data sets were used**

453   NCBI Gene Expression Omnibus

454   1.   Ali Shilatifard (2017)

455   ID GSM2588420. SET1A/COMPASS and shadow enhancers in the regulation of homeotic gene

456   expression

457

458   NCBI Gene Expression Omnibus

459    2. Su G, Guo D, Lu W (2019)

460    ID GSE124306. A distal enhancer maintaining Hoxa1 expression orchestrates retinoic acid-

461    induced early ESCs differentiation.

462

463    NCBI Gene Expression Omnibus

464    3. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos G, Lubling Y, Xu X, Lv

465        X, Hugnot J, Tanay A, Cavalli G (2017)

466    ID GSE96107. Multi-scale 3D genome rewiring during mouse neural development.

467

468    NCBI Gene Expression Omnibus

469    4. Chieh-Chun Chen (2012)

470    ID GSM859491. Genome-wide analysis of K3K27ac, H2AZ and methylation data in mouse

471    embryonic stem cells.

472

473    NCBI Gene Expression Omnibus

474    5. Ali Shilatifard (2018)

475    ID GSM2630487. An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and

476    pluripotency transition in embryonic stem cells.

477

478    NCBI Gene Expression Omnibus

479    6. Pengfei Yu (2012)

480    ID GSM881353. Genome-wide analysis of histone modification, protein-DNA binding, cytosine

481    methylation and transcriptome data in mouse and human ES cells and pig iPS cells.

482

483    NCBI Gene Expression Omnibus

484    7. Hendrik Marks (2014)

485    ID GSM1258237. Mll2 is required for H3K4 trimethylation on bivalent promoters in ES cells

486    whereas Mll1 is redundant.

487

488    NCBI Gene Expression Omnibus

489     8.   Lusy Handoko (2011)

490     ID GSM699165. CTCF-Mediated Functional Chromatin Interactome in Pluripotent Cells.

491

492     NCBI Gene Expression Omnibus

493     9.   Richard A Young (2017)

494     ID GSM2645432. YY1 is a structural regulator of enhancer-promoter loops.

495

496     NCBI Gene Expression Omnibus

497     10.  Richard A Young (2014)

498     ID GSM1439567. he developmental potential of iPSCs is greatly influenced by the selection of

499     the reprogramming factors.

500

501     NCBI Gene Expression Omnibus

502     11.  Richard A Young Consortium (2010)

503     ID GSM560345. Control of Embryonic Stem Cell State by Mediator and Cohesin.

504

505     NCBI Gene Expression Omnibus

506     12.  Hendrik Marks (2014)

507     ID GSM1276711. Mll2 is required for H3K4 trimethylation on bivalent promoters in ES cells

508     whereas Mll1 is redundant.

509

510     NCBI Gene Expression Omnibus

511     13.  ENCODE DCC (2014)

512     ID GSM1014154. A comparative encyclopedia of DNA elements in the mouse genome.

513

514     NCBI Gene Expression Omnibus

515     14.  Xiaoying Bai (2017)

516     ID GSM2651154. Transcription Pausing Regulates Mouse Embryonic Stem Cell Differentiation.

517

523    **COMPETING INTERESTS**

524    The authors declare that they have no conflict of interest.

525

526    **Author contributions**

527    W.L., G.S. and L.Z. conceived and supervised the study and designed the experiments. G.S., J.C.,

528    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.

529    analyzed sequencing data. G.S. and W.W. performed bioinformatic analysis. G.S., W.L. and L.Z

530    wrote the manuscript. All authors have read and approved the final manuscript.

531

532

533 **REFERENCES**

534 Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput
535     sequencing data. *Bioinformatics, 31*(2), 166-169. doi: 10.1093/bioinformatics/btu638

536 Arzate-Mejia, R. G., Recillas-Targa, F., & Corces, V. G. (2018). Developing in 3D: the role of CTCF in cell
537     differentiation. *Nat Rev Mol Cell Biol, 145*(6). doi: 10.1038/s41580-019-0132-4

538 10.1242/dev.137729

539 Ba, Z., Lou, J., Ye, A. Y., Dai, H.-Q., Dring, E. W., Lin, S. G., . . . Alt, F. W. (2020). CTCF orchestrates long-
540     range cohesin-driven V(D)J recombinational scanning. *Nature*. doi: 10.1038/s41586-020-2578-
541     0

542 Bain, G., Kitchens, D., Yao, M., Huettner, J. E., & Gottlieb, D. I. (1995). Embryonic stem cells express
543     neuronal properties in vitro. *Dev Biol, 168*(2), 342-357. doi: 10.1006/dbio.1995.1085

544 Beagan, J. A., Duong, M. T., Titus, K. R., Zhou, L., Cao, Z., Ma, J., . . . Phillips-Cremins, J. E. (2017). YY1 and
545     CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment.
546     *Development, 27*(7), 1139-1152. doi: 10.1242/dev.137729

547 10.1101/gr.215160.116

548 Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., . . . Young, R. A. (2005). Core
549     transcriptional regulatory circuitry in human embryonic stem cells. *Cell, 122*(6), 947-956. doi:
550     10.1016/j.cell.2005.08.020

551 Cao, K., Collings, C. K., Marshall, S. A., Morgan, M. A., Rendleman, E. J., Wang, L., . . . Shilatifard, A. (2017).
552     SET1A/COMPASS and shadow enhancers in the regulation of homeotic gene expression. *Genes
553     Dev, 31*(8), 787-801. doi: 10.1101/gad.294744.116

554 Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., . . . Zhang, F. (2013). Multiplex genome
555     engineering using CRISPR/Cas systems. *Science, 339*(6121), 819-823. doi:
556     10.1126/science.1231143

557 Davies, J. O., Telenius, J. M., McGowan, S. J., Roberts, N. A., Taylor, S., Higgs, D. R., & Hughes, J. R. (2016).
558     Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods,
559     13*(1), 74-80. doi: 10.1038/nmeth.3664

560 De Kumar, B., Parker, H. J., Parrish, M. E., Lange, J. J., Slaughter, B. D., Unruh, J. R., . . . Krumlauf, R. (2017).
561     Dynamic regulation of Nanog and stem cell-signaling pathways by Hoxa1 during early neuro-
562     ectodermal differentiation of ES cells. *Proc Natl Acad Sci U S A, 114*(23), 5838-5845. doi:
563     10.1073/pnas.1610612114

564 De Kumar, B., Parrish, M. E., Slaughter, B. D., Unruh, J. R., Gogol, M., Seidel, C., . . . Krumlauf, R. (2015).
565     Analysis of dynamic changes in retinoid-induced transcription and epigenetic profiles of
566     murine Hox clusters in ES cells. *Genome Res, 25*(8), 1229-1243. doi: 10.1101/gr.184978.114

567 de Wit, E., Vos, E. S., Holwerda, S. J., Valdes-Quezada, C., Verstegen, M. J., Teunissen, H., . . . de Laat, W.
568     (2015). CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell, 60*(4), 676-684. doi:
569     10.1016/j.molcel.2015.09.023

570 Dowen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., . . . Young, R. A. (2014). Control
571     of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell,
572     159*(2), 374-387. doi: 10.1016/j.cell.2014.09.030

573 Engreitz, J. M., Haines, J. E., Perez, E. M., Munson, G., Chen, J., Kane, M., . . . Lander, E. S. (2016). Local
574     regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature,
575     539*(7629), 452-455. doi: 10.1038/nature20149

576 Ferraiuolo, M. A., Rousseau, M., Miyamoto, C., Shenker, S., Wang, X. Q., Nadler, M., . . . Dostie, J. (2010).

577   The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res, 38*(21), 7472-
578       7484. doi: 10.1093/nar/gkq644

579   Freese, N. H., Norris, D. C., & Loraine, A. E. (2016). Integrated genome browser: visual analytics platform
580       for genomics. *Bioinformatics, 32*(14), 2089-2095. doi: 10.1093/bioinformatics/btw069

581   Fujikura, J., Yamato, E., Yonemura, S., Hosoda, K., Masui, S., Nakao, K., . . . Niwa, H. (2002). Differentiation
582       of embryonic stem cells is induced by GATA factors. *Genes Dev, 16*(7), 784-789. doi:
583       10.1101/gad.968802

584   Gajovic, S., St-Onge, L., Yokota, Y., & Gruss, P. (1997). Retinoic acid mediates Pax6 expression during in
585       vitro differentiation of embryonic stem cells. *Differentiation, 62*(4), 187-192. doi:
586       10.1046/j.1432-0436.1998.6240187.x

587   Gaszner, M., & Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms.
588       *Nat Rev Genet, 7*(9), 703-713. doi: 10.1038/nrg1925

589   Ghasemi, R., Struthers, H., Wilson, E. R., & Spencer, D. H. (2020). Contribution of CTCF binding to
590       transcriptional activity at the HOXA locus in NPM1-mutant AML cells. *Leukemia*. doi:
591       10.1038/s41375-020-0856-3

592   Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., . . . Wu, Q. (2015). CRISPR Inversion of CTCF Sites
593       Alters Genome Topology and Enhancer/Promoter Function. *Cell, 162*(4), 900-910. doi:
594       10.1016/j.cell.2015.07.038

595   Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., & Tilghman, S. M. (2000). CTCF
596       mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature,*
597       *405*(6785), 486-489. doi: 10.1038/35013106

598   Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene
599       lists using DAVID bioinformatics resources. *Nat Protoc, 4*(1), 44-57. doi:
600       10.1038/nprot.2008.211

601   Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., . . . Higgs, D. R. (2014).
602       Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-
603       throughput experiment. *Nat Genet, 46*(2), 205-212. doi: 10.1038/ng.2871

604   Jain, S., Ba, Z., Zhang, Y., Dai, H. Q., & Alt, F. W. (2018). CTCF-Binding Elements Mediate Accessibility of
605       RAG Substrates During Chromatin Scanning. *Cell, 174*(1), 102-116.e114. doi:
606       10.1016/j.cell.2018.04.035

607   Ji, X., Dadon, D. B., Powell, B. E., Fan, Z. P., Borges-Rivera, D., Shachar, S., . . . Young, R. A. (2016). 3D
608       Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem Cell, 18*(2), 262-275.
609       doi: 10.1016/j.stem.2015.11.007

610   Kashyap, V., Gudas, L. J., Brenet, F., Funk, P., Viale, A., & Scandura, J. M. (2011). Epigenomic
611       reorganization of the clustered Hox genes in embryonic stem cells induced by retinoic acid. *J*
612       *Biol Chem, 286*(5), 3250-3260. doi: 10.1074/jbc.M110.157545

613   Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., . . . Ren, B. (2007).
614       Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell,*
615       *128*(6), 1231-1245. doi: 10.1016/j.cell.2006.12.048

616   Korkmaz, G., Manber, Z., Lopes, R., Prekovic, S., Schuurman, K., Kim, Y., . . . Agami, R. (2019). A CRISPR-
617       Cas9 screen identifies essential CTCF anchor sites for estrogen receptor-driven breast cancer
618       cell proliferation. *Nucleic Acids Res, 47*(18), 9557-9572. doi: 10.1093/nar/gkz675

619   Kurukuti, S., Tiwari, V. K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., . . . Ohlsson, R. (2006).
620       CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order

621          chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci U S A, 103*(28),
622          10684-10689. doi: 10.1073/pnas.0600326103

623 Levasseur, D. N., Wang, J., Dorschner, M. O., Stamatoyannopoulos, J. A., & Orkin, S. H. (2008). Oct4
624          dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes Dev,*
625          *22*(5), 575-580. doi: 10.1101/gad.1606308

626 Li, Y., Liao, Z., Luo, H., Benyoucef, A., Kang, Y., Lai, Q., . . . Huang, S. (2020). Alteration of CTCF-associated
627          chromatin neighborhood inhibits TAL1-driven oncogenic transcription program and
628          leukemogenesis. *Nucleic Acids Res, 48*(6), 3119-3133. doi: 10.1093/nar/gkaa098

629 Liang, Y. C., Wu, P., Lin, G. W., Chen, C. K., Yeh, C. Y., Tsai, S., . . . Chuong, C. M. (2020). Folding Keratin
630          Gene Clusters during Skin Regional Specification. *Dev Cell, 53*(5), 561-576.e569. doi:
631          10.1016/j.devcel.2020.05.007

632 Liu, G. Y., Zhao, G. N., Chen, X. F., Hao, D. L., Zhao, X., Lv, X., & Liu, D. P. (2016). The long noncoding RNA
633          Gm15055 represses Hoxa gene expression by recruiting PRC2 to the gene cluster. *Nucleic Acids*
634          *Res, 44*(6), 2613-2627. doi: 10.1093/nar/gkv1315

635 Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W., Chen, X., . . . Ng, H. H. (2006). The Oct4 and Nanog
636          transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet, 38*(4),
637          431-440. doi: 10.1038/ng1760

638 Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-
639          seq data with DESeq2. *Genome Biol, 15*(12), 550. doi: 10.1186/s13059-014-0550-8

640 Luo, H., Sobh, A., Vulpe, C. D., Brewer, E., Dovat, S., Qiu, Y., & Huang, S. (2019). HOX Loci Focused
641          CRISPR/sgRNA Library Screening Identifying Critical CTCF Boundaries. *J Vis Exp*(145). doi:
642          10.3791/59382

643 Luo, H., Wang, F., Zha, J., Li, H., Yan, B., Du, Q., . . . Huang, S. (2018). CTCF boundary remodels chromatin
644          domain and drives aberrant HOX gene transcription in acute myeloid leukemia. *Blood, 132*(8),
645          837-848. doi: 10.1182/blood-2017-11-814319

646 Lupianez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., . . . Mundlos, S. (2015).
647          Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer
648          interactions. *Cell, 161*(5), 1012-1025. doi: 10.1016/j.cell.2015.04.004

649 Luppino, J. M., Park, D. S., Nguyen, S. C., Lan, Y., Xu, Z., Yunker, R., & Joyce, E. F. (2020). Cohesin promotes
650          stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nat*
651          *Genet, 52*(8), 840-848. doi: 10.1038/s41588-020-0647-9

652 Magoc, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome
653          assemblies. *Bioinformatics, 27*(21), 2957-2963. doi: 10.1093/bioinformatics/btr507

654 Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium
655          conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A, 78*(12), 7634-7638. doi:
656          10.1073/pnas.78.12.7634

657 Martinez-Ceballos, E., Chambon, P., & Gudas, L. J. (2005). Differences in gene expression between wild
658          type and Hoxa1 knockout embryonic stem cells after retinoic acid treatment or leukemia
659          inhibitory factor (LIF) removal. *J Biol Chem, 280*(16), 16484-16498. doi:
660          10.1074/jbc.M414397200

661 Metcalfe, A. D., & Ferguson, M. W. (2007). Tissue engineering of replacement skin: the crossroads of
662          biomaterials, wound healing, embryonic development, stem cells and regeneration. *J R Soc*
663          *Interface, 4*(14), 413-437. doi: 10.1098/rsif.2006.0179

664 Nanni, L., Ceri, S., & Logie, C. (2020). Spatial patterns of CTCF sites define the anatomy of TADs and their

665  boundaries. *Genome Biol, 21*(1), 197. doi: 10.1186/s13059-020-02108-x

666  Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O., & Reinberg, D. (2015). CTCF
667      establishes discrete functional chromatin domains at the Hox clusters during differentiation.
668      *Science, 347*(6225), 1017-1021. doi: 10.1126/science.1262088

669  Nora, E. P., Goloborodko, A., Valton, A. L., Gibcus, J. H., Uebersohn, A., Abdennur, N., . . . Bruneau, B. G.
670      (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains
671      from Genomic Compartmentalization. *Cell, 169*(5), 930-944.e922. doi:
672      10.1016/j.cell.2017.05.004

673  Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., . . . Gilbert, D. M. (2014). Topologically
674      associating domains are stable units of replication-timing regulation. *Nature, 515*(7527), 402-
675      405. doi: 10.1038/nature13986

676  Qian, Y., Zhang, L., Cai, M., Li, H., Xu, H., Yang, H., . . . Lu, W. (2019). The prostate cancer risk variant
677      rs55958994 regulates multiple gene expression through extreme long-range chromatin
678      interaction to control tumor progression. *Sci Adv, 5*(7), eaaw6710. doi:
679      10.1126/sciadv.aaw6710

680  Rhinn, M., & Dolle, P. (2012). Retinoic acid signalling during development. *Development, 139*(5), 843-
681      858. doi: 10.1242/dev.167502
682  10.1242/dev.065938

683  Rousseau, M., Crutchley, J. L., Miura, H., Suderman, M., Blanchette, M., & Dostie, J. (2014). Hox in
684      motion: tracking HoxA cluster conformation during differentiation. *Nucleic Acids Res, 42*(3),
685      1524-1540. doi: 10.1093/nar/gkt998

686  Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., & Brivanlou, A. H. (2004). Maintenance of pluripotency
687      in human and mouse embryonic stem cells through activation of Wnt signaling by a
688      pharmacological GSK-3-specific inhibitor. *Nat Med, 10*(1), 55-63. doi: 10.1038/nm979

689  Sato, T., Kataoka, K., Ito, Y., Yokoyama, S., Inui, M., Mori, M., . . . Asahara, H. (2020). Lin28a/let-7 pathway
690      modulates the Hox code via Polycomb regulation during axial patterning in vertebrates. *Elife,*
691      *9*, e53608. doi: 10.7554/eLife.53608

692  Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method.
693      *Nat Protoc, 3*(6), 1101-1108. doi: 10.1038/nprot.2008.73

694  Simandi, Z., Horvath, A., Wright, L. C., Cuaranta-Monroy, I., De Luca, I., Karolyi, K., . . . Nagy, L. (2016).
695      OCT4 Acts as an Integrator of Pluripotency and Signal-Induced Differentiation. *Mol Cell, 63*(4),
696      647-661. doi: 10.1016/j.molcel.2016.06.039

697  Su, G., Guo, D., Chen, J., Liu, M., Zheng, J., Wang, W., . . . Lu, W. (2019). A distal enhancer maintaining
698      Hoxa1 expression orchestrates retinoic acid-induced early ESCs differentiation. *Nucleic Acids*
699      *Res, 47*(13), 6737-6752. doi: 10.1093/nar/gkz482

700  Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007).
701      Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell, 131*(5),
702      861-872. doi: 10.1016/j.cell.2007.11.019

703  Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and
704      adult fibroblast cultures by defined factors. *Cell, 126*(4), 663-676. doi:
705      10.1016/j.cell.2006.07.024

706  Tay, Y., Zhang, J., Thomson, A. M., Lim, B., & Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2
707      coding regions modulate embryonic stem cell differentiation. *Nature, 455*(7216), 1124-1128.
708      doi: 10.1038/nature07299

709  Valcarcel, R., Holz, H., Jimenez, C. G., Barettino, D., & Stunnenberg, H. G. (1994). Retinoid-dependent in
710      vitro transcription mediated by the RXR/RAR heterodimer. *Genes Dev, 8*(24), 3068-3079. doi:
711      10.1101/gad.8.24.3068

712  Wang, P., Tang, Z., Lee, B., Zhu, J. J., Cai, L., Szalaj, P., . . . Ruan, Y. (2020). Chromatin topology
713      reorganization and transcription repression by PML-RARalpha in acute promyeloid leukemia.
714      *Genome Biol, 21*(1), 110. doi: 10.1186/s13059-020-02030-2

715  Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., . . . Yue, F. (2018). The 3D Genome Browser: a
716      web-based browser for visualizing 3D genome organization and long-range chromatin
717      interactions. *Genome Biol, 19*(1), 151. doi: 10.1186/s13059-018-1519-9

718  Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., . . . Lu, W. (2013). Klf4 organizes long-range chromosomal
719      interactions with the oct4 locus in reprogramming and pluripotency. *Cell Stem Cell, 13*(1), 36-
720      47. doi: 10.1016/j.stem.2013.05.010

721  Yin, Y., Yan, P., Lu, J., Song, G., Zhu, Y., Li, Z., . . . Shen, X. (2015). Opposing Roles for the lncRNA Haunt
722      and Its Genomic Locus in Regulating HOXA Gene Activation during Embryonic Stem Cell
723      Differentiation. *Cell Stem Cell, 16*(5), 504-516. doi: 10.1016/j.stem.2015.03.007

724  Zhang, L., He, A., Chen, B., Bi, J., Chen, J., Guo, D., . . . Lu, W. (2020). A HOTAIR regulatory element
725      modulates glioma cell sensitivity to temozolomide through long-range regulation of multiple
726      target genes. *Genome Res, 30*(2), 155-163. doi: 10.1101/gr.251058.119

727  Zhao, H., Li, Z., Zhu, Y., Bian, S., Zhang, Y., Qin, L., . . . Hao, B. (2020). A role of the CTCF binding site at
728      enhancer Eα in the dynamic chromatin organization of the Tcra–Tcrd locus. *Nucleic Acids Res*.
729      doi: 10.1093/nar/gkaa711

730  Zheng, H., & Xie, W. (2019). The role of 3D genome organization in development and cell differentiation.
731      *Nature Reviews Molecular Cell Biology, 20*(9), 535-550. doi: 10.1038/s41580-019-0132-4

732  Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., . . . Liu, X. S. (2019). Cistrome Data Browser:
733      expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res, 47*(D1), D729-
734      d735. doi: 10.1093/nar/gky1094

735  Zhou, X., Lowdon, R. F., Li, D., Lawson, H. A., Madden, P. A., Costello, J. F., & Wang, T. (2013). Exploring
736      long-range genome interactions using the WashU Epigenome Browser. *Nat Methods, 10*(5),
737      375-376. doi: 10.1038/nmeth.2440

738  Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., . . . Chanda, S. K. (2019).
739      Metascape provides a biologist-oriented resource for the analysis of systems-level datasets.
740      *Nat Commun, 10*(1), 1523. doi: 10.1038/s41467-019-09234-6

741

742

743

744     **MAIN FIGURE LEGENDS**

745     **Figure 1. Identification of a CBE (CBE$^{+47}$) closest to the 3'-end of *HoxA* within the same TAD**

746     **in ESC.**

747     (**A**) Hi-C interaction map of the ~0.5 Mb region surrounding the *HoxA* in undifferentiated ESC.

748     Data were extracted from Bonev B et al. 2017. (**B**) IGV (Integrative Genomics Viewer) screenshots

749     showing gene tracks of CTCF，MED1, MED12 and YY1 ChIP-seq signal occupancy at Skap2 and

750     *HoxA* loci in ESC. CTCF ChIP-seq shows CTCF binding throughout the locus. Multiple CTCF sites

751     are located at the downstream of *HoxA* locus. (**C, D**) IGV view of selected ChIP-seq tracks at Skap2

752     and *HoxA* loci in ESC. Shown are H3K27ac, H3K4me1, H3K4me2, H3K4me3, Pol ll, ATAC-seq

753     and DNase. Blue shadowing indicates the CBE$^{+47}$ region. Grey shadowing shows the *HoxA* locus.

754

755     **Figure 2. Gene expression analysis in WT and CBE$^{+47}$ deletion cells under self-renewal culture**

756     **condition.**

757     (**A**) Schematic showing CRISPR/Cas9-mediated deletion of the CBE$^{+47}$ (blue shadowing) using two

758     sgRNAs. Indicated primer was used to distinguish CBE$^{+47}$-KO (CBE$^{+47}$ knockout) from wild-type

759     (WT) clones. (**B**) Validation of knockout cell lines by genomic DNA PCR and sequencing. Shown

760     are images from representative clones. DNA sequencing of two CBE$^{+47}$-KO cell clones (#1 and #2)

761     using the primer. (**C**) Overview appearance of WT and CBE$^{+47}$-KO cells with alkaline

762     phosphatase (AP) staining. (**D-K**) qRT-PCR of WT and CBE$^{+47}$-KO cells showing mRNA

763     expression of *Skap2* and *Halr1* (**D**), *HoxA* (**E**), pluripotency genes (**F**), epiblast genes (**G**), endoderm

764     genes (**H**), mesoderm genes (**I**), trophectoderm genes (**J**) and ectoderm genes (**K**) in undifferentiated

765     ESC. In (**D-K**), data are represented as min to max. The expression level in WT ESC was relatively

766     set to 1 (As shown in blue dotted line). Indicated significance is based on Student's *t*-test (Two-

767     tailed, * indicates $P < 0.05$, ** indicates $P < 0.01$, *** indicates $P < 0.001$). In (**D-K**), n = 12,

768     including 2 CBE$^{+47}$-KO cell lines (CBE$^{+47}$-KO1# and CBE$^{+47}$-KO2#), and 2 biological replicates

769     per cell line with 3 technical replicates in 1 biological replicate. In (**B**), M: DNA Marker. In (**C**),

770     scale bar, 50 μm.

771     Source files are available in *Figure 2—source data 1*.

772

773     **Figure 3. CBE$^{+47}$ deletion significantly promotes RA-induced expressions of *HoxA* and early**

774 **ESC abnormal differentiation.**

775 (**A-C**) qRT-PCR of WT and CBE$^{+47}$-KO cells showing expressions of *HoxA* in ESC after RA

776 treatment (12h, 24h and 48h). (**D-F**) qRT-PCR of WT and CBE$^{+47}$-KO ESC showing mRNA

777 expressions of pluripotency genes, epiblast genes, endoderm genes, mesoderm genes,

778 trophectoderm genes and ectoderm genes in RA-induced ESC (12h, 24h and 48h). In (**A-F**), data

779 are represented as min to max. The expression level in WT ESC was relatively set to 1 (As shown

780 in blue dotted line). Indicated significance is based on Student's *t*-test (Two-tailed, * indicates

781 $P < 0.05$, ** indicates $P < 0.01$, *** indicates $P < 0.001$). In (**A-F**), n = 12, including 2 CBE$^{+47}$-KO

782 cell lines (CBE$^{+47}$-KO1# and CBE$^{+47}$-KO2#), and 2 biological replicates per cell line with 3

783 technical replicates in 1 biological replicate.

784 Source files are available in *Figure 3—source data 1*.

785

786 **Figure 4. Transcriptome analysis in WT and CBE$^{+47}$ deletion cells during RA-induced early**

787 **ESC differentiation.**

788 (**A**) Heatmap depicting gene expression changes in WT and CBE$^{+47}$-KO cells treated with RA for

789 24 h (Fold-change ⩾ 2, and $P < 0.05$, as determined by DESeq2). (**B**) RNA-seq results shown the

790 Heatmap of *HoxA*. (**C-D**) GO-BP and KEGG pathway analyses indicated differentially expressed

791 genes. In (**E**), the red font shows the signaling pathways of up-regulated genes enrichment, and the

792 blue font shows the signaling pathways of down-regulated genes enrichment.

793 Source files are available in *Figure 4—source data 1*.

794

795 **Figure 5. The CBE$^{+47}$ is required for proper chromatin interactions between its adjacent**

796 **enhancers and *HoxA*.**

797 (**A**) IGV (Integrative Genomics Viewer) screenshots showing gene tracks of H3K27ac (+RA, 24h)

798 and CTCF ChIP-seq signal occupancy at *Skap2* and *HoxA* loci in WT ESC. (**B-D**) Chromatin

799 interaction profiles by using enhancers as the anchors in WT and CBE$^{+47}$-KO cells under RA

800 treatment for 24h. Y-axis denotes interaction frequency represented by the number. Light green,

801 light purple, light red and light grey shadows mark the E1, E2, E3 and *HoxA* regions, respectively.

802

803 **Figure 6. Multiple enhancers deletion shows synergistic effect on RA-induced *HoxA***

804    **expression. *HoxA* mRNA levels were measured by qRT–PCR and normalized to WT levels in**

805    **ΔE1 (A), ΔE2 (B), ΔE3 (C), ΔE1/2 (D) andΔE1/2/3 (E) cells under RA treatment for 24h.**

806    In (**A-E**), data are represented as min to max. The expression level in WT ESC was relatively set to

807    1 (As shown in blue dotted line). Indicated significance is based on Student's *t*-test (Two-tailed, *

808    indicates $P < 0.05$, ** indicates $P < 0.01$, *** indicates $P < 0.001$). In (**A-C, E**), n = 12, including 2

809    enhancer knockout cell lines, and 2 biological replicates per cell line with 3 technical replicates in

810    1 biological replicate. In (**D**), n = 6, including 1 enhancer knockout cell line, and 2 biological

811    replicates per cell line with 3 technical replicates in 1 biological replicate.

812    Source files are available in *Figure 6—source data 1*.

813

814    **Figure 7. Proposed schematic illustration of the CBE$^{+47}$ that regulates RA-induced *HoxA***

815    **expression and early ESC differentiation through orchestrating long-range chromatin**

816    **interactions between its adjacent enhancers and *HoxA*.**

817    The CBE$^{+47}$ working model is shown: in CBE$^{+47}$-WT cells (Left), the three enhancers near CBE$^{+47}$

818    acted as EEIC to interact with the 3'-end of *HoxA* chromatin to maintain *HoxA* normal expression,

819    thereby promoting early proper differentiation of ESC induced by RA; in CBE$^{+47}$-KO cells (Right),

820    the absence of CBE$^{+47}$ leads to reduced interactions between enhancers, thus forming a relatively

821    loose EEIC, at the same time, interactions between EEIC and the central chromatin region of *HoxA*

822    are increased, then promoting over-expressions of *HoxA*, further resulting in early ESC abnormal

823    differentiation.

824

825    **Figure 1—figure supplement 1. CBE$^{+47}$ is located at the second intron of *Halr1* in ESC.**

826    (**A**) IGV (Integrative Genomics Viewer) screenshot showing refseq gene locus (*Halr1*).

827    (**B**) IGV screenshots showing gene tracks of CTCF，MED1, MED12 and YY1 ChIP-seq signal

828    occupancy at *Halr1* locus in ESC.

829    (**C-D**) IGV view of selected ChIP-seq tracks at Halr1 locus in ESC. Shown are H3K27ac, H3K4me1,

830    H3K4me2, H3K4me3, Pol ll, ATAC-seq, and DNase. Blue shadowing indicates the CBE$^{+47}$ region.

831

832    **Figure 3—figure supplement 1. qRT-PCR analyses *HoxA* expression during RA-induced**

833    **ESC differentiation.**

834      qRT-PCR results showing expressions of *HoxA* in ESC after RA treatment (0h, 12h, 24h and 48h).

835      The expression level in untreated ESC (0h) was relatively set to 1. Data are represented as mean ±

836      SD. N = 6, including 1 cell line (ESC), and 2 biological replicates per cell line with 3 technical

837      replicates in 1 biological replicate.

838      Source files are available in *Figure 3—figure supplement 1—source data 1.*

839

840      **Figure 3—figure supplement 2. qRT-PCR analyses expressions of *HoxA*-related lncRNAs**

841      **in WT and CBE$^{+47}$-KO cells during RA-induced differentiation.**

842      (**A**) Diagram showing RA-induced WT and CBE$^{+47}$-KO cells differentiation.

843      (**B-D**) qRT-PCR analyses *HoxA*-related lncRNAs expression in WT and CBE$^{+47}$-KO cells during

844      RA-induced differentiation.

845      In (**B-D**), data are represented as min to max. The expression level in WT ESC was relatively set to

846      1 (As shown in blue dotted line). Indicated significance is based on Student's *t*-test (Two-tailed, *

847      indicates $P < 0.05$, ** indicates $P < 0.01$, *** indicates $P < 0.001$).

848      In (**B-D**), n = 12, including 2 CBE$^{+47}$-KO cell lines (CBE$^{+47}$-KO1# and CBE$^{+47}$-KO2#), and 2

849      biological replicates per cell line with 3 technical replicates in 1 biological replicate.

850      Source files are available in *Figure 3—figure supplement 2—source data 1.*

851

852      **Figure 4—figure supplement 1. CBE$^{+47}$ served as a *cis*-regulatory element regulates gene**

853      **expression within the same TAD.**

854      (**A**) Hi-C interaction map of the ~ 2.0 Mb region surrounding the *HoxA* in undifferentiated ESC.

855      Data were extracted from Bonev B et al. 2017.

856      (**B**) RNA-seq results shown the Heatmap of *HoxA*.

857      (**C-D**) data show that CBE$^{+47}$ as cis-regulatory element regulates gene expression within the same

858      TAD.

859      In (**D**), NS: not significant. Indicated significance is based on Student's t-test (Two-tailed, *

860      indicates $P < 0.05$, ** indicates $P < 0.01$, *** indicates $P < 0.001$).

861      Source files are available in *Figure 4—figure supplement 1—source data 1.*

862

863      **Figure 4—figure supplement 2.** RNA-seq results shown the Heatmap of *HoxB/C/D* cluster

864     genes (**A**), RA signaling-(**B**) and pluripotency- (**C**) related genes.

865

866     **Figure 4—figure supplement 3. Down-regulated and up-regulated genes to enriched GO-**

867     **BP clusters.**

868

869     **Figure 6—figure supplement 1. CRISPR-cas9 mediated enhancers knockout.**

870     (**A**) Schematic showing CRISPR/Cas9-mediated deletion of the enhancers using specially paired

871     sgRNAs (Red silk arrows). Indicated primers (Blue silk arrows) were used to distinguish enhancers

872     knockout from wild type clones.

873     (**B**) Validation of knockout cell lines by genomic DNA PCR. Agarose gel electrophoresis images

874     from representative clones. The red dashed boxes mark the target genomic DNA PCR products.

875     ΔE3 cell lines have been reported in our previous study (Su et al., 2019).

876

877     **Figure 6—figure supplement 2. CBE$^{+47}$ and enhancers regulate the *HoxA* expression.**

878     (**A**) The data show that CBE$^{+47}$ and enhancers have inhibitory and enhanced regulatory effects on

879     *HoxA* expression, respectively.

880     (**B**) Comparing the difference in regulation of *HoxA* between enhancers. Indicated significance (*P*-

881     value) is based on Student's *t*-test (Two-tailed).

882     Source files are available in *Figure 6—figure supplement 2—source data 1*.

883

884

885    **FIGURE SUPPLEMENTS**

886    **Figure 1—figure supplement 1. CBE$^{+47}$ is located at the second intron of Halr1 in ESC.**

887    **Figure 3—figure supplement 1. qRT-PCR analyses *HoxA* expression during RA-induced**

888    **ESC differentiation.**

889    **Figure 3—figure supplement 2. qRT-PCR analyses expressions of *HoxA*-related lncRNAs**

890    **in WT and CBE$^{+47}$-KO cells during RA-induced differentiation.**

891    **Figure 4—figure supplement 1. CBE$^{+47}$ served as a *cis*-regulatory element regulates gene**

892    **expression within the same TAD.**

893    **Figure 4—figure supplement 2.** RNA-seq results shown the Heatmap of HoxB/C/D cluster

894    genes (**A**), RA signaling-(**B**) and pluripotency- (**C**) related genes.

895    **Figure 4—figure supplement 3. Down-regulated and up-regulated genes to enriched GO-**

896    **BP clusters.**

897    **Figure 6—figure supplement 1. CRISPR-cas9 mediated enhancers knockout.**

898    **Figure 6—figure supplement 2. CBE$^{+47}$ and enhancers regulate the *HoxA* expression.**

899

900    **SUPPLEMENTARY FILES**

901    **Supplementary file 1. Sequences of qRT-PCR primers (5'-3') used in this study**

902    **Supplementary file 2. sgRNAs used in this study**

903    **Supplementary file 3. Primers were used to identify the deletion clones in this study**

904    **Supplementary file 4. Synthetic oligos for Capture-C (5'-3') (5'-biotin)**

905

906    **SOURCE DATA FILES**

907    **Figure 2—source data 1.xlsx**

908    **Figure 3—source data 1 .xlsx**

909    **Figure 4—source data 1.xlsx**

910    **Figure 6—source data 1.xlsx**

911    **Figure 3—figure supplement 1—source data 1.xlsx**
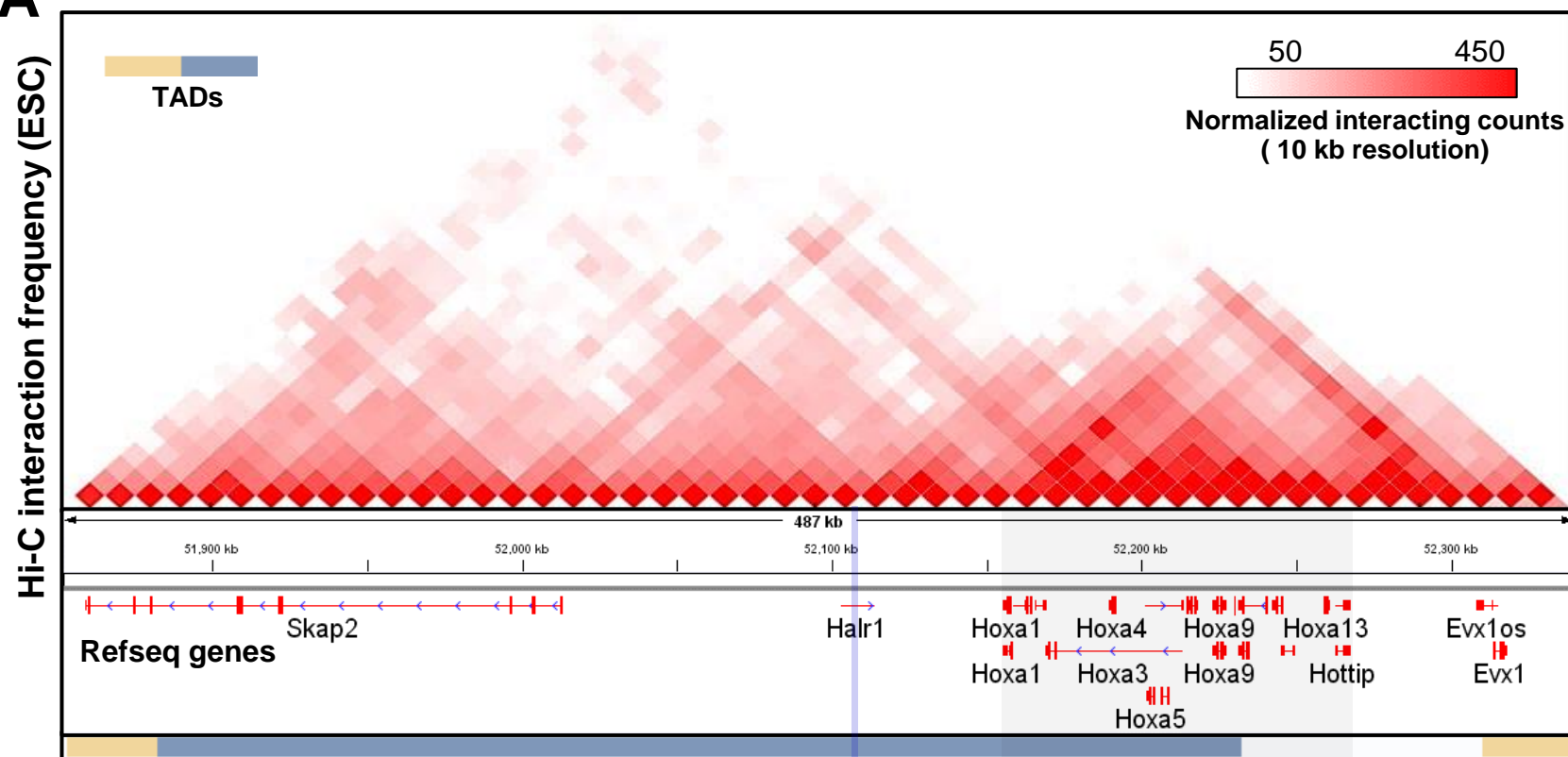
912    **Figure 3—figure supplement 2—source data 1.xlsx**

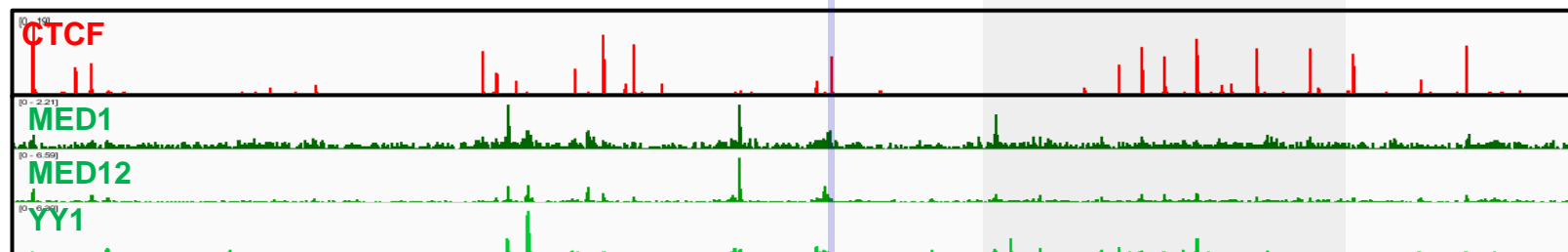913    **Figure 4—figure supplement 1—source data 1.xlsx**
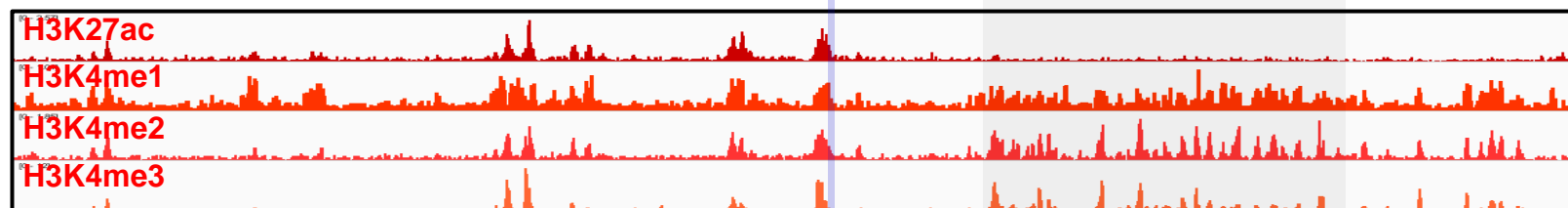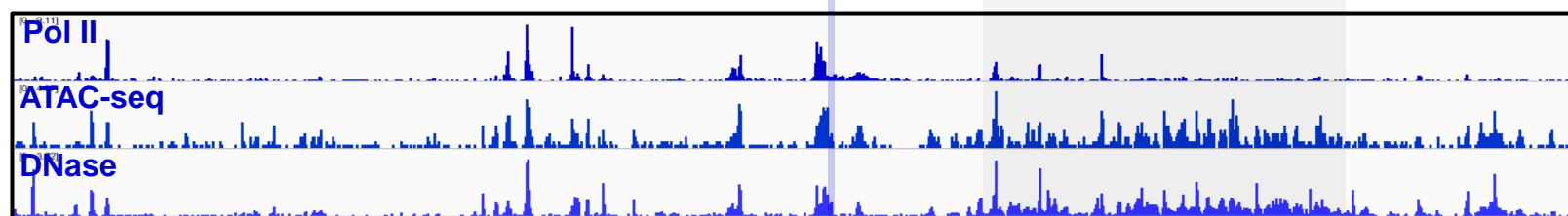
914    **Figure 6—figure supplement 2—source data 1.xlsx**

# Figure 1

# Figure 2

# Figure 3

# Figure 4



**A**   + RA (24h)   WT   CBE$^{+47}$-KO

Up-regulated genes (n = 869)

Down-regulated genes (n = 360)

−1 0 1 Row Z−Score

**B**   RNA-seq   52,160 kb   52,180 kb   52,200 kb   52,220 kb   52,240 kb   52,260 kb

Refseq genes

Hoxa1   Hoxaas2   Hoxa4   Hoxaas3   Hoxa9   Hoxa10   Hoxa13

Hoxa1   Hoxa3   Hoxa9   Hoxa11os   Hottip

Hoxa5

Normalized reads — WT: Rep1, Rep2 [0 - 850]

CBE$^{+47}$-KO: Rep1, Rep2 [0 - 850]

**C** Up-regulated genes (GO-BP terms, top 20)

- regionalization
- gland development
- appendage morphogenesis
- positive regulation of nervous system development
- head development
- chondrocyte development
- cardiac chamber morphogenesis
- regulation of epithelial cell proliferation
- axon development
- regulation of osteoblast differentiation
- positive regulation of receptor binding
- cellular process involved in reproduction in multicellular organism
- regulation of animal organ morphogenesis
- ameboidal-type cell migration
- behavior
- negative regulation of cell differentiation
- cell fate commitment
- supramolecular fiber organization
- dorsal/ventral pattern formation
- regulation of synapse assembly

P-value: $10^{-2}$ $10^{-3}$ $10^{-4}$ $10^{-6}$ $10^{-10}$ $10^{-20}$

**D** Down-regulated genes (GO-BP terms, top 20)

- embryonic placenta development
- positive regulation of vasoconstriction
- regulation of norepinephrine secretion
- blood vessel morphogenesis
- epithelial cell differentiation
- negative regulation of response to external stimulus
- urogenital system development
- regulation of Ras protein signal transduction
- regulation of fibroblast growth factor receptor signalin
- Arachidonic acid metabolism
- negative T cell selection
- T cell proliferation
- regulation of transmembrane transport
- Transcriptional misregulation in cancer
- negative regulation of wound healing
- regulation of fibroblast proliferation
- chemotaxis
- apoptotic mitochondrial changes
- anion transport
- morphogenesis of a branching structure

P-value: $10^{-2}$ $10^{-3}$ $10^{-4}$ $10^{-5}$ $10^{-10}$ $10^{-20}$

**E**

| KEGG-pathway terms | *P*-values | Genes |
|---|---|---|
| cGMP-PKG signaling pathway | 0.00004 | Atp1a2 Atp1b2 Adcy3 Adra2c Agtr1a Bdkrb2 Creb3l4 Itpr1 Irs1 Npr2 Nfatc1 Oprd1 Pde3b Pde5a Prkce Prkg2 Rgs2 Akt3 |
| Neuroactive ligand-receptor interaction | 0.00076 | Htr5a Gpr156 Gpr50 Adra2c Agtr1a Bdkrb2 Cnr1 Cckbr Chrng Gabra1 Gabre Gria4 Grin2c Glrb Lepr Npbwr1 Oprd1 Oprk1 S1pr1 Tacr3 Tspo Vipr1 |
| Axon guidance | 0.0015 | Srgap3 Ablim1 Cxcr4 Dpysl5 Efnb1 Met Nrp1 Nck2 Plxna2 Robo1 Sema3c Sema5a Slit2 |
| Signaling pathways regulating pluripotency of stem cells | 0.0077 | Isl1 Smad3 Dlx5 Fzd1 Fzd4 Fzd8 Id2 Otx1 Pax6 Akt3 Wnt10b Wnt5b |
| Hippo signaling pathway | 0.015 | Patj Rassf6 Smad3 Ccnd2 Fzd1 Fzd4 Fzd8 Gdf6 Id2 Tgfbr2 Wnt10b Wnt5b |
| cAMP signaling pathway | 0.018 | Atp1a2 Atp1b2 Hhip Adcy3 Bdnf Creb3l4 Camk2d Gria4 Grin2c Nfatc1 Ptch1 Pde3b Akt3 Vav3 |
| Wnt signaling pathway | 0.023 | Camk2d Ccnd2 Fzd1 Fzd4 Fzd8 Nkd1 Nfatc1 Prickle2 Sfrp1 Wnt10b Wnt5b |
| Hedgehog signaling pathway | 0.048 | Hhip Dhh Gas1 Ptch1 |
| Proteoglycans in cancer | 0.0035 | Drosha Igf2 Mmp9 Myc Plaur Sdc4 Wnt4 Wnt7b Wnt8a |
| Transcriptional misregulation in cancer | 0.0042 | Cebpb Etv1 Etv4 Mmp9 Myc Per2 Pdgfa Rxrg |
| Arachidonic acid metabolism | 0.0047 | Cyp2b23 Ephx2 Pla2g2e Pla2g4c Ptgis Ptgs1 |
| HTLV-I infection | 0.02 | H2-M2 H2-Eb1 Myc Pdgfa Pold4 Slc25a4 Wnt4 Wnt7b Wnt8a |
| Hippo signaling pathway | 0.039 | Bmp8b Myc Trp73 Wnt4 Wnt7b Wnt8a |
| Thyroid hormone signaling pathway | 0.052 | Myc Plcd1 Rxrg Slco1c1 Wnt4 |

# Figure 5



**A**

E3　E2　E1　HoxA

432 kb

51,900 kb　52,000 kb　52,100 kb　52,200 kb

Refseq genes

Skap2　Halr1

Hoxa1　Hoxa4　Mira　Hoxa11
Hoxa1　Hoxa3　Hoxa9　Hottip
Hoxa5

ESC (WT)

H3K27ac ChIP-seq (+RA 24h)

CTCF ChIP-seq　CBE$^{+47}$

**B**　Anchor: E1

WT　12　0

CBE$^{+47}$-KO　12　0

**C**　Anchor: E2

Capture-C (ESC, +RA 24h)

WT　12　0

CBE$^{+47}$-KO　12　0

**D**　Anchor: E3

WT　8　0

CBE$^{+47}$-KO　8　0

# Figure 6

**Figure 7**