CHD4 Conceals Aberrant CTCF-Binding Sites at TAD Interiors by Regulating Chromatin Accessibility in Mouse Embryonic Stem Cells

Sungwook Han[1,4], Hosuk Lee[1,2,4], Andrew J. Lee[1], Seung-Kyoon Kim[3], Inkyung Jung[1], Gou Young Koh[2], Tae-Kyung Kim[3], and Daeyoup Lee[1,]*

[1]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea, [2]Center for Vascular Research, Institute for Basic Sciences, Daejeon 34141, Korea, [3]Department of Life Sciences, Pohang University of Science and Technology, Pohang 37673, Korea, [4]These authors contributed equally to this work.
*Correspondence: daeyoup@kaist.ac.kr
https://doi.org/10.14348/molcells.2021.0224
www.molcells.org

CCCTC-binding factor (CTCF) critically contributes to 3D chromatin organization by determining topologically associated domain (TAD) borders. Although CTCF primarily binds at TAD borders, there also exist putative CTCF-binding sites within TADs, which are spread throughout the genome by retrotransposition. However, the detailed mechanism responsible for masking the putative CTCF-binding sites remains largely elusive. Here, we show that the ATP-dependent chromatin remodeler, chromodomain helicase DNA-binding 4 (CHD4), regulates chromatin accessibility to conceal aberrant CTCF-binding sites embedded in H3K9me3-enriched heterochromatic B2 short interspersed nuclear elements (SINEs) in mouse embryonic stem cells (mESCs). Upon CHD4 depletion, these aberrant CTCF-binding sites become accessible and aberrant CTCF recruitment occurs within TADs, resulting in disorganization of local TADs. RNA-binding intrinsically disordered domains (IDRs) of CHD4 are required to prevent this aberrant CTCF binding, and CHD4 is critical for the repression of B2 SINE transcripts. These results collectively reveal that a CHD4-mediated mechanism ensures appropriate CTCF binding and associated TAD organization in mESCs.

Keywords: 3D chromatin organization, ATP-dependent

chromatin remodeler, B2 short interspersed nuclear elements, chromodomain-helicase-DNA binding protein 4, intrinsically disordered domains, topologically associated domains

INTRODUCTION

The mammalian genome is spatially arranged into three-dimensional (3D) chromatin organizations that allow for proper gene regulation; these include chromosome territories (Cremer and Cremer, 2010), compartments (the A and B compartments are enriched for active/open chromatin and inactive/closed chromatin, respectively) (Lieberman-Aiden et al., 2009; Simonis et al., 2006), topologically associated domains (TADs) (Dixon et al., 2012; Nora et al., 2012), contact/loop domains (Rao et al., 2014; 2017), and insulated neighborhoods (Dowen et al., 2014; Ji et al., 2016). Previous studies found that the insulator protein, CCCTC-binding factor (CTCF), coexists with the cohesin complex in chromatin (Wendt et al., 2008) and localizes to the anchors/borders of chromatin loops (Dixon et al., 2012; Nora et al., 2012; Rao et al., 2014; Splinter et al., 2006). Furthermore, removal of the cohesin loader, Nipbl (Schwarzer et al., 2017), loss of cohesin itself (Rao et al., 2017), deletion of CTCF-binding sites (de

Wit et al., 2015; Guo et al., 2015; Sanborn et al., 2015), or loss of CTCF itself (Nora et al., 2017) have all been shown to interfere with the maintenance of 3D chromatin organizations, suggesting that these factors all play critical functions. Notably, recent loss-of-function studies using single-cell analysis performed in mammalian cells yielded two interesting observations (Luppino et al., 2020; Szabo et al., 2020). First, the loss of CTCF was found to result in the loss of TAD borders and the subsequent merging of two previously insulated TADs, but did not significantly affect intra-TAD interactions. Second, and in contrast, the loss of the cohesin subunit, RAD21, significantly disrupted intra-TAD interactions but did not significantly affect TAD borders. These findings demonstrated that 3D chromatin organizations are regulated by the combined action of CTCF and cohesin via distinct mechanisms: Cohesin generates intra-TAD contacts, whereas CTCF prevents inter-TAD contacts by determining the TAD borders.

Importantly, in addition to the CTCF-binding sites (or motifs) that exhibit convergent orientation at the border of 3D domains (Rao et al., 2014), there also exist putative CTCF-binding sites at the TAD interiors, which are spread throughout the genome by retrotransposition of short interspersed nuclear element (SINE) retrotransposons (Bourque et al., 2008; Schmidt et al., 2012). From an evolutionary perspective, conserved CTCF-binding sites are enriched at the borders of conserved 3D domains, whereas CTCF-binding sites that diverge between species (divergent CTCF-binding sites) drive local 3D structural changes at the domain interiors (Vietri Rudan et al., 2015). Thus, it is critical to mask the putative CTCF-binding sites that lie within the TADs; otherwise, similar to the divergent CTCF-binding sites, these sites could cause aberrant CTCF recruitment and disrupt the normal 3D chromatin organizations. Although most studies to date have focused on the function of CTCF at the TAD borders, the prevention of inappropriate CTCF binding at TAD interiors is equally essential for the maintenance of 3D chromatin organizations. However, the detailed mechanism responsible for masking these putative CTCF-binding sites remains poorly understood.

To address this issue, we focused on the members of the chromodomain helicase DNA-binding protein (CHD) family, especially the well-known ATP-dependent chromatin remodeler, CHD4. The CHD family is evolutionarily conserved (Hall and Georgel, 2007), and its members contribute to diverse cellular processes by assembling nucleosomes (de Dieuleveult et al., 2016; Gaspar-Maia et al., 2009; Micucci et al., 2015; Skene et al., 2014). Since the majority of CTCF-binding sites are located at nucleosome-free regions (Cuddapah et al., 2009), it has been proposed that chromatin remodelers could govern CTCF recruitment by regulating nucleosome occupancy at certain CTCF-binding sites (Ong and Corces, 2014). Furthermore, the ability of chromatin remodelers to regulate access to DNA through nucleosome positioning may contribute to obscuring the putative CTCF-binding sites within TADs.

Here, we show that CHD4 regulates chromatin accessibility to conceal aberrant CTCF-binding sites within TADs, thereby preventing aberrant CTCF recruitment and securing the 3D chromatin organizations in mESCs. We used various next-generation sequencing (NGS) assays, including *in situ* Hi-C, MNase-seq, ATAC-seq, and H3/CTCF ChIP-seq, and per-

formed temporal depletion/restoration of CHD4 to confirm the order of events. We found that CHD4 initially assembles core histones to conceal aberrant CTCF-binding sites, and thereby prevents aberrant CTCF binding. We observed that RNA-binding intrinsically disordered domains (IDRs) of CHD4 are required to prevent aberrant CTCF recruitment. Finally, we discovered that the CHD4-regulated aberrant CTCF-binding sites are embedded in H3K9me3-enriched heterochromatic B2 SINE retrotransposons, and CHD4 is required for the repression of B2 SINE transcripts. Together, our results demonstrate the detailed biological functions of CHD4 and reveal a CHD4-modulated mechanism that secures appropriate CTCF recruitment and intact TAD organization in mESCs.

MATERIALS AND METHODS

Mouse ES cell culture
E14Tg2a mESCs were maintained under feeder-free conditions. Briefly, cells were cultured on gelatin-coated cell culture dishes in an mESC culture medium consisting of Glasgow's minimum essential medium (GMEM) containing 10% knock-out serum replacement, 1% non-essential amino acids, 1% sodium pyruvate, 0.1 mM β-mercaptoethanol (all from Gibco, USA), 1% fetal bovine serum, 0.5% antibiotic-antimycotic (both from Hyclone, USA) and 1,000 units/ml LIF (ESG1106; Millipore, USA). mESCs were maintained at 37°C with 5% CO_2 in humidified air.

RNA interference
The siRNAs against *EGFP* and *Chd4* were synthesized and annealed by ST Pharm (Korea). Their sequences are presented in Supplementary Table S1. mESCs were transfected with 50 nM of the indicated siRNA using DharmaFECT I (T-2001-03; Dharmacon, USA) according to the manufacturer's protocol. Briefly, mESCs were seeded to 6-well plates. One day later, 50 nM of siRNAs and DharmaFECT reagent were separately diluted in Opti-MEM (Gibco) and incubated at 25°C for 5 min, and then mixed together. The mixtures were incubated at 25°C for 20 min and added to the mESC cultures. The culture medium was replaced after 24 h. Transfected mESCs were harvested at 48 h after transfection, and knockdown efficiency was analyzed by real-time quantitative polymerase chain reaction (RT-qPCR).

RNA purification and reverse transcription
Total RNAs were purified from mESCs using the TRIzol reagent (Invitrogen, USA) according to the manufacturer's protocol. Briefly, mESCs cultured in 6-well plates were harvested and homogenized with 1 ml of TRIzol reagent. Chloroform (200 μl/sample) was added, and the samples were mixed vigorously by hand for 15 s and incubated at 25°C for 2 min. The mixtures were centrifuged at 12,000 rpm for 15 min at 4°C, and 500 μl of each aqueous phase was transferred to a new Eppendorf tube and mixed with the same volume of isopropanol. The mixtures were incubated at 25°C for 10 min to precipitate total RNAs. The samples were centrifuged at 12,000 rpm for 10 min at 4°C, washed with 75% ethanol, and centrifuged again at 10,000 rpm for 5 min at 4°C. The RNA pellets were dried and dissolved in RNase-free water,

and 1 μg of DNase-treated total RNA was applied for cDNA synthesis using an Improm-II Reverse transcription system (A3802; Promega, USA) according to the manufacturer's protocol. For analysis of retrotransposon expression, the cDNA synthesis step was primed with random hexamers.

Real-time quantitative polymerase chain reaction

The generated cDNAs were amplified using a BioFact Real-time PCR kit (BIOFACT, Korea) according to the manufacturer's manual. The primer sequences used for RT-qPCR are presented in Supplementary Table S2. Briefly, 20-μl reactions containing 1× EvaGreen, 10 mM tetraethylammonium chloride, and 10 pmol of primers were analyzed with a CFX96 system (Bio-Rad, USA) under the following conditions: 95°C for 12 min (initial melting), followed by 40 cycles of 95°C for 20 s (denaturation), 57°C for 30 s (annealing), and 72°C for 30 s (extension). The relative expression levels of Chd family members and retrotransposons were quantified with respect to those of β-actin and the 28S rRNA, respectively.

Generating CHD4-mAID mESCs

We employed the auxin-inducible degron (AID) system to the *Chd4* gene and generated the stable cell line for CHD4-mAID E14Tg2a mESCs as previously described (Natsume et al., 2016; Nora et al., 2017). Briefly, we first generated an OsTIR1 parental cell line (Supplementary Figs. S1G and S1H) by transfecting pEN396-pCAGGS-Tir1-V5-2A-PuroR (*Tigre* donor, #92142; Addgene, USA) and pX330-EN1201 (spCas9nuclease with *Tigre* sgRNA, #92144; Addgene) into E14Tg2a mESCs. After confirming the successful generation of the OsTIR1 parental cell line (Supplementary Fig. S1H), we cloned a *Chd4* donor based on pMK293-mAID-mCherry2-Hygro (#72831; Addgene) and spCas9nuclease with *Chd4* sgRNA based on PX458-pSpCas9(BB)-2A-GFP (#48138; Addgene). We then generated CHD4-mAID E14Tg2a by transfecting these two vectors into OsTIR1 parental cells. Degradation of AID-tagged CHD4 was induced by the addition of 500 μM 3-indoleacetic acid (I2886; Sigma, USA).

CUT&RUN

Cleavage under targets and release using nuclease (CUT&RUN) assays were performed as previously described (Meers et al., 2019; Skene and Henikoff, 2017), with minor modification. Briefly, 4 million mESCs were harvested and washed thrice with 1.5 ml wash buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM spermidine). Cells were bound to activated concanavalin A-coated magnetic beads (at 25°C for 10 min on a nutator), then permeabilized with antibody buffer (wash buffer containing 0.05% digitonin and 4 mM EDTA). The bead-cell slurry was incubated with 3 μl of the appropriate antibody (see below) in a 150 μl volume at 25°C for 2 h on a nutator. After two washes in 1 ml Dig-wash buffer (wash buffer containing 0.05% digitonin), the beads were resuspended in 150 μl pAG/MNase and incubated at 4°C for 1 h on a nutator. After two washes in 1 ml Dig-wash buffer, the beads were gently vortexed with 100 μl Dig-wash buffer. The tubes were chilled to 0°C for 5 min and ice-cold 2.2 mM CaCl₂ was added with gentle vortexing. The tubes were immediately placed on ice and incubated at 4°C for 1

h on a nutator; thereafter, 100 μl 2xSTOP buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.05% digitonin, 0.1 mg/ml RNase A, 50 μg/ml glycogen) was added and the tubes were incubated at 37°C for 30 min on a nutator. The beads were secured using a magnet stand, and the liquid was removed to a fresh tube, combined with 2 μl 10% SDS and 2.5 μl proteinase K (20 mg/ml), and incubated at 50°C for 1 h. DNA was extracted using phenol chloroform as described at https://www.protocols.io/view/cut-amp-run-targeted-in-situ-genome-wide-profiling-zcpf2vn. CUT&RUN libraries were prepared using an Accel-NGS 2S Plus DNA Library Kit (21024; Swift BioSciences, USA), according to the manufacturer's guidelines. The libraries were then sequenced using an Illumina Novaseq 6000 platform (Illumina, USA). The libraries were generated from two sets of biological replicates.

In-situ Hi-C

For Control and Chd4KD cells, *in-situ* Hi-C was performed as previously described (Rao et al., 2014). Briefly: 1.5 million mESCs were crosslinked with 1% formaldehyde; nuclei were isolated; chromatin was digested with MboI (R0147; New England Biolabs [NEB], USA); 5' ends were filled by incorporation of biotinylated-dCTP (19524-016; Life Technologies, USA); proximity ligation, reverse-crosslinking, and DNA shearing were performed; biotinylated junctions were isolated with streptavidin beads (65601; Life Technologies); and DNA libraries were prepared. Each Hi-C library was amplified for six cycles and then subjected to deep sequencing on a Hiseq4000 Illumina platform. For −Aux and +Aux cells, *in situ* Hi-C was performed using the Arima Hi-C kit following the manufacturer's protocols. Briefly, 1.5 million mESCs were crosslinked with 2% formaldehyde, and 150 ng of sheared DNA was used for biotin enrichment. Hi-C libraries were prepared using a Kapa Library Prep Kit with a modification of the protocol provided by Arima. Each Hi-C library was amplified for seven cycles and then subjected to deep sequencing on a Illumina Novaseq 6000 platform. All libraries were generated from two sets of biological replicates.

ChIP-seq

ChIP assays were performed as previously described (Kim et al., 2012), with minor modification. Briefly, mESCs were washed twice with phosphate-buffered saline (PBS) and incubated with 1% formaldehyde for 15 min at 25°C. This crosslinking was quenched with 125 mM glycine at 25°C for 5 min, and the cells were harvested with cold PBS and suspended in SDS lysis buffer (50 mM Tris-HCl, pH 8.0, 1% SDS, 10 mM EDTA). The chromatin was sheared to mono- or dinucleosome sizes using a focused ultrasonicator (S220; Covaris, USA). The sonicates were incubated overnight with the relevant antibodies (see below) and Protein A, G sepharose (17-1279-03 and 17-0618-05; GE Healthcare, USA) at 4°C on a nutator. The immune complexes were washed for 10 min each with the following wash buffers: low-salt wash buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA), high-salt wash buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA), and LiCl wash buffer (10 mM Tris, pH 8.0, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA).

The immune complexes were further washed twice with TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). Finally, the immune complexes were eluted with elution buffer (1% SDS, 0.1 M NaHCO₃) and reverse-crosslinked overnight at 68°C. The immunoprecipitated DNA was treated with proteinase K and RNase A and recovered by phenol-chloroform-isoamyl alcohol precipitation. ChIP-seq libraries were prepared using a NEXTflex ChIP-seq kit (5143-02; Bioo Scientific, USA) according to the manufacturer's guidelines. The libraries were then sequenced using an Illumina HiSeq 2500 platform. The libraries were generated from two sets of biological replicates.

Antibodies
Antibodies against CHD4 (ab70469; Abcam [UK] and in-house generated), CTCF (ab37477; Abcam), α-tubulin (2144S; Cell Signaling Technology, USA), V5 (R960-25; Thermo Fisher Scientific), mCherry2 (M11217; Invitrogen), β-actin (SC-47778; Santa Cruz Biotechnology), and FLAG (F7425; Sigma) were used for immunoblotting. Antibodies against CHD4 (ab70469; Abcam and in-house generated) and IgG (12-371; Millipore) were used for immunoprecipitation. Antibodies against CTCF (07-729; Millipore), RAD21 (ab992; Abcam), H3 (ab1791; Abcam), H3K9me3 (ab8898; Abcam), H3K4me1 (in-house generated), H3K4me3 (in-house generated), H3K27ac (ab4729; Abcam), and H3K27me3 (07-449; Millipore) were used for ChIP-seq. Antibodies against CHD4 (ab70469; Abcam and in-house generated), RAD21 (ab992; Abcam), and IgG (12-371; Millipore and in-house generated) were used for CUT&RUN.

Immunoblotting
E14Tg2a mESCs were transfected with siRNAs against *EGFP* and *Chd4*. The transfected cells were harvested, washed with cold PBS, and lysed with EBC buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 0.5% NP-40, 1 mM PMSF). The lysates were boiled for 5 min with SDS sample buffer, resolved by SDS-PAGE, and subjected to immunoblotting.

ATAC-seq
ATAC-seq libraries were prepared as previously described (Buenrostro et al., 2013; 2015; Choi et al., 2020), with minor modification. Briefly, 50,000 mESCs were harvested, washed with cold PBS, lysed with cold lysis buffer, and immediately centrifuged. The nuclear pellets were resuspended in 25 μl of 2× tagmentation reaction buffer (10 mM Tris, pH 8.0, 5 mM MgCl₂, 10% dimethylformamide), 23 μl of nuclease-free water, and 2 μl of Tn5 transposase (in-house generated), and incubated at 37°C for 30 min. The samples were then immediately purified using a QIAquick PCR purification kit (28106; Qiagen, Germany). The libraries were pre-enriched for five cycles using the KAPA HiFi Hotstart ready mix (KK2601; Kapa Biosystems, USA), and the threshold cycle (Ct) was monitored using qPCR to determine the additional enrichment cycles, which were then applied. The final libraries were purified again with a QIAquick PCR purification kit, and sequenced using an Illumina HiSeq 2500 platform. The libraries were generated from two sets of biological replicates.

Micrococcal nuclease digestion and MNase-seq
Cells cultured in 60Φ culture dishes were washed with PBS and incubated with modified PBS (150 mM NaCl, 0.02% Tween-20, 0.02% Triton X-100) at 25°C for 5 min. The cells were then fixed with 1% formaldehyde at 25°C for 10 min, quenched with 125 mM glycine at 25°C for 5 min, washed, and suspended with lysis buffer (18% Ficoll-400, 10 mM KH₂PO₄, 10 mM K₂HPO₄, 1 mM MgCl₂, 250 nM EGTA). The lysates were centrifuged at 12,000 rpm for 40 min at 4°C, and the pelleted chromatin was resuspended in A buffer (10 mM Tris-Cl, pH 7.4, 150 mM NaCl, 5 mM KCl, 1 mM EDTA). CaCl₂ (5 mM) was added to the samples, which were immediately digested with MNase (M0247S; NEB) at 37°C for 60 min. The reaction was inactivated with 1% SDS and 500 mM EDTA, and reverse-crosslinking was performed overnight at 68°C. Proteinase K and RNase were added sequentially to the samples, and DNAs were recovered using phenol-chloroform-isoamyl alcohol extraction. The MNase-digested DNAs were separated on a 1.5% agarose gel, and mono-nucleosomal DNAs were extracted using a QIAEX II gel extraction kit (20021; Qiagen) according to the manufacturer's manual. The purified mono-nucleosomal DNAs were subjected to sequencing using a TruSeq DNA library prep kit (FC-121-2001; Illumina). The final libraries were sequenced using an Illumina HiSeq 2500 platform. The libraries were generated from two sets of biological replicates.

Total RNA-seq and mRNA-seq
DNase-treated total RNAs were purified from mESCs and used for total RNA-seq. Total RNA-seq libraries were prepared using a TruSeq stranded RNA kit (RS-122-2301; Illumina) according to the manufacturer's manual. For mRNA-seq library preparation, mRNAs were isolated from total RNA using a Magnetic mRNA isolation kit (S1550S; NEB), and libraries were prepared using a NEXTflex Rapid directional RNA-seq kit (5138-08; Bioo Scientific). The libraries were sequenced using an Illumina HiSeq 2500 platform. The libraries were generated from two sets of biological replicates.

Immunoprecipitation
E14Tg2a mESCs were harvested, washed with cold PBS, and lysed with EBC buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5% NP-40, 1 mM PMSF). The lysates were incubated for 1 h with the indicated antibodies and Protein A, G sepharose (17-1279-03 and 17-0618-05; GE Healthcare) at 4°C with agitation. The immune complexes were washed three times with EBC buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5% NP-40, 1 mM PMSF). The lysates were boiled for 5 min with SDS sample buffer, resolved by SDS-PAGE, and subjected to immunoblotting.

RNA immunoprecipitation assays
RNA immunoprecipitation (RIP) assays were conducted as previously described (Heo and Sung, 2011; Kim et al., 2014), with minor modification. Briefly, cultured mESCs were crosslinked with 1% formaldehyde for 15 min at 25°C and quenched with 125 mM glycine. The cells were harvested, washed with cold PBS, and then lysed with RIPA buffer (50 mM Tris-Cl, pH 8.0, 150 mM NaCl, 0.05% SDS, 1 mM EDTA,

1% NP-40, and 0.05% sodium deoxycholate) containing an RNase inhibitor (M007L; Enzynomics, Korea). The samples were sonicated, treated with DNase I (18068-015; Invitrogen) for 10 min at 37°C, and cleared by centrifugation. The extracts were bound with anti-Chd4 (ab70469; Abcam) for 2 h at 4°C in the presence of RNase inhibitor. The immune complexes were further incubated with Dynabeads Protein A and G (10001D, 10004D; Invitrogen) for 1 h at 4°C, and then washed for 10 min each with the following wash buffers: low-salt wash buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA), high-salt wash buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA), and LiCl wash buffer (10 mM Tris, pH 8.0, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA). The immune complexes were further washed twice with TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), eluted, and reverse-crosslinked with 400 mM NaCl for 2 h at 65°C. RNAs were purified from the eluates using an RNA Clean & Concentrator™-5 kit (R1015; Zymo Research, USA) according to the manufacturer's manual, and then incubated again with DNase I (18068-015; Invitrogen) for 30 min at 37°C. Finally, the DNase I was inactivated and removed using a DNA-free™ kit (AM1906; Invitrogen) according to the manufacturer's manual.

In vitro transcription and biotinylation of RNAs
The plasmids used for *in vitro* transcription were constructed as previously described (Espinoza et al., 2004), with minor modification. The primer sequences used for *in vitro* transcription are presented in Supplementary Table S3. Briefly, we used the pUC vector to construct pUC-T7-yGAL1, pUC-T7-EGFP1/2, and pUC-T7-Nanog. The indicated RNAs were transcribed *in vitro* using a TranscriptAid T7 High Yield Transcription kit (K0441; Thermo Fisher Scientific) and then purified using an RNA Clean & Concentrator™-5 kit according to the manufacturer's manual. The purified RNAs were resolved on 5% denaturing polyacrylamide mini gels at 70 V and 4°C for 2 h. The gels were cut for size selection and appropriately sized fragments were eluted with elution buffer (10 mM Tris-HCl, pH 7.5, 300 mM NaCl, 0.1% SDS, 1 mM EDTA) overnight at 4°C with agitation. The eluted RNAs were purified using an RNA Clean & Concentrator™-5 kit. The purified RNAs of this step were used for the nucleosome-sliding assay. For other experiments, the purified RNAs were labeled with biotin using a Pierce RNA 3' End Biotinylation kit (20160; Thermo Fisher Scientific) according to the manufacturer's manual. The biotinylated RNAs were purified using an RNA Clean & Concentrator™-5 kit and then resolved on a 5% denaturing polyacrylamide mini gel at 70 V and 4°C for 2 h. The gels were cut for size selection and eluted with elution buffer (10 mM Tris-HCl, pH 7.5, 300 mM NaCl, 0.1% SDS, 1 mM EDTA) overnight at 4°C with agitation. The eluted biotinylated RNAs were purified once more using an RNA Clean & Concentrator™-5 kit, and then subjected to electrophoretic mobility shift assay (EMSA).

Electrophoretic mobility shift assay
EMSA was carried out as previously described (Kim et al., 2014), with minor modifications. Briefly, the indicated

amount of purified Chd4 protein was incubated with 5 nM of biotinylated RNAs in binding buffer (50 mM Tris-Cl, pH 7.5, 100 mM NaCl, 10 mM β-mercaptoethanol, 5% glycerol) for 40 min at 25°C. Bound protein-RNA complexes were resolved on 5% native polyacrylamide mini gels at 70V and 4°C for 2 h. The complexes were transferred to an Amersham Hybond-N+ nylon membrane (RPN303B; GE Healthcare), UV-crosslinked, and visualized using a Chemiluminescent Nucleic Acid Detection Module kit (89880; Thermo Fisher Scientific) according to the manufacturer's manual.

Nucleosome reconstitution and nucleosome-sliding assay
Nucleosome reconstitution was performed as previously described (Owen-Hughes et al., 1999). Briefly, two types (lateral and central forms) of a 147-bp DNA fragment were amplified from pGEM-3z/601, and *X. laevis* core histones were purified using bacterial expression systems and nickel-affinity chromatography. The core histones and DNA fragments were mixed at a 1:1 ratio in initial buffer (10 mM HEPES, pH 7.9, 1 mM EDTA, 5 mM DTT, 0.5 mM PMSF), brought to 2 M NaCl with 2 μg bovine serum albumin (BSA) in each 10 μl reaction, and incubated for 15 min at 37°C. The reaction was serially diluted with 3.6 μl, 6.7 μl, 5 μl, 3.6 μl, 4.7 μl, 6.7 μl, 10 μl, 30 μl, and 20 μl of initial buffer, and each dilution was incubated for 15 min at 30°C. Lastly, the reaction was diluted with 100 μl of final buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 0.1% NP-40, 5 mM DTT, 0.5 mM PMSF, 20% glycerol, 100 μg/ml BSA) and incubated for 15 min at 30°C. The lateral and central nucleosome forms were reconstituted and used for the nucleosome-sliding assay.

For the nucleosome-sliding assay, reconstituted nucleosomes (75 ng, 0.36 pmol) were incubated with purified Chd4 (50 ng, 0.23 pmol), varying concentrations of purified RNAs (0 pmol, 0.1 pmol, 0.2 pmol, 0.5 pmol, and 1 pmol), and ATP (0.1 pmol) in sliding buffer (50 mM KCL, 20 mM HEPES, pH 7.9, 2 mM DTT, 0.5 mM PMSF, 0.05% NP40, 10% glycerol, 100 μg/ml BSA, 10 mM MgCl$_2$) for 45 min at 30°C. The reaction was stopped by the addition of 500 ng of plasmid (size > 10 kb) and 0.1 pmol of ATP-γ-S tetralithium salt (10102342001-Roche; Sigma) followed by incubation for 30 min at 30°C. The samples were resolved on 5% native polyacrylamide mini gels at 65 V and 4°C for 2.5 h.

GRO-seq
Circularized GRO-seq was performed as previously described (Danko et al., 2013; Lam et al., 2013). Briefly, around 10 million nuclei per sample were purified and used for global run-on and base hydrolysis, which were carried out as previously described (Core et al., 2008). BrU-labeled nascent RNAs were immunoprecipitated twice with anti-BrdU antibody-conjugated agarose (SC-32323ac; Santa Cruz Biotechnology). Between the two immunoprecipitations, the BrU-precipitated RNAs were subjected to polyA tailing with poly(A)-polymerase (P7460L; Enzymatics, USA). The RNAs were subjected to first-strand cDNA synthesis using Superscript III Reverse Transcriptase (18080-044; Invitrogen) and the oNTI223 RT primer (IDT). Excess RT primers were removed with exonuclease I (M0293S; NEB), and the cDNAs were size-selected (120-220 nt) in a 6% polyacrylamide TBE-urea gel. The selected

cDNAs were circularized using CircLigase (CL4111K; Epicentre, USA) and re-linearized with ApeI (M0282S; NEB). The single-stranded DNA templates were amplified using Phusion High-Fidelity DNA Polymerase (M0530S; NEB) and Illumina TruSeq small-RNA sample barcoded primers. The generated PCR products were isolated and size-selected (190-290 bp) by electrophoresis on a 6% native polyacrylamide TBE gel. The final libraries were sequenced on an Illumina HiSeq 2500 platform. The libraries were generated from two sets of biological replicates.

Data processing and analysis

For Hi-C analysis, the *in situ* Hi-C data were analyzed using HiC-Pro (Servant et al., 2015). This included alignment to the mouse genome (mm10), extraction of valid Hi-C interactions, merging of replicates, building of the Hi-C matrix, and performance of iterative correction and eigenvector decomposition (ICE) normalization. Using HiC-Pro, we generated ICE-normalized Hi-C matrixes for 20-kb and 40-kb resolutions and annotation files that indicated the genomic bins. Chromatin loops (or peaks in APA analysis) were identified using the HiCCUPS algorithm with default options in the Juicer tool (Durand et al., 2016). To identify TAD positions, we applied reciprocal insulation (RI) scores, which were calculated from the ICE-normalized 20-kb resolution Hi-C matrix using CaTCH (Zhan et al., 2017). Next, we determined the RI score that exhibited maximal CTCF enrichment at the RI score-based TAD borders. In this study, we used TAD positions with an RI score of 0.635 in Control/Chd4KD cells and an RI score of 0.750 in −Aux/+Aux cells. For most of the Hi-C related analyses, including relative contact probability (RCP), aggregate TAD analysis (ATA), and aggregate peak analysis (APA), we used GENOVA (van der Weide et al., 2021).

For the MNase-seq, ATAC-seq, ChIP-seq, CUT&RUN, and GRO-seq analyses, the raw reads were aligned to the mouse genome (mm10) using Bowtie2 (ver. 2.2.9) with default parameters (Langmead and Salzberg, 2012). For the total RNA-seq and mRNA-seq analyses, the raw reads were aligned to the mouse genome (mm10) using STAR (ver. 2.5.2a) with default parameters (Dobin et al., 2013).

Generally, we used MACS2 (Zhang et al., 2008) to convert the aligned BAM files into bedGraph files, and normalized the data with respect to the total read counts. Then, we used bedGraphToBigWig (Kent et al., 2010) to convert the bedGraph files into bigWig files. The bigWig files were used as input files for bwtool (Pohl and Beato, 2014) (matrix and aggregate option) to quantify the intensity (e.g., using heatmaps or average line plots) of the relevant sequencing data. Further analyses were performed using in-house scripts, MACS2 (Zhang et al., 2008), and HOMER (Heinz et al., 2010). All of our raw data (fastq files) were confirmed to be of good quality using FastQC (http://www.bioinformatics. babraham.ac.uk/projects/fastqc/).

For our CUT&RUN and ChIP-seq analyses, we used MACS2 (callpeak option, $P < 0.005$) to identify the peaks (or binding sites) of proteins of interest by using input (for ChIP-seq) or IgG (for CUT&RUN) as the background. The CUT&RUN and ChIP-seq data were also subjected to motif analysis, which was performed using the HOMER findMotifsGenome.pl op-

tion. Analyses of the genomic contents within CTCF-binding sites were performed using the HOMER annotatPeaks.pl option (Heinz et al., 2010).

For our mRNA-seq and total RNA-seq analyses, we used Cufflinks (Cuffdiff option, fr-firststrand) to assess the expression levels. For our analysis of SINEs, we built a STAR-index exclusively containing the previously identified 1.5 million SINE copies (RepeatMasker in UCSC table browser), and performed STAR and Cufflinks analyses as described above.

For our GRO-seq analysis, we used HOMER (findPeaks option, -style groseq) to examine the number of *de novo*-produced SINE transcripts (Heinz et al., 2010).

Box plots, volcano plots, and other plots were drawn with R (ggplot2) (Wickham, 2009) and heatmaps were drawn with Java TreeView (Saldanha, 2004). The examples of our genome-wide data were visualized using the Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

Public data acquisition

The publicly released ChIP-seq data were downloaded from NCBI GEO DataSets and the ENCODE Consortium. These data were downloaded as sra or fastq formats, and the sra files were converted to fastq files using the SRA Toolkit (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software). Thus, we analyzed the public datasets and our own results using the same methods. Details of the analyzed data are summarized in Supplementary Table S4.

RESULTS

CHD4 and CTCF are closely linked in mESCs

We first determined the expression of *Chd* family members using RT-qPCR. The expression levels for the majority of *Chd* family members were higher in mESCs than in lineage-specified cells, such as mouse embryonic fibroblasts (MEFs) or NIH3T3 cells (Supplementary Fig. S1A). Among the *Chd* family members, *Chd4* was the most abundant in mESCs (Supplementary Fig. S1A), indicating that it could play an important role in these cells. Therefore, we focused on the biological function of CHD4 in mESCs.

To elucidate the function of CHD4, we first identified CHD4-binding sites by performing CUT&RUN (Skene and Henikoff, 2017) and also by analyzing the public chromatin immunoprecipitation sequencing (ChIP-seq) data against CHD4 (GSE64825, GSE61188, and GSE27844, see also Supplementary Table S4). Then, we analyzed genomic features within the CHD4-binding sites (CUT&RUN and ChIP-seq peaks). Notably, we found that CTCF motifs were highly enriched within CHD4-binding sites relative to CHD1-binding sites and randomly selected sites (Supplementary Fig. S1B, Supplementary Table S5). Interestingly, we also observed that CHD4 localized to the borders of TADs (Fig. 1A, Supplementary Fig. S1C) and the anchors of insulated neighborhoods (Fig. 1B, Supplementary Fig. S1D), where CTCF is known to be localized (Figs. 1A and 1B) and play a boundary function (Dixon et al., 2012; Dowen et al., 2014; Luppino et al., 2020; Nora et al., 2012; Rao et al., 2014; Szabo et al., 2020). Furthermore, 34% of the CTCF ChIP-seq peaks coincided with the CHD4 CUT&RUN peaks (Fig. 1C), suggesting that
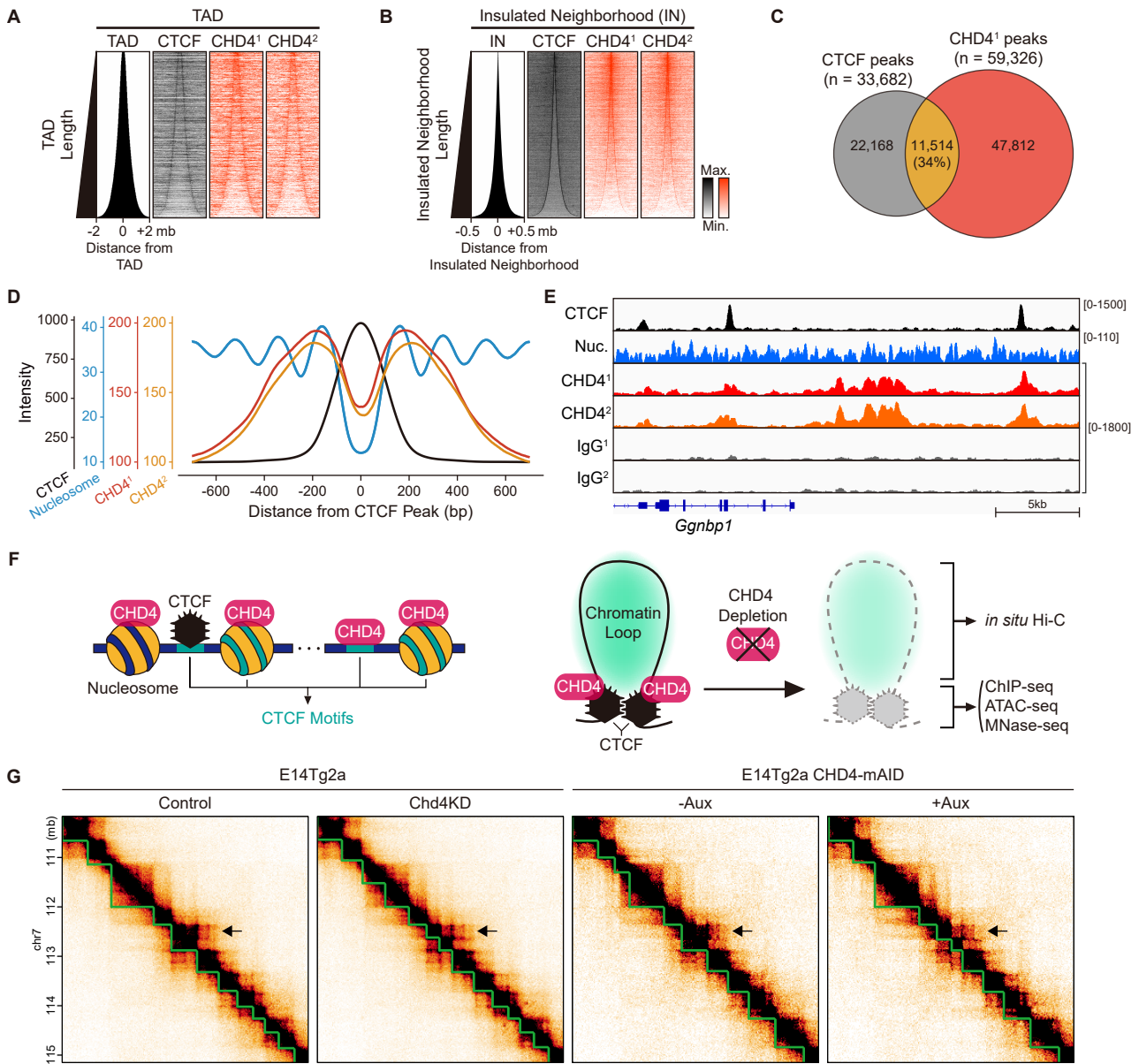
Fig. 1. CHD4 localizes near CTCF and marginally organizes the global 3D chromatin organizations. (A and B) Heatmaps of CTCF and CHD4 aligned at 2,153 TADs (Dixon et al., 2012) (A) and 23,726 insulated neighborhoods (Dowen et al., 2014) (B). All heatmaps were sorted in ascending order by the length of the TAD or insulated neighborhood. CHD4[1] (in-house generated antibody) and CHD4[2] (Abcam antibody) denote CUT&RUN data obtained using different antibodies against CHD4. See also Supplementary Figs. S1C and S1D. (C) Venn diagrams representing the overlap of CTCF and CHD4[1] peaks (binding sites). (D) Line plots showing average enrichments of CTCF, nucleosomes, and CHD4 at CTCF peaks. (E) Examples representing the position/enrichment of CTCF, nucleosomes (Nuc.), CHD4[1,2], and IgG[1,2] (which are relevant to CHD4[1,2]). (F) Model in which CTCF motifs are enriched within CHD4-binding sites, and CTCF is localized at nucleosome-free regions flanking nucleosomes and CHD4 (left). Model showing that both CHD4 and CTCF are localized at the boundary regions of chromatin loops (right). After CHD4 depletion, we conducted experiments to detect changes in 3D chromatin organization (*in situ* Hi-C), CTCF enrichment (ChIP-seq), and chromatin accessibility (ATAC-seq and MNase-seq). (G) Examples of Hi-C interactions (heatmaps) and TADs (green lines). Black arrows indicate the region where 3D chromatin organizations are severely disrupted upon CHD4 depletion. (H and I) Heatmaps of CTCF and CHD4[1,2] aligned at 5,452 TADs in Control cells (H) and at 4,525 TADs in auxin-untreated (-Aux) cells (I). All heatmaps were sorted in ascending order by the lengths of the TADs. (J) Examples of differential Hi-C interactions and CHD4 localization at chromosome 17. Distribution of PC1 (first principal component, equivalent to the first eigenvector) values across the chromosomes and A (red)/B (green) compartments are shown (bottom). Black arrow and dotted box indicate the CHD4-enriched regions where 3D chromatin organizations are disrupted upon CHD4 depletion. (K) *cis* PC1 values in 40-kb genomic bins are ranked into 50 quantiles (A and B compartments), and pairwise Hi-C enrichment (compartment strength) was calculated between each of the 50 quantiles. The difference in compartment strength upon CHD4 depletion was calculated (right).
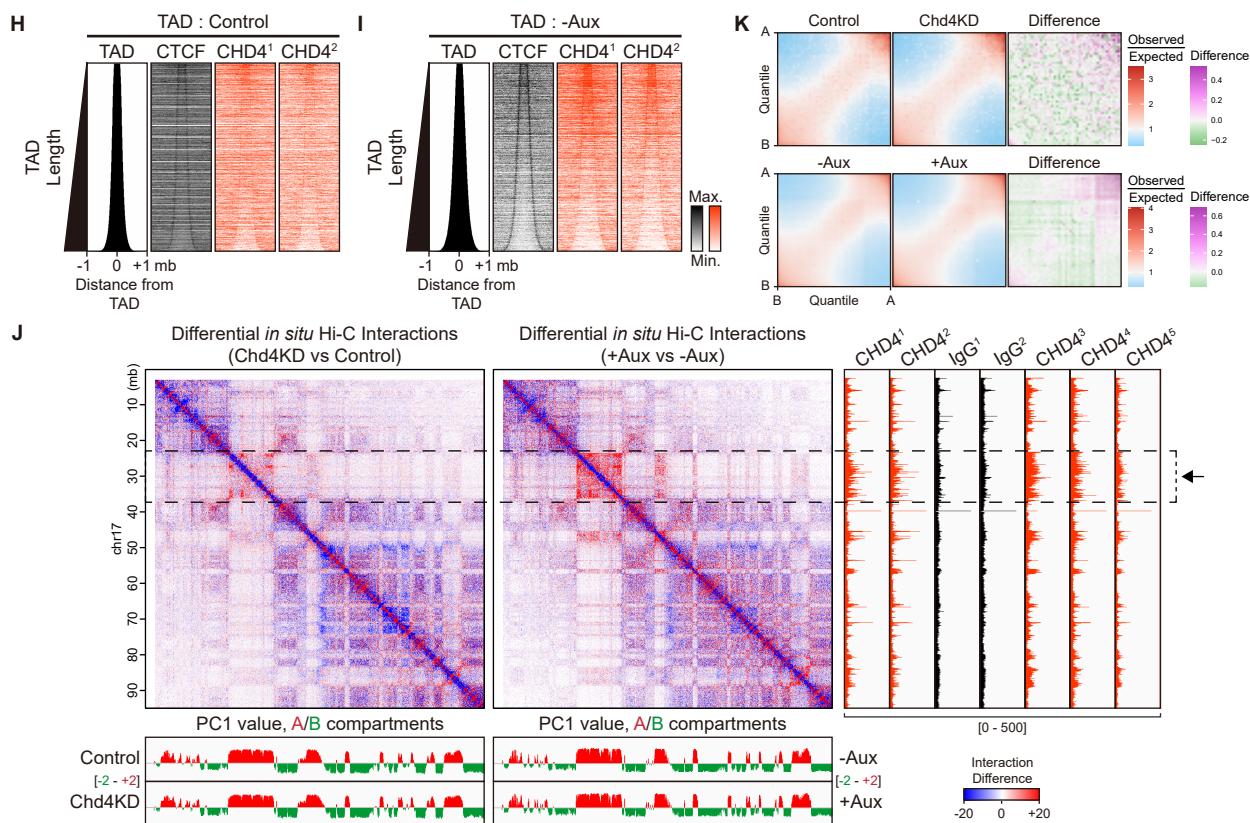
Fig. 1. Continued.

CHD4 and CTCF resided near one another. To evaluate this possibility, we analyzed the localization of CHD4 near the CTCF ChIP-seq peaks. As expected, we found that CHD4 resided near CTCF (Figs. 1D and 1E). Intriguingly, although the CHD4-binding sites were enriched for CTCF motifs (Supplementary Fig. S1B, Supplementary Table S5), we observed that CHD4 was also present at locations lacking CTCF (Figs. 1C and 1E), suggesting that CHD4 may conceal CTCF motifs (Fig. 1F). We also observed that CTCF localized to exposed DNA sequences surrounded by well-positioned nucleosomes (Figs. 1D and 1E), consistent with a previous report (Cuddapah et al., 2009). The presence of CTCF at nucleosome-free regions flanking nucleosomes and CHD4 (Fig. 1D) may suggest that CHD4 could regulate nucleosome occupancy near CTCF-binding sites, and thereby control interactions between CTCF and DNA sequences harboring CTCF motifs. Taken together, these results strongly indicate that CHD4 and CTCF are closely linked in mESCs. Thus, we hypothesized that CHD4 might control CTCF-mediated 3D chromatin organization by regulating chromatin accessibility near CTCF-binding sites (Fig. 1F).

CHD4 depletion marginally affects genome-wide 3D chromatin organization

To test our hypothesis, we first designed and applied siRNAs against *EGFP* (Control) and *Chd4* (Chd4KD). We confirmed the successful knockdown of *Chd4* in mESCs using RT-qPCR (Supplementary Fig. S1E) and immunoblotting (Supplemen-

tary Fig. S1F). Furthermore, we employed an AID system (Natsume et al., 2016) to degrade CHD4 in mESCs (CHD4-mAID) (Supplementary Figs. S1G-S1L). We confirmed that CHD4 was efficiently degraded after 24 h of auxin treatment, as assessed by immunoblotting (Supplementary Fig. S1J).

To investigate whether CHD4 depletion disturbed the 3D chromatin organization, we performed *in situ* Hi-C upon *Chd4* knockdown (Control and Chd4KD cells) and CHD4 depletion (auxin-untreated and -treated cells, termed −Aux and +Aux cells, respectively). We generated a total of ~2.1 billion valid Hi-C interactions (Supplementary Table S6) and comprehensively analyzed the *in situ* Hi-C data using HiC-Pro (Servant et al., 2015) and GENOVA (van der Weide et al., 2021). The Hi-C experiments were performed in two biological replicates and showed high reproducibility according to the Pearson correlation coefficient (Supplementary Fig. S1M). Furthermore, we applied RI scores to identify TADs (Fig. 1G) by using CaTCH (Zhan et al., 2017). Consistent with the above-described results (Figs. 1A and 1B), we observed that CHD4 and CTCF were both localized at the border of TADs in Control cells (Fig. 1H) and −Aux cells (Fig. 1I). Notably, we found changes in TAD positions (Fig. 1G, arrow) and Hi-C interactions at the CHD4-enriched regions (Fig. 1J, dotted box with arrow) upon CHD4 loss. However, we did not detect any major change in the RCP (Supplementary Fig. S1N) or compartment domains (Fig. 1J, Supplementary Fig. S1O). Indeed, only ~2.3%-2.5% of compartment domains (in 100-kb bins) were switched (from A to B or B to A) upon CHD4

depletion (Supplementary Fig. S1P). Interestingly, we detected a slight increase in the AA compartmentalization strength upon CHD4 loss (Fig. 1K). Our observation that there was relatively little change in compartmentalization despite changes in Hi-C interactions is consistent with the findings of previous studies (Nora et al., 2017; Rao et al., 2017). Based on these findings, we conclude that CHD4 depletion marginally affects global 3D chromatin organizations.

CHD4 maintains the local 3D chromatin organization by preventing aberrant CTCF recruitment at TAD interiors

To investigate the role of CHD4 in local 3D chromatin organization, we performed ATA upon CHD4 depletion at four TADs obtained from each cell (Control, Chd4KD, −Aux, and +Aux cells). Notably, we detected genome-wide decreases of Hi-C interactions within the TADs of Control and −Aux cells upon CHD4 depletion (Figs. 2A and 2B, arrow). In contrast, Hi-C interactions within the TADs of Chd4KD and +Aux cells were largely unaltered by CHD4 loss (Figs. 2A and 2B). These results suggest that CHD4 maintains the local interactions within the original TADs of wild-type mESCs.

To address this further, we identified four types of differential TADs seen upon CHD4 depletion: We compared the changes in Hi-C interactions to identify weakened or strengthened TADs, and compared the changes in TAD positions to identify separating and merging TADs (Figs. 2C-2G). Weakened TADs were defined as original TADs of wild-type mESCs (Control or −Aux cells) whose Hi-C interactions were decreased upon CHD4 loss. Strengthened TADs were defined as TADs of CHD4-depleted cells (Chd4KD or +Aux cells) whose Hi-C interactions were increased upon CHD4 loss. Separating TADs were defined as original TADs of wild-type mESCs (Control or −Aux cells) that were split into two or more smaller TADs upon CHD4 loss. Finally, merging TADs were defined as TADs of CHD4-depleted cells (Chd4KD or +Aux cells) that contained two or more original TADs of wild-type mESCs (Control or −Aux cells). The ATA results indicated that Hi-C interactions showed genome-wide decreases at weakened TADs (Fig. 2D) and separating TADs (Fig. 2E), but genome-wide increases at the strengthened TADs (Fig. 2F) and merging TADs (Fig. 2G) upon CHD4 loss. The decreasing TADs (weakened and separating TADs) shared more TADs than the increasing TADs (strengthened and merging TADs) (Supplementary Fig. S2A), indicating that weakened and separating TADs were more closely associated. Interestingly, comparing the numbers of the four types of differential TAD seen upon CHD4 loss, revealed that decreasing TADs occurred more frequently than increasing TADs (Fig. 2H). This was consistent with our above-described observation that Hi-C interactions showed genome-wide decreases within the original TADs of wild-type mESCs (Figs. 2A and 2B, arrow). Thus, we focused on decreasing TADs (weakened and separating TADs).

To elucidate how CHD4 maintains local interactions within TADs, we performed CTCF ChIP-seq upon CHD4 loss. By comparing the CTCF intensity, we defined CTCF signals that were gained or lost upon CHD4 depletion (Fig. 2I). Next, we divided the decreasing TADs (weakened and separating TADs) into borders and interiors (Fig. 2J, top right). We observed that both weakened and separating TADs exhibited similar

distribution patterns of gained/lost CTCF at the borders/interiors of TADs (Fig. 2J), suggesting that these two types of decreasing TAD may be fundamentally the same and/or formed similarly upon CHD4 loss. We did not detect any significant difference in the distribution of gained and lost CTCF at TAD borders, whereas gained CTCF was significantly enriched at TAD interiors compared to lost CTCF (Fig. 2J). These results indicate that gained CTCF may be responsible for decreasing interactions within weakened and separating TADs upon CHD4 loss.

To further support this notion, we comprehensively analyzed Hi-C interactions, TAD positions, insulations scores, and CTCF ChIP-seq at specific regions of the genome upon CHD4 loss. Consistent with our genome-wide ATA results (Figs. 2D and 2E), we found a notable decrease in Hi-C interactions within the weakened and separating TADs (Fig. 2K, Supplementary Figs. S2B and S2C, black arrow). Importantly, we observed that the presence of a gained CTCF (marked with *) greatly reduced the local insulation, thereby generating a novel TAD border (separating TAD, Fig. 2K, Supplementary Fig. S2B) and/or markedly decreasing the Hi-C interactions within TADs (weakened TAD, Supplementary Figs. S2B and S2C). In general, we observed the same types of decreasing TAD in the two cell models CHD4 loss (Chd4KD and +Aux cells); at some genomic loci, however, we observed a weakened TAD in Chd4KD cells but a separating TAD in +Aux cells (Supplementary Fig. S2B). This observation supports the idea that these two types of decreasing TAD are closely associated, and that the gained CTCF causes both types of decreasing TAD upon CHD4 depletion. When considering the splitting feature of separating TADs (Fig. 2C), one can speculate the reduced TAD sizes and an elevated number of chromatin loops (peaks) coupled with gained CTCF upon CHD4 loss. As expected, we observed reduced TAD sizes (Supplementary Fig. S2D) and an elevated number of total chromatin loops (peaks) and cell-type-specific peaks (Supplementary Figs. S2E and S2F) upon CHD4 depletion (Supplementary Figs. S2G and S2H). Importantly, the insulation scores at the gained CTCF sites were globally reduced upon CHD4 depletion (Fig. 2L), indicating that the CHD4 depletion-triggered gain of CTCF is directly responsible for generating TAD borders in a genome-wide manner. Furthermore, our APA results showed that Hi-C interactions were increased at the gained CTCF-associated peaks upon CHD4 loss (Supplementary Fig. S2I), supporting the direct relationship between gained CTCF and changes in Hi-C interactions. Collectively, our results strongly imply that the decreasing TADs (weakened and separating TADs) are directly associated with gained CTCF upon CHD4 loss. Therefore, these findings suggest that CHD4 regulates the local 3D chromatin organization by preventing the aberrant recruitment of CTCF to TAD interiors.

CHD4 conceals aberrant CTCF-binding sites by regulating chromatin accessibility at heterochromatic regions

To investigate how CHD4 prevents aberrant CTCF binding, we analyzed the genome-wide changes in CTCF upon CHD4 depletion at CHD4 CUT&RUN peaks (or CHD4-binding sites) (Figs. 3A and 3B, Supplementary Figs. S3A and S3B). We classified the CHD4 CUT&RUN peaks into those that coincid-
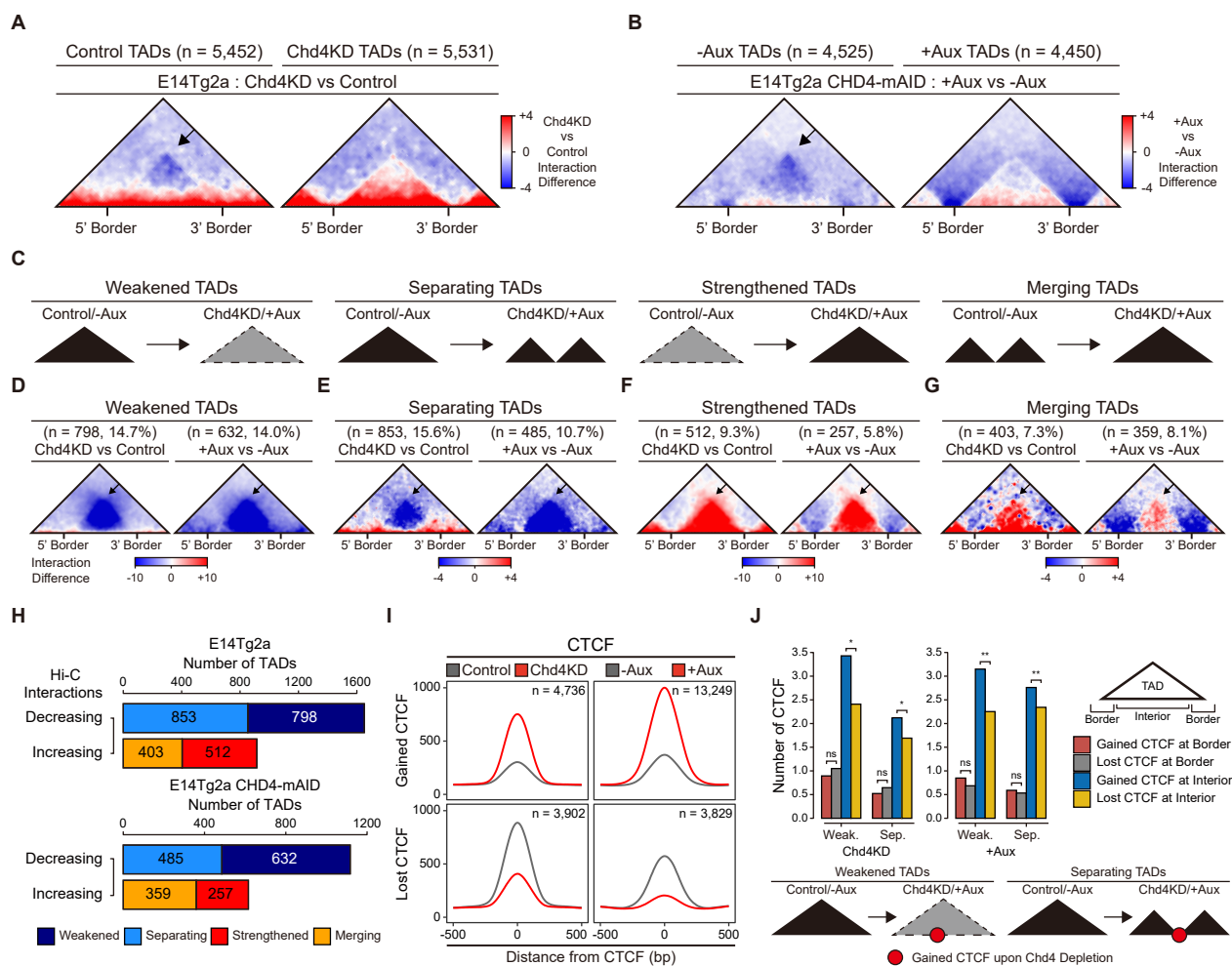
Fig. 2. CHD4 maintains the local 3D chromatin organizations by ensuring proper CTCF recruitment. (A and B) Aggregate TAD analysis (ATA) comparing Hi-C interactions of Chd4KD and Control cells at 5,452 Control TADs (left) and 5,531 Chd4KD TADs (right) (A) and in +Aux and -Aux cells at 4,525 -Aux TADs (left) and 4,450 +Aux TADs (right) (B). (C) Schematic diagrams showing the four types of modified TADs observed upon CHD4 depletion. (D-G) ATA results at weakened (D), separating (E), strengthened (F), and merging (G) TADs. (H) Stacked bar graphs showing the number of decreasing and increasing TADs upon *Chd4* knockdown (top) and CHD4 depletion (bottom). (I) Line plots showing average enrichments of gained CTCF (top) and lost CTCF (bottom) upon CHD4 depletion. (J) Bar graphs showing the average number of gained and lost CTCF at the TAD border and interior. The average number of CTCF was normalized with respect to the total number of relevant CTCF types. P values were derived using the Wilcoxon signed rank test ($*P < 1 \times 10^{-10}$; $**P < 1 \times 10^{-50}$; ns, not significant). (K) Example of separating TADs upon CHD4 depletion. Hi-C interactions (heatmaps), TADs (green lines and black triangles), insulation heatmaps, CTCF ChIP-seq peaks (red and blue triangles indicate + and - directionalities of CTCF motifs), and genes are shown for Control, Chd4KD, -Aux, and +Aux cells. Black arrows indicate the region where 3D chromatin organizations are severely disrupted upon CHD4 depletion. The zoom-in of CTCF ChIP-seq data from gained CTCFs (1*) residing at newly-established borders of separating TADs (dotted boxes) upon CHD4 depletion. See also Supplementary Figs. S2B and S2C. (L) Line plots showing average insulation scores obtained at gained CTCF upon CHD4 depletion.

ed with CTCF ChIP-seq peaks of wild-type mESCs (Control or −Aux cells) and those that did not coincide with these peaks. We did not detect any major change in CTCF upon CHD4 depletion at CHD4 peaks that coincided with CTCF peaks (Figs. 3A and 3B, Supplementary Figs. S3A and S3B), but observed aberrant gain of CTCF upon CHD4 depletion at CHD4 peaks that did not coincide with CTCF peaks (Figs. 3A and 3B, Supplementary Figs. S3A and S3B, arrow). Since CHD4 CUT&RUN peaks are highly enriched for CTCF motifs (Sup-

plementary Fig. S1B, Supplementary Table S5), these results indicate that CHD4 conceals putative CTCF motifs (Fig. 1F, left). The observed changes in CTCF binding were not derived from any change in *CTCF* expression (Supplementary Figs. S3C and S3D).

To elucidate how CHD4 conceals putative CTCF motifs, we comprehensively analyzed CTCF ChIP-seq data obtained from Control and Chd4KD cells. We first determined the CTCF-binding sites (or ChIP-seq peaks) in Control and Ch-
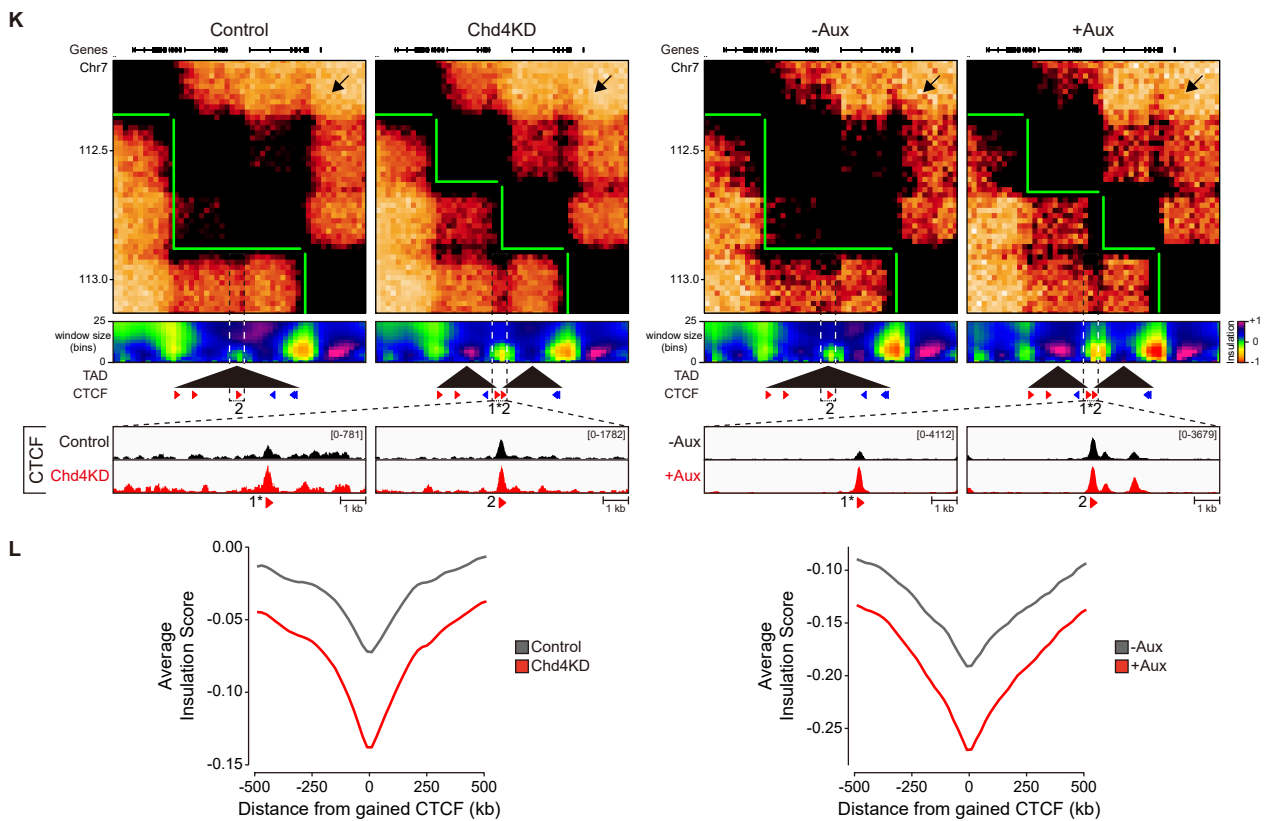
Fig. 2. Continued.

d4KD cells. We then merged the CTCF-binding sites (n = 39,800) and sorted them in descending order of their relative CTCF intensity (Chd4KD/Control) (Figs. 3C and 3F, Supplementary Figs. 3E, S3H-S3J). According to this relative CTCF intensity, we defined two groups of CTCF-binding sites: Group 1 (n = 4,736, 11.9%) represents CTCF-binding sites at which more CTCF was recruited in Chd4KD cells compared to Control cells, while Group 2 (n = 35,064, 88.1%) represents the remainder of the CTCF-binding sites, at which CTCF binding was unchanged or reduced in Chd4KD cells. Consistent with our above-described results (Figs. 2I, 2K, 3A, and 3B, Supplementary Figs. S2B, S2C, S2E, S3A, and S3B), we observed that CHD4 depletion-triggered gain and loss of CTCF binding occurred at the Group 1 and Group 2 CTCF-binding sites, respectively (Fig. 3C). Furthermore, we found that RAD21 was aberrantly gained along with CTCF at the Group 1 sites upon CHD4 depletion (Supplementary Fig. S3E, left). This supports our observation that the CTCF that is aberrantly gained upon CHD4 depletion is directly involved in forming new chromatin loops (Supplementary Fig. S2I), and thereby disrupting local 3D chromatin organizations.

Previously, we observed that CTCF binds to exposed DNA sequences surrounded by well-positioned nucleosomes (Fig. 1D) (Cuddapah et al., 2009), and herein found that CHD4 also localizes near these sites (Fig. 1D). In addition, we found that CTCF was aberrantly gained upon CHD4 depletion at the CHD4 CUT&RUN peaks (Figs. 3A and 3B, Supplementary Figs. S3A and S3B), which harbor CTCF motifs. Considering

these results, we hypothesized that CHD4 conceals putative CTCF motifs by regulating chromatin accessibility at specific DNA sequences with embedded CTCF motifs (Fig. 1F, left). To test this hypothesis, we analyzed MNase-seq (chromatin digestion with micrococcal nuclease combined with sequencing), DNase-seq (DNase I hypersensitive site sequencing), and ATAC-seq (assay for transposase-accessible chromatin using sequencing) data and aligned them with the same order as previously defined relative CTCF intensity (Group 1 and 2 sites). In Control cells, CTCF mainly localized to Group 2 sites (Fig. 3C), which were characterized by an accessible chromatin region with depleted nucleosomes, enriched DNase I hypersensitive sites, and enriched ATAC-seq signals (Fig. 3C). In accordance with this, Control cells exhibited far less binding of CTCF at Group 1 sites (Fig. 3C), which were characterized by an inaccessible chromatin region with enriched nucleosomes, depleted DNase I hypersensitive sites, and depleted ATAC-seq signals (Fig. 3C). Importantly, we detected a significant decrease in nucleosomes and a significant increase in ATAC-seq signals upon CHD4 depletion at Group 1 sites (Figs. 3C-3E), indicating that CHD4 depletion caused the Group 1 sites to change from inaccessible chromatin regions to accessible chromatin regions. In contrast, we did not detect any significant change in chromatin accessibility at Group 2 sites (Figs. 3C and 3D).

To examine why nucleosomes were depleted specifically at Group 1 sites, we analyzed the features of these DNA sequences. Notably, we found that nucleosome-repelling
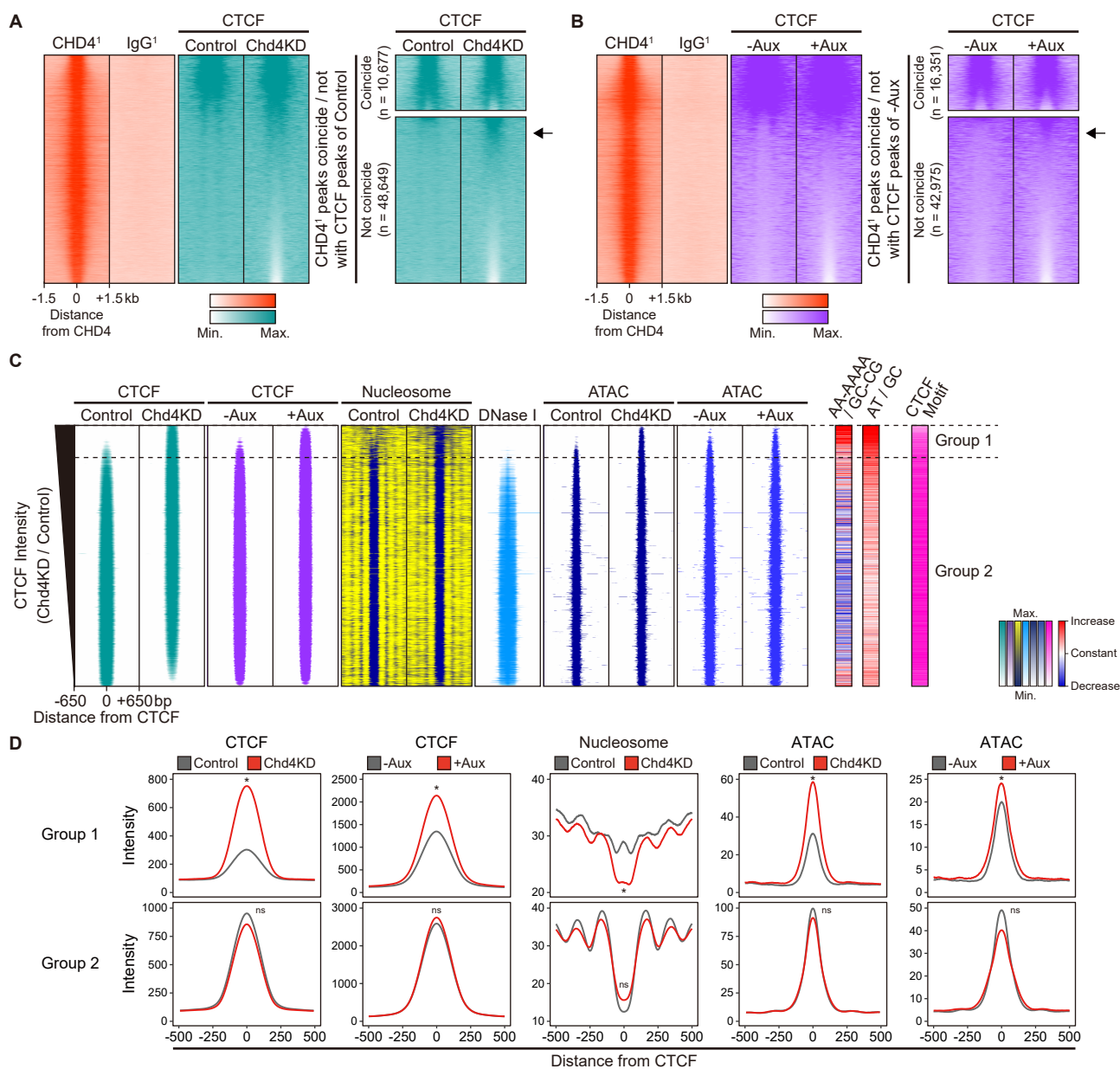
Fig. 3. CHD4 conceals aberrant CTCF-binding sites at TAD interiors by regulating chromatin accessibility. (A and B) Heatmaps representing CHD4[1], IgG[1] CUT&RUN, CTCF ChIP-seq of Control and Chd4KD cells (A), and CTCF ChIP-seq of -Aux and +Aux cells (B) at CHD4 peaks (left), and CHD4 peaks that coincide (top right) or do not coincide with CTCF peaks (bottom right). (C) Heatmaps displaying CTCF, nucleosomes, DNase I hypersensitive sites (DNase I), ATAC-seq signals (ATAC), the ratio of nucleosome-repelling sequences (AA-AAAA) or DNA bases (AT) to nucleosome-preferring sequences (GC-CG) or DNA bases (GC), and CTCF motifs at all CTCF peaks in Control and Chd4KD cells, as indicated at the top. See also Supplementary Fig. S3F. (D) Line plots showing average enrichments of CTCF, nucleosomes, and ATAC at Group 1 and Group 2 CTCF-binding sites. P values were derived using the Wilcoxon signed rank test (∗P < 1 × 10^{-100}; ns, not significant). (E) Examples of Group 1 sites, including representative results for CTCF, nucleosomes (Nuc.), ATAC, CHD4, and IgG. (F) Heatmaps displaying various histone H3 modifications in Control cells (black) and the heterochromatin marker, H3K9me3, in Control and Chd4KD cells (red), as indicated at the top. See also Supplementary Fig. S3H. (C and F) All heatmaps were aligned at 39,800 CTCF peaks (rows) and sorted in descending order by the relative CTCF intensity (Chd4KD/Control). The notation "4,736 Group 1 (11.9%)" represents CTCF-binding sites at which more CTCF was recruited in Chd4KD cells compared to Control cells (1.75-fold), while "35,064 Group 2 (88.1%)" represents the remainder of the CTCF-binding sites, at which CTCF binding was unchanged or reduced in Chd4KD cells. Heatmaps that are not labeled as Control or Chd4KD represent wild-type mESCs. (G) Heatmaps representing TAD positions (black), Group 1 (red), and Group 2 (blue) CTCF-binding sites at Control TADs (left) and -Aux TADs (right). All heatmaps were sorted in ascending order by the lengths of the TADs. (H) Percentage bar graph showing the distribution of Group 1 CTCF-binding sites at the interiors and borders of Control TADs (top) and -Aux TADs (bottom).
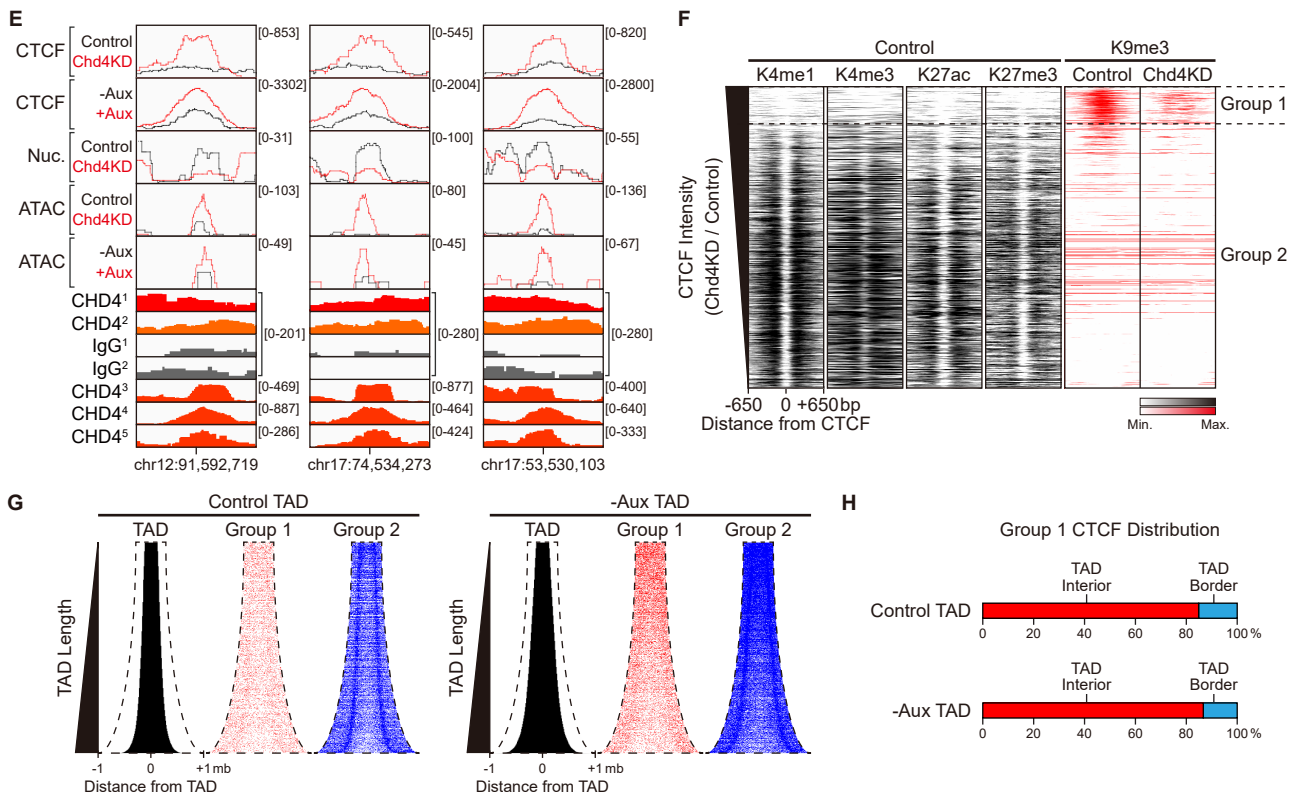
Fig. 3. Continued.

sequences (AA-AAAA) and DNA bases (AT) (Kaplan et al., 2009; Segal and Widom, 2009; Tillo and Hughes, 2009; Valouev et al., 2011) were more abundant within Group 1 sites compared to Group 2 sites (Fig. 3C). Moreover, CTCF motifs were also present within Group 1 sites (Fig. 3C), even though these sites were not typically targeted by CTCF in Control cells (Fig. 3C). Consistently, when we performed the same analysis using CTCF ChIP-seq peaks of −Aux and +Aux cells, we observed similar changes in CTCF, RAD21, and ATAC-seq signals upon CHD4 depletion at Group 1* sites (Supplementary Figs. S3F and S3G). The Group 1* sites also exhibited patterns similar to those of Group 1 sites in terms of depleted DNase I hypersensitive sites, enriched nucleosome-repelling sequences, and enriched CTCF motifs (Fig. 3C). Taken together, our results suggest that CHD4 maintains the inaccessible/closed chromatin state and conceals CTCF motifs at Group 1 sites by assembling nucleosomes, thereby preventing aberrant CTCF recruitment. Upon CHD4 depletion, however, nucleosomes are no longer assembled by CHD4 and instead tend to be removed from Group 1 sites due to the high-level enrichment of nucleosome-repelling sequences, resulting in the open/accessible chromatin states, exposure of CTCF motifs, and the mislocalization of CTCF. Thus, our data support the idea that the CHD4-mediated regulation of chromatin accessibility controls the appropriate recruitment of CTCF in mESCs.

We next determined the epigenetic features of the Group 1 sites using ChIP-seq data generated against various factors, including histone modifications. Consistent with our

observation of nucleosome enrichment (Fig. 3C), we observed that histone H3.1/2 was enriched at Group 1 sites in wild-type mESCs (Supplementary Fig. S3H). Remarkably, among the various histone modifications, H3K9me3 (Fig. 3F, Supplementary Fig. S3H) and the H3K9 methyltransferases, SUV39h1 and SUV39h2, but not SETDB1 (Supplementary Fig. S3H), were the only marks enriched at Group 1 sites in wild-type mESCs. As expected, given that the chromatin structure changed from the closed (nucleosome-enriched) to open (nucleosome-depleted) chromatin state at Group 1 sites upon CHD4 depletion, H3K9me3 was also diminished at these regions in Chd4KD cells (Fig. 3F); this reflected the absence of core histones, which are the substrates for H3K9 methyltransferases. This reduction was not related to any change in the expression of genes encoding H3K9 methyltransferases (Supplementary Table S7). Whereas CHD4 (Supplementary Fig. S3E) and ChAHP complexes (Supplementary Fig. S3I) were abundant at Group 1 sites, most of the analyzed chromatin remodelers (Supplementary Fig. S3E), mediators, histone variants, histone modifications (Supplementary Fig. S3H), nucleosome remodeling deacetylase (NuRD) complexes, architectural proteins, and transcription factors (Supplementary Fig. S3J) were sparse at the Group 1 sites. Together, these results indicate that CHD4 assembles core histones to conceal putative CTCF motifs at heterochromatic regions.

Lastly, we analyzed the positions of the Group 1 and Group 2 CTCF-binding sites with respect to the TADs. We observed that the Group 2 sites were mostly localized at the

TAD borders, whereas the Group 1 sites (equivalent to gained CTCF in Fig. 2I) were not (Fig. 3G). Instead, the latter sites were mainly located at the interior regions of TADs in Control cells (84.8%) and −Aux cells (86.7%) (Fig. 3H), which is consistent with our earlier observations (Figs. 2J and 2K, Supplementary Figs. S2B and S2C). In light of the reduced insulation scores (Fig. 2L) and APA results (Supplementary Fig. S2I), these data collectively imply that CHD4 depletion-triggered gained CTCF (Group 1 sites) directly contributes to disrupting local TAD organizations by forming a TAD borders and/or chromatin loops at TAD interiors.

CHD4 initially assembles core histones to conceal aberrant CTCF-binding sites and thereby prevents aberrant CTCF binding

Our results indicated that CHD4 depletion increased both chromatin accessibility (decreased nucleosome levels and increased ATAC-seq signals) and CTCF occupancy at Group 1 sites (Fig. 3, Supplementary Fig. S3). Since CHD4 is a chromatin remodeler, we speculated that CHD4 depletion would first alter chromatin accessibility, leading to the exposure of CTCF motifs, whereupon CTCF would be newly recruited to the Group 1 sites, resulting in aberrant CTCF binding (Fig. 3, Supplementary Fig. S3, see also Fig. 4A). To confirm this, we carried out two types of temporal experiments to determine the order of events (Figs. 4A and 4H). We first performed temporal depletion of CHD4 by varying the RNAi (siRNA against *Chd4*) treatment time (Fig. 4B). As expected, we observed a gradual decrease in *Chd4* expression as we increased the RNAi treatment time (Fig. 4C). At the same time points, we performed ChIP-seq against CTCF and histone H3. Consistent with our earlier results (Fig. 3, Supplementary Fig. S3), we did not detect any major change in the levels of CTCF or H3 at Group 2 sites upon gradual depletion of CHD4 (Fig. 4D). Notably, however, the CTCF level gradually increased (gained CTCF) and the H3 level gradually decreased (reflecting increased chromatin accessibility) at Group 1 sites upon temporal depletion of CHD4 (Fig. 4D). To assess these changes more accurately, we calculated the average intensity of CTCF and H3 at Group 1 sites at each time point (Figs. 4E and 4F). Importantly, consistent with our original speculation, we found that the H3 level decreased first, between 6-12 h of CHD4 depletion (Fig. 4F, red arrow), and the CTCF level increased thereafter between 12-48 h of CHD4 depletion (Fig. 4E, red arrow). These results were confirmed when we examined the individual loci of Group 1 sites (Fig. 4G, Supplementary Fig. S4A). Our findings indicate that, upon CHD4 depletion, changes in chromatin accessibility (closed to open chromatin state) occur first, followed by the aberrant gain of CTCF binding.

To further validate our findings, we performed temporal restoration of CHD4 levels (Fig. 4H). We first confirmed that CHD4 was completely degraded after 6 h of auxin treatment in CHD4-mAID cells (Supplementary Fig. S4B). Then, we removed the auxin by changing the cell culture medium. We found that CHD4 protein was gradually restored during 12 h of auxin withdrawal (Figs. 4I and 4J). To determine whether core histones assembly precedes CTCF eviction (or vice versa) at Group 1 sites upon CHD4 restoration, we varied the auxin

withdrawal time (Fig. 4I) and performed ChIP-seq against CTCF and histone H3. Consistent with our earlier results (Figs. 3 and 4D, Supplementary Fig. S3), we found that CHD4 depletion, which is equivalent to 6 h of auxin treatment without auxin withdrawal ("+Aux (6 h), withdraw 0 h"), decreased the H3 level and increased the CTCF level at Group 1 sites (Supplementary Fig. S4C). Upon temporal restoration of CHD4, the CTCF level gradually decreased (gained CTCF) and the H3 level gradually increased (reflecting increased chromatin accessibility) at Group 1 sites (Supplementary Fig. S4C). When we performed the same analysis using the CTCF ChIP-seq peaks of −Aux and +Aux cells ("+Aux (6 h), withdraw 0 h"), we observed the same rescue patterns for CTCF and H3 (i.e., a gradual decrease and increase in the levels of CTCF and H3, respectively) upon temporal restoration of CHD4 at Group 1* sites (Fig. 4K). To quantitatively assess these changes, we calculated the average intensities of CTCF and H3 at Group 1* sites for each time point (Figs. 4L and 4M). Importantly, we found that the H3 level increased between 0.5-3 h of auxin withdrawal (Fig. 4M, red arrow) and the CTCF level thereafter decreased between 3-12 h of auxin withdrawal (Fig. 4L, red arrow). These results were further confirmed when we examined individual loci of Group 1* sites (Fig. 4N, Supplementary Fig. S4D). Thus, CHD4 restoration first rescues chromatin accessibility (open to closed chromatin state), then rescues CTCF binding. Collectively, these results enabled us to determine the order of events and conclude that CHD4 directly regulates chromatin accessibility at Group 1 sites to conceal putative CTCF motifs and thereby prevent aberrant CTCF binding.

An RNA-binding intrinsically disordered region of CHD4 is required to prevent aberrant CTCF binding

We next questioned why CHD4 depletion majorly affects chromatin accessibility at Group 1 sites but has only modest effects at Group 2 sites, even though it binds to both types of site (Supplementary Fig. S3E). The most plausible explanation would be that there is a between-site-type difference in the presence of various factors. Indeed, such differences were seen among chromatin remodelers (Supplementary Fig. S3E), mediators, histone variants, histone modifications (Supplementary Fig. S3H), NuRD complexes, architectural proteins, and transcription factors (Supplementary Fig. S3J). Such factors, particularly the chromatin remodelers (CHD1 and brahma-related gene-1 [BRG1]), may compensate for the loss of CHD4 at Group 2 sites. At Group 1 sites, in contrast, these factors are sparse, while CHD4, ChAHP complex, and SUV39H1/2 are abundant. Together, these differences may explain why CHD4 depletion primarily affects the chromatin states of putative CTCF motifs at Group 1 sites.

Alternatively, differences in the chromatin state between Group 1 and 2 sites may explain the differential effects of CHD4 depletion. As described above, Group 1 sites represent heterochromatic regions, while Group 2 sites represent relatively euchromatic regions (Figs. 3C and 3F, Supplementary Fig. S3H) where RNAs are abundant. Since CHD4 is an RNA-binding protein (He et al., 2016; Hendrickson et al., 2016), we hypothesized that RNAs might inhibit the catalytic activity of CHD4 (given that CHD4 functions majorly at
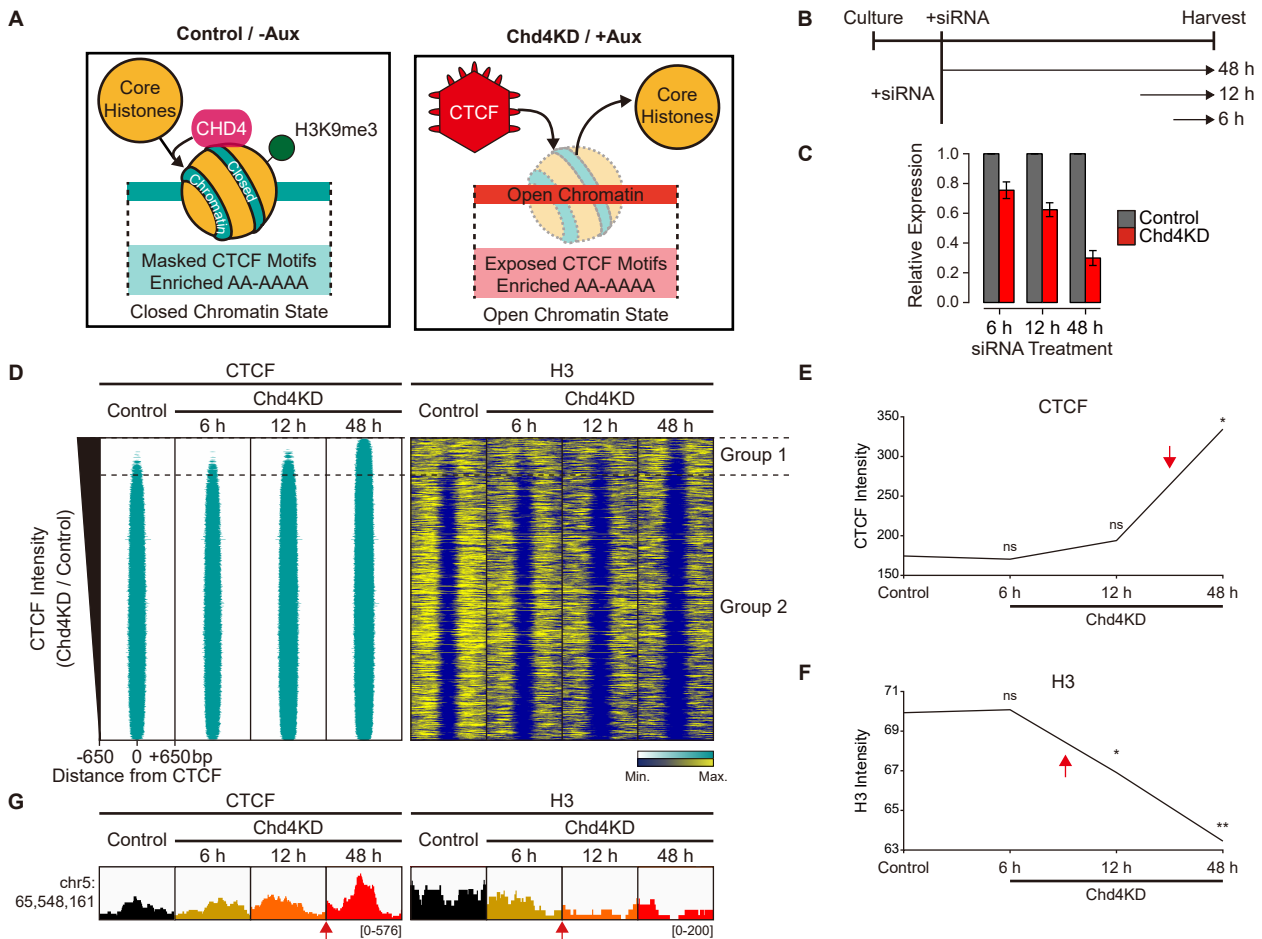
Fig. 4. CHD4 first assembles core histones and thereby prevents aberrant CTCF binding at the Group 1 sites. (A) Model illustrating the role of CHD4 at the Group 1 sites based on our results presented in Fig. 3 and Supplementary Fig. S3. (B) Schematic representation of temporal aspect experiments performed using samples treated with siRNAs against *Chd4* for various durations. (C) RT-qPCR analysis of *Chd4* expression levels observed after various siRNA treatment times. Error bars denote the standard deviations obtained from three biological replicates. The expression levels were normalized with respect to that of β-actin. (D) Heatmaps displaying CTCF and H3 ChIP-seq data obtained for various siRNA treatment times at all CTCF peaks in Control and Chd4KD cells, as indicated at the top. All heatmaps were aligned and sorted as described in Fig. 3C. See also Fig. 3C (siRNA against *Chd4* treated for 48 h). (E) Line plots showing average enrichments of CTCF at Group 1 sites. *P* values were derived using the Wilcoxon signed rank test (*$P < 1 \times 10^{-100}$; ns, not significant). (F) Line plots showing average enrichments of H3 at Group 1 sites. *P* values were derived using the Wilcoxon signed rank test (*$P < 1 \times 10^{-3}$; **$P < 1 \times 10^{-15}$; ns, not significant). (E and F) *P* values were calculated by comparing the Control and each Chd4KD time point. (G) Examples of Group 1 sites, including representative results for CTCF and H3 across a time course of siRNA treatment. See also Supplementary Fig. S4A. (H) Model representing aberrant CTCF binding upon CHD4 depletion (left) and our prediction after CHD4 restoration via auxin withdrawal (right). (I) Schematic representation of auxin withdrawal experiments performed for various durations. (J) Immunoblots of CHD4 after auxin treatment (6 h) followed by various auxin withdrawal times in CHD4-mAID cells and E14Tg2a wild-type (E14). α-Tubulin was detected as a loading control. (K) Heatmaps displaying CTCF and H3 ChIP-seq data for -Aux and +Aux followed by various auxin withdrawal times at all CTCF peaks in -Aux and +Aux (auxin withdrawal 0 h) cells, as indicated at the top. All heatmaps were aligned at 59,327 CTCF peaks (rows) and sorted in descending order by the relative CTCF intensity (auxin withdrawal 0 h/-Aux). The notation "7,225 Group 1* (12.2%)" indicates CTCF-binding sites at which more CTCF was recruited in +Aux (auxin withdrawal 0 h) compared to -Aux cells (1.5-fold), while "52,102 Group 2* (87.8%)" indicates the remainder of the CTCF-binding sites. See also Supplementary Figs. S3F (auxin was treated for 24 h) and S4C. (L) Line plots showing average enrichments of CTCF at Group 1* sites. *P* values were derived using the Wilcoxon signed rank test (*$P < 1 \times 10^{-100}$; **$P < 1 \times 10^{-300}$; ns, not significant). (M) Line plots showing average enrichments of H3 at Group 1* sites. *P* values were derived using the Wilcoxon signed rank test (*$P < 1 \times 10^{-30}$; **$P < 1 \times 10^{-50}$; ns, not significant). (L and M) *P* values were calculated by comparing results obtained from -Aux and +Aux at each auxin withdrawal time point. (N) Examples of Group 1* sites, including representative results for CTCF and H3 across a time course of auxin withdrawal. See also Supplementary Fig. S4D. (E-G and L-N) Red arrows indicate the time points at which the intensity (CTCF or H3) changes most rapidly after siRNA treatment (E-G) or auxin withdrawal (L-N).
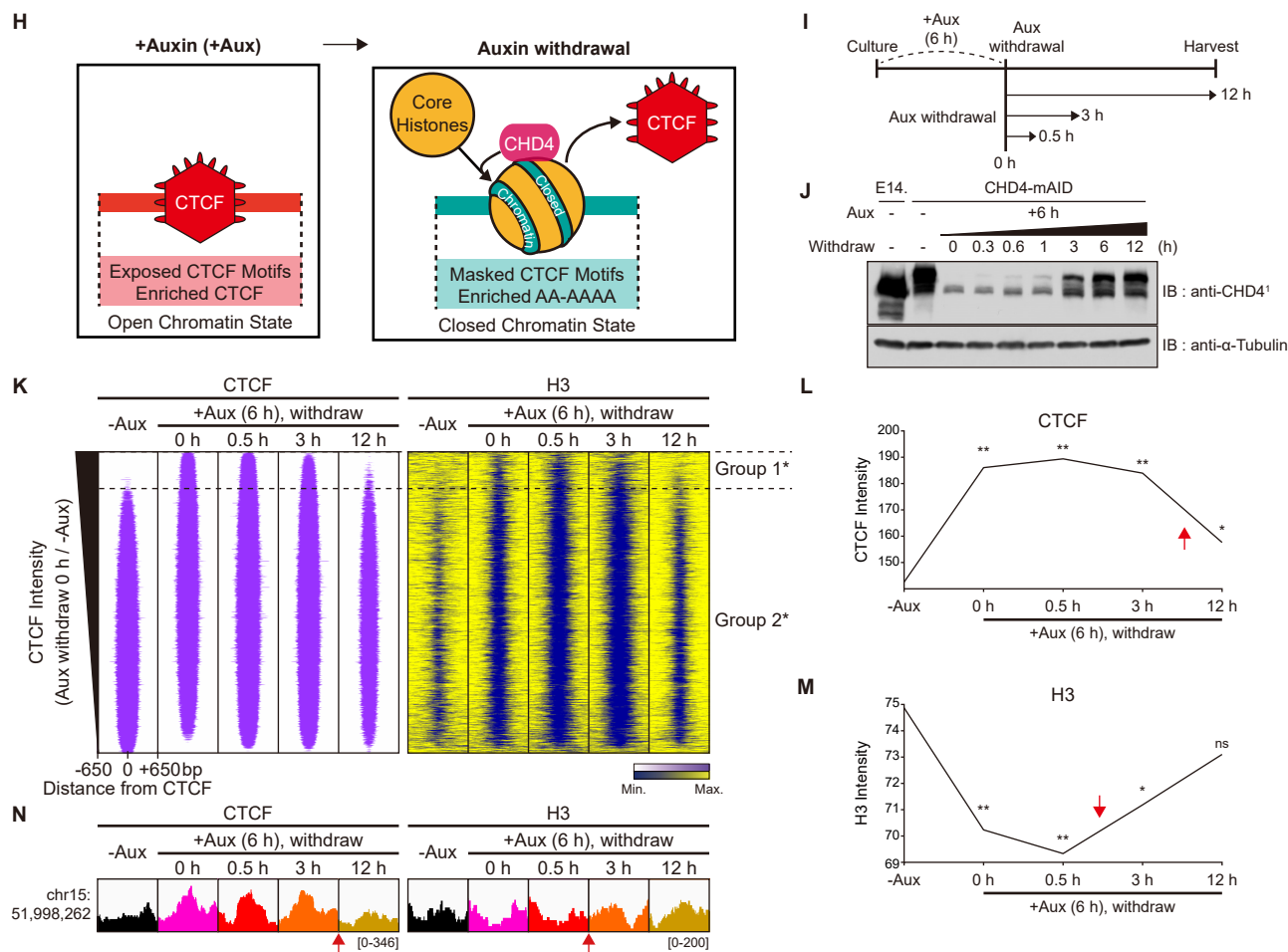
Fig. 4. Continued.

RNA-depleted heterochromatic Group 1 sites in wild-type mESCs), which could explain why CHD4 depletion primarily affects Group 1 sites. We first confirmed that CHD4 antibodies could be used for immunoprecipitation (Supplementary Fig. S5A), and then performed RIP. We found that CHD4 bound the *Oct4* and *Nanog* RNAs (Fig. 5A), which are very abundant in mESCs (Ang et al., 2011). Furthermore, we confirmed the RNA binding ability of purified CHD4 (Supplementary Fig. S5B) using RNA EMSAs (Fig. 5B, Supplementary Fig. S5C). Our results demonstrated that CHD4 binds to RNAs, which is consistent with previous reports (He et al., 2016; Hendrickson et al., 2016). To determine the influence of RNA binding on the activity of CHD4, we performed *in vitro* nucleosome-sliding assays with various concentrations of RNA. Interestingly, we observed that the nucleosome-sliding activity of CHD4 decreased as the RNA concentration increased (Fig. 5C, Supplementary Fig. S5D). Collectively, our data suggest that RNAs bind to CHD4 and inhibit its catalytic activity, which may explain (along with the other possibilities mentioned above) why CHD4 plays only a modest role at RNA-abundant euchromatic Group 2 sites.

To investigate the importance of the RNA-binding ability of CHD4 in the regulation of CTCF binding, we first sought to identify the specific RNA-binding domain(s) of CHD4. Many RNA-binding proteins contain IDRs that can bind to RNAs (Lin et al., 2016; Molliex et al., 2015). Here, we used MobiDB (Piovesan et al., 2018; 2021) to determine the RNA-binding IDRs of CHD4. Our MobiDB analysis predicted four IDRs (IDR 1-4) within *Chd4* (Fig. 5D). To test these predictions, we first generated and purified proteins containing IDR 1/2, IDR 3/4, and a negative control region (N.C.), and then performed RNA EMSAs. We found that only IDR 1/2 (1-1059 bp of *Chd4*) exhibited RNA-binding ability (Fig. 5E). Our further experiments showed that both IDR 1 (1-450 bp) and IDR 2 (706-1059 bp) could bind to RNAs (Fig. 5F). Thus, we specifically identified the RNA binding domains of CHD4. To exclude off-target effects, we narrowed the relevant regions within the IDRs. Since protein-RNA interaction interfaces are known to be formed by clusters of positively charged residues that are scattered on protein surfaces (Shazman and Mandel-Gutfreund, 2008; Shazman et al., 2007), we identified four clusters of positively charged residues (charged domains, CDs 1-4) within IDR 1 and IDR 2 (Fig. 5D). We next created DNA constructs encoding RNA-binding-defective *Chd4* mutants harboring deletion of all regions four CDs in IDR 1 and IDR 2 (Fig. 5D, CDΔ[1]), and also generated deletion of
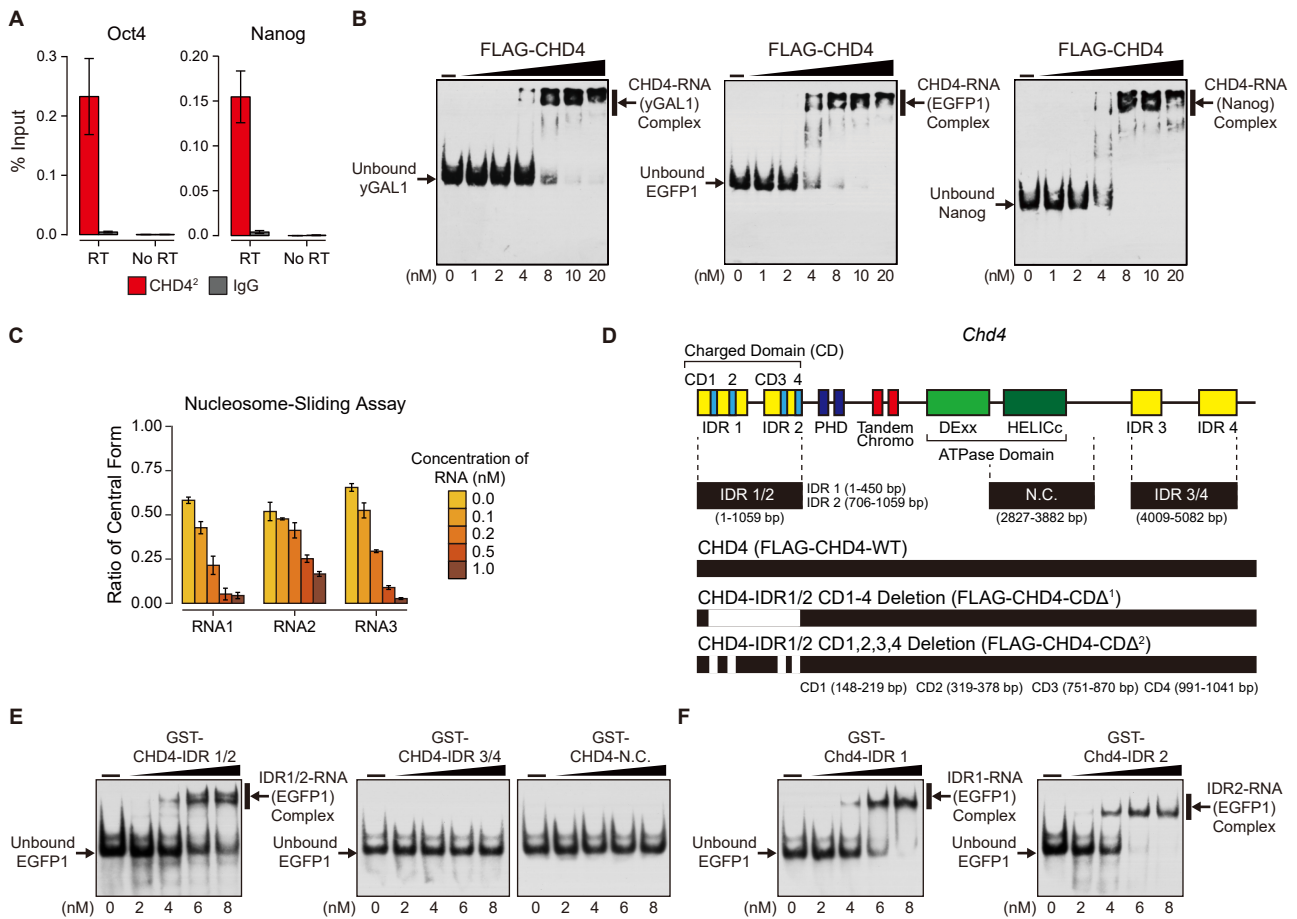
Fig. 5. RNA binding ability of CHD4 is required to prevent aberrant CTCF binding at the Group 1 sites. (A) RNA immunoprecipitation (RIP) assay results for CHD4. RIP was carried out with the indicated antibodies. All signals were normalized with respect to the input signal. Error bars denote the standard deviations obtained from three biological replicates. (B) Electrophoretic mobility shift assays (EMSAs) for the *yGAL1* (left), *EGFP1* (middle), and *Nanog* (right) RNAs were performed with purified FLAG-CHD4. See also Supplementary Fig. S5C. (C) Bar graphs showing the ratio of the central form to total nucleosomes (central and lateral forms) according to the results of nucleosome-sliding assays (Supplementary Fig. S5D). RNA1, -2, and -3 denote *yGAL1*, *EGFP1*, and *EGFP2*, respectively. Error bars denote the standard deviations obtained from three biological replicates. (D) Schematic representation of various domains within *Chd4* genes. Specific positions of domains used in the present work are indicated. (E and F) EMSAs for the *EGFP1* RNAs were performed with purified GST-CHD4-IDR 1/2 (E, left), GST-CHD4-IDR 3/4 (E, middle), GST-CHD4-N.C. (E, right), GST-CHD4-IDR 1 (F, left), and GST-CHD4-IDR 2 (F, right). (B, E, and F) "Unbound RNA" indicates free RNA, and "CHD4 (or IDR)-RNA Complex" indicates the binding of the purified protein to each RNA species. (G) Heatmaps displaying CTCF ChIP-seq data for -Aux, ++Aux, and ++Aux followed by ectopic transfection of various FLAG-CHD4 constructs at all CTCF peaks in ++Aux and ++Aux-FLAG-CHD4-WT cells, as indicated on top. All heatmaps were aligned at 57,631 CTCF peaks (rows) and sorted in descending order by the relative CTCF intensity (++Aux/++Aux-FLAG-CHD4-WT). The notation "4,326 Group 1* (7.5%)" indicates CTCF-binding sites at which more CTCF was recruited in ++Aux compared to ++Aux-FLAG-CHD4-WT cells (1.5-fold), while "53,305 Group 2* (92.5%)" represents the remainder of the CTCF-binding sites. See also Supplementary Figs. S5E-S5G. (H) Box plots showing the enrichments of CTCF at Group 1* sites. *P* values were calculated by comparing the results from -Aux with those from each of the other samples. The horizontal line and the white rhombus in the box denote the median and mean, respectively. *P* values were derived using the Wilcoxon signed rank test (*$P < 1 \times 10^{-100}$; **$P < 1 \times 10^{-200}$; ns, not significant). (I) Examples of Group 1* sites, including representative results for CTCF. See also Supplementary Fig. S5H.

each individual CD (Fig. 5D, CDΔ²) to minimize the off-target effects. To perform rescue experiments, we transfected cells with ectopic DNA constructs encoding FLAG-tagged full-length CHD4 (FLAG-CHD4-WT) or RNA-binding-defective CHD4 (FLAG-CHD4-CDΔ¹ or -CDΔ²), and then treat auxin to deplete endogenous CHD4 proteins (Supplementary Fig. S5E, left, "+Auxin 24 h (+Aux)"). Interestingly, although FLAG-

CHD4-CDΔ¹,² was successfully overexpressed (Supplementary Fig. S5F, left, see lanes 6 and 7), we failed to overexpress FLAG-CHD4-WT (Supplementary Fig. S5F, left, see lane 3). We found that this failure was due to the presence of endogenous CHD4-mAID proteins, which probably formed dimers with ectopic FLAG-CHD4-WT proteins, leading to their mutual degradation. In contrast, FLAG-CHD4-CDΔ¹,² failed
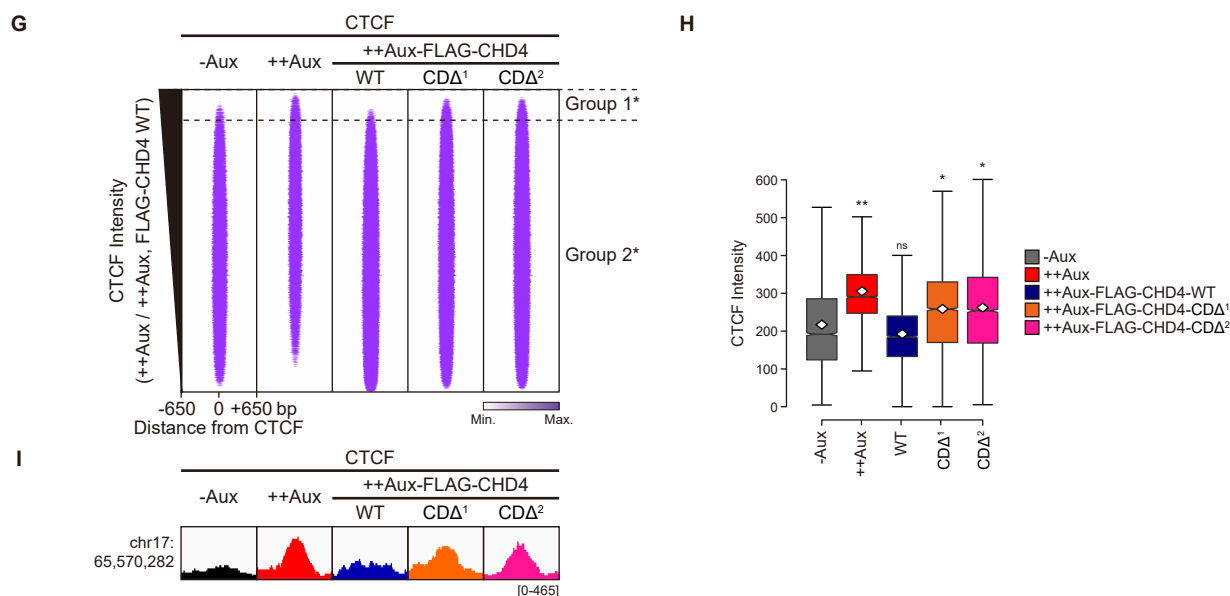
Fig. 5. Continued.

to form dimers with endogenous CHD4-mAID proteins, and thus could be stably expressed. To resolve this, we treated auxin before transfection (Supplementary Fig. S5E, right, "++Auxin 48 h (++Aux)"), so the endogenous CHD4-mAID proteins were degraded prior to the expression of FLAG-CHD4-WT. This enabled us to observe the stable expression of FLAG-CHD4-WT proteins (Supplementary Fig. S5F, right, see lane 10). To elucidate the role of the RNA-binding CDs of CHD4 in the regulation of CTCF binding, we performed rescue experiments by completely depleting endogenous CHD4 proteins and then transfecting FLAG-CHD4-WT/-CDΔ^1/-CDΔ^2 at a ratio that ensured their equal expression (Supplementary Fig. S5G). We then performed CTCF ChIP-seq and observed that, consistent with our earlier results (Figs. 3 and 4), CTCF was aberrantly gained at Group 1* sites upon CHD4 depletion (++Aux) (Fig. 5G). Importantly, we found that ectopic transfection of FLAG-CHD4-WT (full-length CHD4) rescued the CTCF levels at Group 1* sites (Figs. 5G-5I, Supplementary Fig. S5H), indicating that CHD4 is directly associated with preventing aberrant CTCF bindings at these sites. However, strikingly, we observed that CTCF levels were not entirely restored in either RNA-binding-defective *Chd4* mutant (FLAG-CHD4-CDΔ^{1,2}) (Figs. 5G-5I, Supplementary Fig. S5H). This indicates that the RNA-binding CDs within the IDRs of CHD4 are required to prevent aberrant CTCF binding in mESCs.

CHD4 is required for the repression of B2 SINEs in mESCs

Having defined the epigenetic features of H3K9me3-enriched heterochromatic aberrant CTCF-binding sites (Group 1) and determined the role of CHD4 at these sites, we next sought to elucidate the genomic features of Group 1 sites by examining the distribution of genomic regions via genome annotation. Notably, we found that SINEs (particularly B2 SINEs), which are well-known retrotransposons (McClintock, 1950), were enriched specifically at Group 1 sites (Fig. 6A); this is consistent with previous reports that H3K9me3 sup-

presses SINEs (Martens et al., 2005; Varshney et al., 2015), which harbor CTCF motifs (Bourque et al., 2008; Schmidt et al., 2012). Since CHD4 depletion caused the loss of nucleosomes/core histones (Figs. 3 and 4, Supplementary Figs. S3 and S4), resulting in the reduction of H3K9me3 levels (Fig. 3F) at Group 1 sites, we next investigated whether this affected the transcription of B2 SINE by performing mRNA-seq, total RNA-seq, and global run-on sequencing (GRO-seq). Surprisingly, CHD4 depletion elevated the signals for all three assays at Group 1 sites (Fig. 6B, Supplementary Fig. S6A), where B2 SINEs are enriched (Fig. 6A). Our GRO-seq analysis further revealed that the number of *de novo* SINE transcripts increased more than 2-fold upon CHD4 depletion (Supplementary Fig. S6B). Notably, our total RNA-seq analysis showed that the expression levels of B2 SINEs localized at the Group 1 sites were up-regulated upon CHD4 depletion (Supplementary Fig. S6C), strongly implying that CHD4 directly represses B2 SINE transcripts at Group 1 sites. As B2 SINE is mainly suppressed by H3K9me3 via SUV39h1 and SUV39h2 in mESCs (Martens et al., 2005; Varshney et al., 2015), our results indicate that B2 SINEs within Group 1 sites are repressed by enriched H3K9me3 and its methyltransferases in Control cells, but become de-repressed upon CHD4 depletion due to the reduction of the repressive mark, H3K9me3.

To confirm this, we determined the expression levels of various retrotransposons, including long interspersed nuclear elements (LINEs), endogenous retroviruses (ERVs), and SINEs, using RT-qPCR. Consistently, the expression level of B2 SINE, a major mouse SINE, was dramatically up-regulated in Chd4KD cells, whereas the other retrotransposons (e.g., LINE-1 and IAP) were not significantly altered (Fig. 6C). B2 SINE transcripts are generally transcribed by RNA polymerase III (RNAPIII), but some B2 SINEs are located within introns and may be affected by RNA polymerase II (RNAPII). To exclude the possible relevance of RNAPII-affected B2 SINES, we applied the RNAPII inhibitor, α-amanitin, and determined
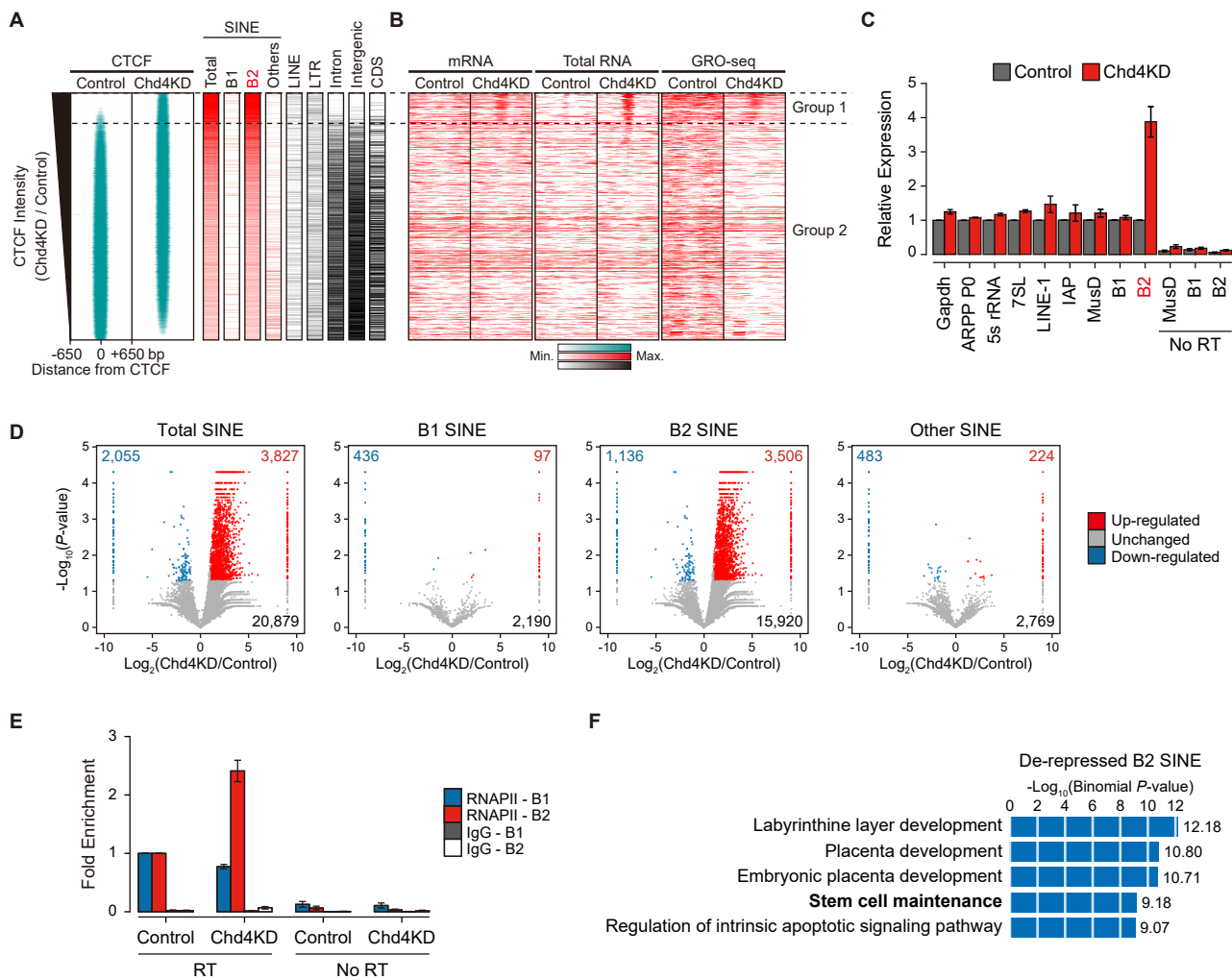
Fig. 6. CHD4 represses B2 SINEs in mESCs. (A) Heatmaps displaying the genomic contents of the mouse genome. Coding sequence (CDS) includes the exon, promoter, transcription start site (TSS), and transcription termination site (TTS). (B) Heatmaps representing mRNA-seq, total RNA-seq, and GRO-seq data in Control and Chd4KD cells. (A and B) All heatmaps were aligned and sorted as described in Fig. 3C. See also Fig. 3 and Supplementary Fig. S3. (C) RT-qPCR analysis of retrotransposons in mESCs. *Gapdh* and *ARPP* are transcribed by RNAPII. The 5s rRNA, 7SL, and retrotransposons are mainly transcribed by RNAPIII. The obtained expression levels were normalized with respect to that of the 28S rRNA, which is transcribed by RNA polymerase I. Error bars denote the standard deviations obtained from three biological replicates. (D) Volcano plots showing the differential expression levels of SINEs (Total, B1, B2, and other SINEs), as measured using total RNA-seq. Only SINEs localized at intergenic regions were analyzed. Blue and red dots indicate the expressions of SINEs that are significantly decreased and increased, respectively, upon *Chd4* knockdown ($P \leq 0.05$ and $\log_2$ (Chd4KD/Control) $\leq$ -1 or $\geq$ +1). (E) RIP assay of B1 and B2 SINEs bound by RNAPII. Blue, red, gray, and white boxes indicate RNAPII (8WG16)-bound B1 SINE, RNAPII (8WG16)-bound B2 SINE, IgG-bound B1 SINE, and IgG-bound B2 SINE, respectively. Immunoprecipitated SINEs were normalized with respect to the levels of B1 or B2 SINEs bound by RNAPII in Control cells. Error bars denote the standard deviations obtained from three biological replicates. (F) Results of genomic regions enrichment of annotations tool (GREAT) analysis, which was used to identify genes located near de-repressed B2 SINEs (see Materials and Methods section). (G) Examples of de-repressed B2 SINEs (- and + strand total RNA-seq) and disrupted localization of RNAPII (ChIP-exo) at various stem cell maintenance-related genes. Orange, green, and blue boxes highlight changes in total RNA-seq (- strand), total RNA-seq (+ strand), and RNAPII (ChIP-exo), respectively. Genes, enhancers, and B2 SINE loci and their strand directions are marked at the bottom. See also Supplementary Fig. S6F.

the expression of B2 SINEs in mESCs. Consistent with the above-described results, B2 SINEs were specifically up-regulated by *Chd4* knockdown even when RNAPII was inhibited (Supplementary Fig. S6D). These results were also confirmed by our total RNA-seq analysis of SINEs localized at intergenic regions of the mouse genome: CHD4 depletion specifically up-regulated B2 SINEs but did not affect B1 SINEs or other SINEs (Fig. 6D). Together, these data indicate that CHD4 represses global B2 SINE transcripts in mESCs.
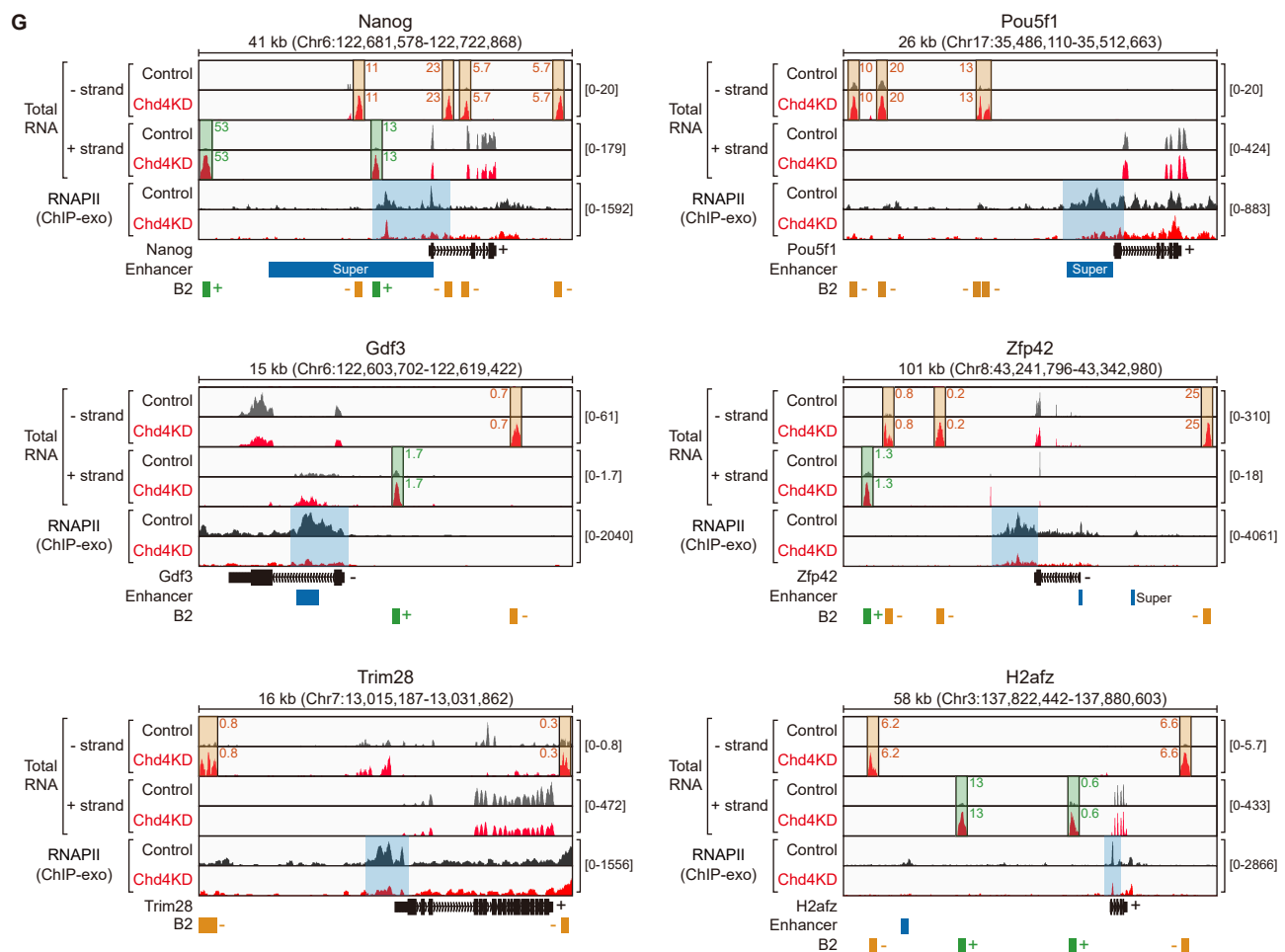
Fig. 6. Continued.

CHD4 depletion-triggered de-repressed B2 SINE may hinder RNA polymerase II recruitment at the transcription start sites of pluripotent genes

We next investigated the role of CHD4 depletion-induced de-repressed B2 SINE in mESCs. As up-regulated B2 SINE RNA has been reported to directly bind RNAPII and inhibit transcription (Allen et al., 2004; Espinoza et al., 2004), we examined whether CHD4 depletion-triggered de-repressed B2 SINE RNA could do the same. We first performed RIP and confirmed that de-repressed B2 SINE transcripts have a binding preference for RNAPII in Chd4KD cells (Fig. 6E). We then used genomic regions enrichment of annotations tool (GREAT) analysis (McLean et al., 2010) to identify genes located in the vicinity of de-repressed B2 SINEs. Interestingly, our results revealed that genes important for stem cell maintenance (pluripotent genes) were located near de-repressed B2 SINEs (Fig. 6F) and Group 1 CTCF-binding sites (Supplementary Fig. S6E). To investigate whether de-repressed B2 SINE transcripts hinder the recruitment of RNAPII to the transcription start sites (TSSs) of genes involved in stem cell maintenance, we analyzed our total RNA-seq data and publicly available RNAPII ChIP-exo data (GSE64825, see also Supplementary Table S4). For a detailed analysis, we separat-

ed the total RNA-seq data into those for the −/+ strands and confirmed the presence of de-repressed B2 SINEs in Chd4KD cells (Fig. 6G). Notably, Chd4KD cells exhibited a significant reduction of RNAPII recruitment at TSSs when abundant de-repressed B2 SINEs were located nearby (Fig. 6G), whereas no reduction of RNAPII recruitment was observed at TSSs without a nearby de-repressed B2 SINE (Supplementary Fig. S6F). The observed changes in RNAPII recruitment were not derived from any change in the expression levels of the RNAPII subunits (Supplementary Fig. S6G). To determine whether this reduction of RNAPII recruitment at TSSs could dysregulate the expression levels of stem cell maintenance-related genes, we analyzed total RNA-seq and mRNA-seq data. Notably, we observed that genes with abundantly de-repressed B2 SINEs, which exhibited reduced RNAPII recruitment at their TSSs (Fig. 6G), were down-regulated (Supplementary Fig. S6H, red) compared to genes that lacked a nearby de-repressed B2 SINE (Supplementary Fig. S6H, gray). Taken together, our data indicate that *Chd4* knockdown induces the de-repression of B2 SINEs, thereby preventing the proper recruitment of RNAPII to the TSSs of stem cell maintenance genes (e.g., *Nanog* and *Oct4*) and resulting in the down-regulation of these genes. Thus, CHD4 is required
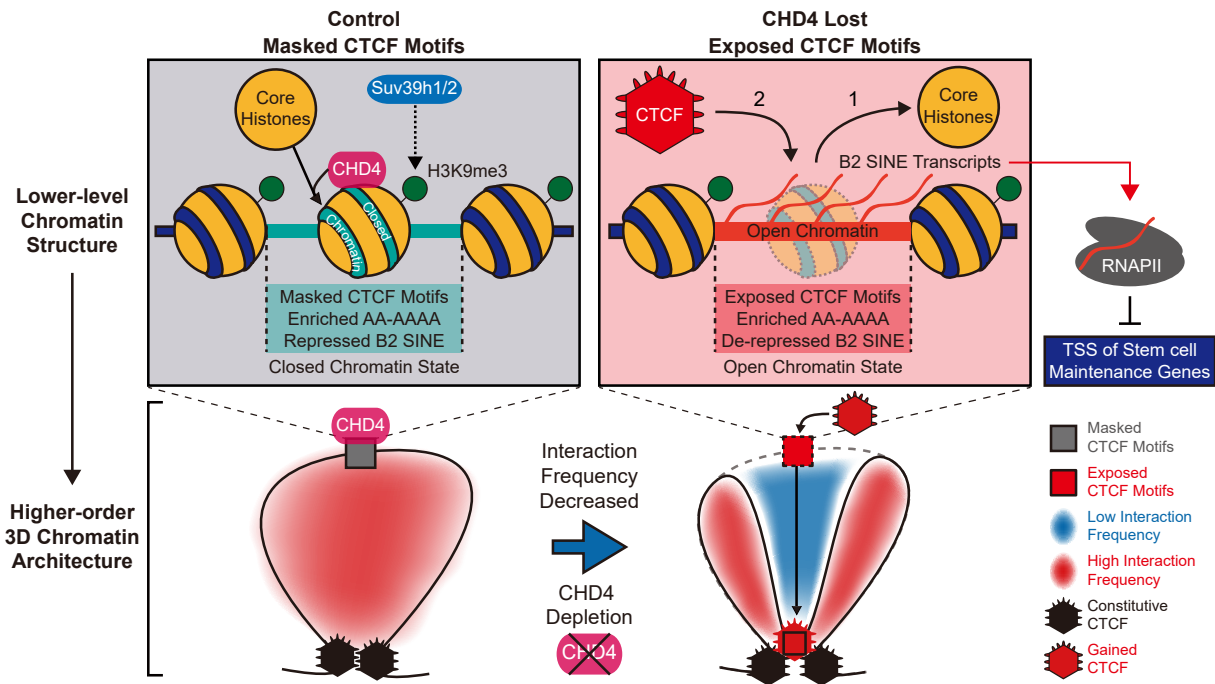
Fig. 7. Proposed model illustrating the function of CHD4 in regulating lower-level chromatin structures (top) to maintain the higher-order 3D chromatin organizations (bottom).

to repress B2 SINEs and thereby hinder the recruitment of RNAPII to the TSSs of pluripotent genes.

DISCUSSION

Here, we explored the changes in the 3D chromatin organizations of mESCs upon CHD4 depletion using *in situ* Hi-C. We found that CHD4 critically contributes to the maintenance of TADs (Hi-C interactions within TADs and TAD positions) by preventing aberrant CTCF binding at the TAD interior. Most importantly, we discovered that CHD4 conceals aberrant CTCF-binding sites (Group 1 sites) by regulating chromatin accessibility, thereby maintaining these sites as inaccessible, H3K9me3-enriched heterochromatic regions. Upon CHD4 depletion, the aberrant CTCF-binding sites become accessible, exposing the putative CTCF motifs, and aberrant CTCF recruitment occurs; this results in weakened/separating TADs. We also showed that these aberrant CTCF-binding sites are embedded in B2 SINEs, which are normally repressed by CHD4 in wild-type mESCs. Upon CHD4 depletion, these B2 SINEs are de-repressed (due to reduction of the B2 SINE-repressive mark, H3K9me3) and interact with RNAPII; this disrupts the recruitment of RNAPII to the TSSs of stem cell maintenance-related genes, resulting in dysregulated expression of the relevant genes. Collectively, our results reveal a novel mechanism through which CHD4 regulates chromatin accessibility at aberrant CTCF-binding sites to safeguard appropriate CTCF binding and maintain the local TAD organizations in mESCs (Fig. 7).

Since the genome is packaged into 3D chromatin organizations and most of the biological processes, including transcription, occur in this context, it is important to investigate the 3D organization of the genome. Previous studies characterized the 3D chromatin organization in terms of compartments (Lieberman-Aiden et al., 2009; Simonis et al., 2006), TADs (Dixon et al., 2012; Nora et al., 2012), contact/loop domains (Rao et al., 2014; 2017), and insulated neighborhoods (Dowen et al., 2014; Ji et al., 2016). Recent reports have mainly addressed the importance of CTCF and cohesin in 3D genome topology by observing the disruption of 3D chromatin structures after CRISPR/Cas9 system-mediated deletion/orientation change of CTCF-binding sites (de Wit et al., 2015; Guo et al., 2015; Sanborn et al., 2015), deletion of cohesion (Luppino et al., 2020; Rao et al., 2017; Szabo et al., 2020), and deletion of CTCF (Luppino et al., 2020; Nora et al., 2017; Szabo et al., 2020). These studies mainly focused on the properties and formation of the 3D chromatin organization and the key factors (e.g., CTCF and cohesin) that compose its structures. However, intriguing questions remained, such as: (1) How are TADs established or maintained? (2) How are aberrant CTCF binding and associated Hi-C interactions at TAD interiors prevented? (3) How is CTCF recruitment at the TAD borders regulated? (4) Are chromatin remodelers involved in CTCF recruitment? (5) What is the role of chromatin remodelers in 3D chromatin organizations? We believe that our present work provides clues to these questions by elucidating a CHD4-mediated mechanism that conceals aberrant CTCF-binding sites at the TAD interiors by regulating chromatin accessibility, thereby preventing inappropriate CTCF binding and safeguarding both proper CTCF recruitment and 3D chromatin organization in mESCs.

It is interesting to set our results in the context of evolutionary aspects. A previous report compared Hi-C data from four mammals and identified two types of CTCF sites: (1)

conserved CTCF sites, which are conserved across species and enriched at strong TAD borders; and (2) divergent CTCF sites, which lack conservation across species, are newly evolved, and localize at TAD interiors (Vietri Rudan et al., 2015). The authors also demonstrated that the evolutionary dynamics of intra-TAD interactions, which reflect the actions of divergent CTCF sites, could play a critical role in enhancer-promoter interactions within TADs (Vietri Rudan et al., 2015). In support of this, a recent review suggested that 3D chromatin organization may have significantly impacted the evolution of transcriptional regulation by increasing regulatory complexity (the combination of enhancer and promoter contacts) (Acemel et al., 2017). Taken together, these findings indicate that the evolution of gene regulation is strongly correlated with 3D chromatin organization, especially intra-TAD interactions involving divergent CTCF sites. Interestingly, the divergent CTCF sites that primarily reside at TAD interiors have been spread through genomes by retrotransposition of B2 SINE transposable elements, which contain CTCF motifs in their consensus sequences (Bourque et al., 2008; Schmidt et al., 2012). Our present data illustrate that the chromatin remodeler, CHD4, conceals the B2 SINE-originated divergent/inappropriate CTCF-binding sites by regulating chromatin accessibility, thereby preventing three events: (1) aberrant CTCF recruitment; (2) aberrant CTCF binding-associated changes in local TADs (separating and weakened TADs); and (3) overexpression of B2 SINE transcripts in mESCs.

The previous studies on chromatin remodeler mainly focused on examining ATP catalysis-based nucleosome sliding *in vitro* (Blosser et al., 2009; Narlikar et al., 2001; Rippe et al., 2007; van Vugt et al., 2009). More recent reports have investigated the distribution of chromatin remodelers throughout the genome (Morris et al., 2014), especially at promoter-TSSs (de Dieuleveult et al., 2016), and their functional roles in diverse cellular processes (de Dieuleveult et al., 2016; Gaspar-Maia et al., 2009; Micucci et al., 2015; Skene et al., 2014). Such studies showed that CHD1 plays a critical role in maintaining the open chromatin structure by preventing heterochromatinization and is required for pluripotency in mESCs (Gaspar-Maia et al., 2009). In addition, various chromatin remodelers (CHD1, CHD2, CHD4, CHD6, CHD8, CHD9, BRG1, and EP400) have been shown to act together at specific nucleosomes near promoters and regulate transcription in mESCs (de Dieuleveult et al., 2016). CHD4 is known to be part of the protein network underlying pluripotency (PluriNetwork) (Som et al., 2010) and critically ensures mESC identity by regulating pluripotency- and differentiation-associated genes (Zhao et al., 2017). However, the detailed molecular mechanisms through which chromatin remodelers function in mESCs have remained elusive, especially given that the prior studies have mainly focused on chromatin remodelers in transcribed regions, which account for only about 2% of the mammalian genome.

Among the studies that did not focus specifically on transcribed regions, several reports showed that chromatin remodelers are closely associated with CTCF. For example, RNAi-mediated knockdown of *Chd8* revealed that CHD8 interacts with CTCF and is required for the CTCF-dependent insulator function at specific CTCF-binding sites (*H19*, the β-globin control region, and promoter regions of the *BRCA1* and *c-myc* genes) in HeLa and Hep3B cells (Ishihara et al., 2006). Experiments involving RNAi-mediated knockdown of *Snf2h* (*Smarca5*) uncovered that SNF2H promotes CTCF binding (i.e., maintains CTCF occupancy) and organizes the nucleosome positioning adjacent to CTCF-binding sites, but does not act as a CTCF-loading factor, in HeLa cells (Wiechens et al., 2016). Interestingly, the same authors examined the knockdown effects of other chromatin remodelers, including CHD4, but did not detect major changes in CTCF occupancy or nucleosome occupancy/positions near CTCF-binding sites. This apparent discrepancy with our present results may reflect the use of different cell lines (mESC vs HeLa) and/or that the previous study measured changes in CTCF/nucleosome occupancy/position as average line plots, which only exhibit general changes rather than context-specific changes. Recently, a study involving the conditional knockout of *Chd4* in the granule neurons of mouse cerebellum revealed that CHD4 regulates chromatin accessibility and cohesin binding to coordinate intra-domain loop strength and gene expression in mouse brain (Goodman et al., 2020). Furthermore, knockout of *Adnp*, which forms ChAHP complex with CHD4 and HP1β/γ (Ostapcuk et al., 2018), in mESCs revealed that ChAHP complex, which mainly localizes to less diverged B2 SINEs, prevents aberrant CTCF recruitment and maintains TAD organizations (Kaaij et al., 2019). However, although the results obtained from *Adnp* knockout and our CHD4 depletion seem very similar, the primary mechanisms used to maintain proper CTCF binding are explicitly different. In the case of ADNP, the ChAHP complex competes with CTCF for common binding sequences (motifs), and thereby prevents aberrant CTCF binding. In the previous study, the authors focused primarily on ADNP and only peripherally addressed the functional role of CHD4 (all of the data were obtained by the deletion of ADNP, not CHD4). Conversely, our primary target was CHD4. Notably, we found that CHD4 directly regulates chromatin accessibility at putative/aberrant CTCF motifs (Group 1 CTCF-binding sites), concealing these sites as H3K9me3-enriched heterochromatic regions to prevent aberrant CTCF binding. When CHD4 is intact, the aberrant CTCF-binding sites are concealed by nucleosomes, exhibiting the closed chromatin states, as shown by our analyses of enriched nucleosomes (Figs. 3C-3E), core histones (Fig. 4, Supplementary Fig. S3H), the constitutive heterochromatin marker H3K9me3 (Fig. 3F, Supplementary Fig. S3H), and depleted DNaseI and ATAC-seq signals (Figs. 3C-3E). Upon CHD4 depletion, the concealed aberrant CTCF-binding sites become accessible, as shown by our findings that enriched nucleosomes (Figs. 3C-3E), core histones (Fig. 4), and H3K9me3 (Fig. 3F) are lost while ATAC-seq signals are increased (Figs. 3C-3E). We performed temporal depletion/restoration of CHD4 to determine the order of events (Fig. 4, Supplementary Fig. S4), and report that CHD4 first assembles core histones at the aberrant CTCF-binding sites and thereby prevents aberrant CTCF recruitment. We also performed rescue experiments by transfecting CHD4-depleted cells with an expression vector for ectopic CHD4, and our results confirmed that the aberrant CTCF binding is direct effect of CHD4 depletion (Fig. 5, Supplementary Fig. S5). Lastly, while

the ChAHP complex study found that ADNP specifically binds to less diverged B2 SINEs, our present data provide direct evidence that CHD4 regulates chromatin accessibility at B2 SINE-embedded aberrant CTCF-binding sites (Group 1 sites), and that CHD4 depletion triggered the overexpression of B2 SINE transcripts in mESCs (Fig. 6, Supplementary Fig. S6). Thus, there are significant differences in the major mechanisms through which cells regulate the CTCF-binding processes that involve ADNP (ChAHP complex) versus CHD4.

Based on our findings, we propose a mechanism wherein the ATP-dependent chromatin remodeler, CHD4, maintains the lower-level chromatin structure (including chromatin accessibility and nucleosome positioning/occupancy) and prevents the aberrant recruitment of CTCF, thereby safeguarding appropriate CTCF binding and associated higher-order TAD organization by suppressing aberrant interactions at TAD interiors (Fig. 7). Our results reveal how the primary structure of chromatin could facilitate higher-order 3D chromatin organization in mESCs. Collectively, our findings provide a novel perspective on the functional role of chromatin remodelers in mESCs, thereby adding an extra layer of information in stem cell biology and greatly facilitating our understanding of how chromatin remodelers function in embryonic stem cells.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

S.H., H.L., and D.L. designed the concepts and experiments. S.H. analyzed most of the NGS data. S.H. and H.L. performed most of the experiments. S.H. performed the Hi-C experiments for –Aux and +Aux cells. A.J.L. performed the Hi-C experiments for Control and Chd4KD cells under the guidance of I.J. S.K.K. performed the GRO-seq experiments under the guidance of T.K.K. S.H. and H.L. wrote the manuscript under the mentorship and technical supervision of G.Y.K. and D.L.

CONFLICT OF INTEREST

The authors have no potential conflicts of interest to disclose.

ORCID

Sungwook Han   https://orcid.org/0000-0003-3944-3266
Hosuk Lee   https://orcid.org/0000-0003-0377-5293
Andrew J. Lee   https://orcid.org/0000-0001-9917-8834
Seung-Kyoon Kim   https://orcid.org/0000-0001-9776-4047
Inkyung Jung   https://orcid.org/0000-0002-5885-2754
Gou Young Koh   https://orcid.org/0000-0002-1231-1485
Tae-Kyung Kim   https://orcid.org/0000-0001-5161-9864
Daeyoup Lee   https://orcid.org/0000-0003-2006-1823

REFERENCES

Acemel, R.D., Maeso, I., and Gomez-Skarmeta, J.L. (2017). Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. Wiley Interdiscip. Rev. Dev. Biol. *6*, e265.

Allen, T.A., Von Kaenel, S., Goodrich, J.A., and Kugel, J.F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. Nat. Struct. Mol. Biol. *11*, 816-821.

Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., et al. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell *145*, 183-197.

Blosser, T.R., Yang, J.G., Stone, M.D., Narlikar, G.J., and Zhuang, X. (2009). Dynamics of nucleosome remodelling by individual ACF complexes. Nature *462*, 1022-1027.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., et al. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. *18*, 1752-1762.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213-1218.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. *109*, 21.29.1-21.29.9.

Choi, W.Y., Hwang, J.H., Cho, A.N., Lee, A.J., Jung, I., Cho, S.W., Kim, L.K., and Kim, Y.J. (2020). NEUROD1 intrinsically initiates differentiation of induced pluripotent stem cells into neural progenitor cells. Mol. Cells *43*, 1011-1022.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Cremer, T. and Cremer, M. (2010). Chromosome territories. Cold Spring Harb. Perspect. Biol. *2*, a003889.

Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res. *19*, 24-32.

Danko, C.G., Hah, N., Luo, X., Martins, A.L., Core, L., Lis, J.T., Siepel, A., and Kraus, W.L. (2013). Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. Mol. Cell *50*, 212-222.

de Dieuleveult, M., Yen, K., Hmitou, I., Depaux, A., Boussouar, F., Bou Dargham, D., Jounier, S., Humbertclaude, H., Ribierre, F., Baulard, C., et al. (2016). Genome-wide nucleosome specificity and function of chromatin remodellers in ES cells. Nature *530*, 113-116.

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.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

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

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., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell 159, 374-387.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95-98.

Espinoza, C.A., Allen, T.A., Hieb, A.R., Kugel, J.F., and Goodrich, J.A. (2004). B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. Nat. Struct. Mol. Biol. 11, 822-829.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J., Heidersbach, A., Ramalho-Santos, J., McManus, M.T., Plath, K., Meshorer, E., et al. (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature 460, 863-868.

Goodman, J.V., Yamada, T., Yang, Y., Kong, L., Wu, D.Y., Zhao, G., Gabel, H.W., and Bonni, A. (2020). The chromatin remodeling enzyme Chd4 regulates genome architecture in the mouse brain. Nat. Commun. 11, 3419.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell 162, 900-910.

Hall, J.A. and Georgel, P.T. (2007). CHD proteins: a diverse family with strong ties. Biochem. Cell Biol. 85, 463-476.

He, C., Sidoli, S., Warneford-Thomson, R., Tatomer, D.C., Wilusz, J.E., Garcia, B.A., and Bonasio, R. (2016). High-resolution mapping of RNA-binding regions in the nuclear proteome of embryonic stem cells. Mol. Cell 64, 416-430.

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

Hendrickson, D., Kelley, D.R., Tenen, D., Bernstein, B., and Rinn, J.L. (2016). Widespread RNA binding by chromatin-associated proteins. Genome Biol. 17, 28.

Heo, J.B. and Sung, S. (2011). Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. Science 331, 76-79.

Ishihara, K., Oshimura, M., and Nakao, M. (2006). CTCF-dependent chromatin insulator is linked to epigenetic remodeling. Mol. Cell 23, 733-742.

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., et al. (2016). 3D chromosome regulatory landscape of human pluripotent cells. Cell Stem Cell 18, 262-275.

Kaaij, L.J.T., Mohn, F., van der Weide, R.H., de Wit, E., and Bühler, M. (2019). The ChAHP complex counteracts chromatin looping at CTCF sites that emerged from SINE expansions in mouse. Cell 178, 1437-1451.e14.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J., et al. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458, 362-366.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 26, 2204-2207.

Kim, S.K., Jung, I., Lee, H., Kang, K., Kim, M., Jeong, K., Kwon, C.S., Han, Y.M., Kim, Y.S., Kim, D., et al. (2012). Human histone H3K79 methyltransferase DOT1L protein [corrected] binds actively transcribing RNA polymerase II to regulate gene expression. J. Biol. Chem. 287, 39698-39709.

Kim, S.K., Lee, H., Han, K., Kim, S.C., Choi, Y., Park, S.W., Bak, G., Lee, Y., Choi, J.K., Kim, T.K., et al. (2014). SET7/9 methylation of the pluripotency factor LIN28A is a nucleolar localization mechanism that blocks let-7 biogenesis in human ESCs. Cell Stem Cell 15, 735-749.

Lam, M.T., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., et al. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature 498, 511-515.

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

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289-293.

Lin, Y.H., Forman-Kay, J.D., and Chan, H.S. (2016). Sequence-specific polyampholyte phase separation in membraneless organelles. Phys. Rev. Lett. 117, 178101.

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.

Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. 24, 800-812.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. U. S. A. 36, 344-355.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495-501.

Meers, M.P., Bryson, T.D., Henikoff, J.G., and Henikoff, S. (2019). Improved CUT&RUN chromatin profiling tools. Elife 8, e46314.

Micucci, J.A., Sperry, E.D., and Martin, D.M. (2015). Chromodomain helicase DNA-binding proteins in stem cells and human developmental diseases. Stem Cells Dev. 24, 917-926.

Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A.P., Kim, H.J., Mittag, T., and Taylor, J.P. (2015). Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. Cell 163, 123-133.

Morris, S.A., Baek, S., Sung, M.H., John, S., Wiench, M., Johnson, T.A., Schiltz, R.L., and Hager, G.L. (2014). Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions. Nat. Struct. Mol. Biol. 21, 73-81.

Narlikar, G.J., Phelan, M.L., and Kingston, R.E. (2001). Generation and interconversion of multiple distinct nucleosomal states as a mechanism for catalyzing chromatin fluidity. Mol. Cell 8, 1219-1230.

Natsume, T., Kiyomitsu, T., Saga, Y., and Kanemaki, M.T. (2016). Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. Cell Rep. 15, 210-218.

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.e22.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381-385.

Ong, C.T. and Corces, V.G. (2014). CTCF: an architectural protein bridging genome topology and function. Nat. Rev. Genet. 15, 234-246.

Ostapcuk, V., Mohn, F., Carl, S.H., Basters, A., Hess, D., Iesmantavicius, V., Lampersberger, L., Flemr, M., Pandey, A., Thomä, N.H., et al. (2018). Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. Nature 557, 739-743.

Owen-Hughes, T., Utley, R.T., Steger, D.J., West, J.M., John, S., Cote, J., Havas, K.M., and Workman, J.L. (1999). Analysis of nucleosome disruption by ATP-driven chromatin remodeling complexes. Methods Mol. Biol. 119,

319-331.

Piovesan, D., Necci, M., Escobedo, N., Monzon, A.M., Hatos, A., Mičetić, I., Quaglia, F., Paladin, L., Ramasamy, P., Dosztányi, Z., et al. (2021). MobiDB: intrinsically disordered proteins in 2021. Nucleic Acids Res. 49(D1), D361-D367.

Piovesan, D., Tabaro, F., Paladin, L., Necci, M., Micetic, I., Camilloni, C., Davey, N., Dosztanyi, Z., Meszaros, B., Monzon, A.M., et al. (2018). MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. Nucleic Acids Res. 46(D1), D471-D476.

Pohl, A. and Beato, M. (2014). bwtool: a tool for bigWig files. Bioinformatics 30, 1618-1619.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665-1680.

Rao, S.S.P., Huang, S.C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin loss eliminates all loop domains. Cell 171, 305-320.e24.

Rippe, K., Schrader, A., Riede, P., Strohner, R., Lehmann, E., and Langst, G. (2007). DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. Proc. Natl. Acad. Sci. U. S. A. 104, 15635-15640.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. 29, 24-26.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics 20, 3246-3248.

Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. U. S. A. 112, E6456-E6465.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Goncalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 148, 335-348.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L., et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature 551, 51-56.

Segal, E. and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr. Opin. Struct. Biol. 19, 65-71.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 16, 259.

Shazman, S., Celniker, G., Haber, O., Glaser, F., and Mandel-Gutfreund, Y. (2007). Patch Finder Plus (PFplus): a web server for extracting and displaying positive electrostatic patches on protein surfaces. Nucleic Acids Res. 35(Web Server issue), W526-W530.

Shazman, S. and Mandel-Gutfreund, Y. (2008). Classifying RNA-binding proteins based on electrostatic properties. PLoS Comput. Biol. 4, e1000146.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. 38, 1348-1354.

Skene, P.J. and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife 6, e21856.

Skene, P.J., Hernandez, A.E., Groudine, M., and Henikoff, S. (2014). The nucleosomal barrier to promoter escape by RNA polymerase II is overcome by the chromatin remodeler Chd1. Elife 3, e02042.

Som, A., Harder, C., Greber, B., Siatkowski, M., Paudel, Y., Warsow, G., Cap, C., Scholer, H., and Fuellen, G. (2010). The PluriNetWork: an electronic representation of the network underlying pluripotency in mouse, and its applications. PLoS One 5, e15165.

Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. Genes Dev. 20, 2349-2354.

Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G.L., Cheutin, T., Bonev, B., Nora, E.P., Bruneau, B.G., Bantignies, F., and Cavalli, G. (2020). Regulation of single-cell genome organization into TADs and chromatin nanodomains. Nat. Genet. 52, 1151-1157.

Tillo, D. and Hughes, T.R. (2009). G+C content dominates intrinsic nucleosome occupancy. BMC Bioinformatics 10, 442.

Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. Nature 474, 516-520.

van der Weide, R.H., van den Brand, T., Haarhuis, J.H.I., Teunissen, H., Rowland, B.D., and de Wit, E. (2021). Hi-C analyses with GENOVA: a case study with cohesin variants. NAR Genom. Bioinform. 3, lqab040.

van Vugt, J.J., de Jager, M., Murawska, M., Brehm, A., van Noort, J., and Logie, C. (2009). Multiple aspects of ATP-dependent nucleosome translocation by RSC and Mi-2 are directed by the underlying DNA sequence. PLoS One 4, e6345.

Varshney, D., Vavrova-Anderson, J., Oler, A.J., Cowling, V.H., Cairns, B.R., and White, R.J. (2015). SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. Nat. Commun. 6, 6569.

Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. 10, 1297-1309.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451, 796-801.

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

Wiechens, N., Singh, V., Gkikopoulos, T., Schofield, P., Rocha, S., and Owen-Hughes, T. (2016). The chromatin remodelling enzymes SNF2H and SNF2L position nucleosomes adjacent to CTCF and other transcription factors. PLoS Genet. 12, e1005940.

Zhan, Y., Mariani, L., Barozzi, I., Schulz, E.G., Blüthgen, N., Stadler, M., Tiana, G., and Giorgetti, L. (2017). Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. Genome Res. 27, 479-490.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.

Zhao, H., Han, Z., Liu, X., Gu, J., Tang, F., Wei, G., and Jin, Y. (2017). The chromatin remodeler Chd4 maintains embryonic stem cell identity by controlling pluripotency- and differentiation-associated genes. J. Biol. Chem. 292, 8507-8519.