ADNP modulates SINE B2-derived CTCF-binding sites during blastocyst formation in mice

Wen Wang,[1,2,8] Rui Gao,[1,8] Dongxu Yang,[1,2,8] Mingli Ma,[1,2,8] Ruge Zang,[1] Xiangxiu Wang,[3] Chuan Chen,[4] Xiaochen Kou,[1] Yanhong Zhao,[1] Jiayu Chen,[1,5] Xuelian Liu,[2,6,7] Jiaxu Lu,[2] Ben Xu,[1] Juntao Liu,[2] Yanxin Huang,[1] Chaoqun Chen,[2] Hong Wang,[1] Shaorong Gao,[1,2,5] Yong Zhang,[2,5] and Yawei Gao[2,5]

[1]Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, Shanghai Key Laboratory of Signaling and Disease Research, Frontier Science Center for Stem Cell Research, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China; [2]State Key Laboratory of Cardiology and Medical Innovation Center, Institute for Regenerative Medicine, Shanghai East Hospital, Frontier Science Center for Stem Cell Research, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China; [3]Key Laboratory of Biorheological and Technology of Ministry of Education, State and Local Joint Engineering Laboratory for Vascular Implants, Modern Life Science Experiment Teaching Center at Bioengineering College of Chongqing University, Chongqing 400030, China; [4]Women's Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang 310006, China; [5]Shanghai Institute of Stem Cell Research and Clinical Translation, Shanghai 200120, China; [6]Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China; [7]University of Chinese Academy of Sciences, Beijing 100049, China

CTCF is crucial for chromatin structure and transcription regulation in early embryonic development. However, the kinetics of CTCF chromatin occupation in preimplantation embryos have remained unclear. In this study, we used CUT&RUN technology to investigate CTCF occupancy in mouse preimplantation development. Our findings revealed that CTCF begins binding to the genome prior to zygotic genome activation (ZGA), with a preference for CTCF-anchored chromatin loops. Although the majority of CTCF occupancy is consistently maintained, we identified a specific set of binding sites enriched in the mouse-specific short interspersed element (SINE) family B2 that are restricted to the cleavage stages. Notably, we discovered that the neuroprotective protein ADNP counteracts the stable association of CTCF at SINE B2-derived CTCF-binding sites. Knockout of *Adnp* in the zygote led to impaired CTCF binding signal recovery, failed deposition of H3K9me3, and transcriptional derepression of SINE B2 during the morula-to-blastocyst transition, which further led to unfaithful cell differentiation in embryos around implantation. Our analysis highlights an ADNP-dependent restriction of CTCF binding during cell differentiation in preimplantation embryos. Furthermore, our findings shed light on the functional importance of transposable elements (TEs) in promoting genetic innovation and actively shaping the early embryo developmental process specific to mammals.

[*Keywords*: ADNP; blastocyst; CTCF; short interspersed element (SINE) family B2]

Supplemental material is available for this article.

Received September 18, 2023; revised version accepted February 20, 2024.

CCCTC-binding factor (CTCF) is a highly conserved, ubiquitously expressed DNA-binding protein with multivalent properties, including transcriptional regulation (Filippova et al. 1996; Ling et al. 2006), chromatin insulation (Bell et al. 1999; Cuddapah et al. 2009), genomic imprinting (Hark et al. 2000), and X-chromosome inactivation (Chao et al. 2002). Recent evidence has also revealed CTCF's role in organizing the three-dimensional (3D) chromatin architecture through the formation of short-range chromatin loops (Handoko et al. 2011; Dixon et al. 2012) or the novel phase separation behavior for long-range interactions (Lee et al. 2022; Wei et al. 2022). Composed of 11 zinc finger (ZF) domains, CTCF recognizes DNA sequence diversity through the deployment of its distinct ZF clusters (Nakahashi et al. 2013). Sequence

[8]**These authors contributed equally to this work.**
Corresponding authors: gaoshaorong@tongji.edu.cn, gaoyawei@tongji.edu.cn, yzhang@tongji.edu.cn
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.351189.123.

motif analysis has identified two major parts of the CTCF motif: the core and upstream bound by ZF3–7 and ZF9–11, respectively. The core motif is present in most CTCF-binding sites (CBSs) (Rhee and Pugh 2011), and the methylation status of the cytosine located inside the core motif can influence the binding affinity of CTCF (Bell and Felsenfeld 2000; Hashimoto et al. 2017).

Many studies have shown that CTCF plays a role during early mammalian development (Carmona-Aldana et al. 2018). Upon fertilization, terminally differentiated gametes are reprogrammed into totipotent cells through zygotic genome activation (ZGA) at the maternal-to-zygotic transition (MZT) (Lee et al. 2014; Schulz and Harrison 2019). After the first lineage segregation, pluripotent embryonic and extraembryonic trophectoderm lineages are generated (Hackett and Surani 2014). Appropriate CTCF occupancy in preimplantation embryos is required for successful cell lineage segregation, as the absence of CTCF results in embryonic lethality in blastocysts (Wan et al. 2008; Chen et al. 2019; Andreu et al. 2022). In our previous study, we discovered nucleosome depletion regions (NDRs) around CTCF motif sites in both pronuclei, which suggested that CTCF may bind to the mouse genome prior to the timing of ZGA (Wang et al. 2022a). However, information about the dynamics of CTCF-binding sites (CBSs) during early embryogenesis, particularly for preimplantation embryos, is currently unavailable.

Transposable elements (TEs) are a rich source of diverse *cis*-regulatory regions for mammalian transcription regulation, including promoters, enhancers, and transcription factor binding sites (Sundaram et al. 2014; Trizzino et al. 2017; Hermant and Torres-Padilla 2021; Fueyo et al. 2022). In humans, only 11% of CBSs are derived from TEs (Kunarso et al. 2010). In contrast, in mice, TEs contribute to 28% of CBSs. Specifically, short interspersed element (SINE) B2, a type of TE, constitutes nearly a quarter of these CBSs (Bourque et al. 2008). Despite this difference, the 3D chromatin architectures are highly conserved between humans and mice (Harmston et al. 2017). Two models have been proposed to explain this discrepancy. The first model suggests that species-specific CTCF motif sites are suppressed by the activity-dependent neuroprotector homeobox protein (ADNP) and its cofactors to prevent unwanted CBSs (Kaaij et al. 2019). The second model proposes that newly derived CBSs located near existing CBSs can act as a backup and cause no significant changes in 3D chromatin structures (Choudhary et al. 2020).

Here, we used the recently developed cleavage under targets and release using nuclease (CUT&RUN) (Skene and Henikoff 2017) technique to produce CTCF binding profiles throughout early embryogenesis. We examined the intrinsic sequence features and epigenetic factors that impact CTCF rebinding events following fertilization. CTCF anchor sites were preferentially established during embryogenesis. Furthermore, we identified >1000 cleavage-specific CBSs (cs-CBSs), which are highly enriched for SINE B2 repetitive elements, by comparing CBSs between eight-cell and blastocyst stage embryos.

Results

CTCF rebinds to the genome prior to ZGA

To detect the dynamic of CBSs in early embryos, we used CUT&RUN technology to capture the genomic occupation of CTCF using low-input cells. We first generated CTCF CUT&RUN data in mouse embryonic stem cells (mESCs) using 1×10^5 cells with two biological replicates (Supplemental Table S1). The high reproducibility of CTCF CUT&RUN data was demonstrated by the highly correlated signal on potential CBSs (Pearson's correlation coefficients = 0.94) (Fig. 1A; Supplemental Fig. S1A–C; Supplemental Tables S2, S3; for a potential CBS definition, see the Materials and Methods). Next, we generated CTCF occupancy data using 5×10^3 and 1×10^3 cells with two highly reproducible biological replicates each (Pearson's correlation coefficients $\geq$ 0.97) (Supplemental Fig. S1B,C; Supplemental Table S1). The signal of merged replicates from different input materials highly correlated on potential CBSs (Pearson's correlation coefficients $\geq$ 0.89) (Supplemental Fig. S1D). We further identified CBSs through a peak calling process on the merged signal using MACS (version 2.1.3; Zhang et al. 2008; Feng et al. 2012). Most of the CBSs were identified in all three different input samples (Supplemental Fig. S1E; see the Materials and Methods for details). Moreover, we performed de novo motif finding using MEME (version 5.0.5; Bailey et al. 2006). The CTCF motif was ranked as the top motif for each sample (Supplemental Fig. S1F; see the Materials and Methods for details). These results suggest that CUT&RUN can produce high-quality CTCF binding profiles using low-input materials.

Subsequently, we generated CTCF occupancy data in mouse gametes and early embryos with two highly correlated biological replicates at each stage (Supplemental Figs. S1B,C, S2A; Supplemental Table S1). In gametes, we identified thousands of CBSs that exist in the germinal vesicle (GV) but are absent in metaphase II (MII) oocytes (Fig. 1B,C). This phenomenon can be attributed to the differences in chromatin status across the cell cycle (Oomen et al. 2019). Following fertilization, we observed a resurgence of CTCF binding signal at the pronucleus (PN)-3 stage (6 h postfertilization [hpf]), which is consistent with the well-organized NDRs around CTCF motif sites in both pronuclei observed in our previous work (Fig. 1B, C; Adenot et al. 1997; Wang et al. 2022a). To confirm these CTCF rebinding events, we analyzed the dynamics of NDRs around PN-3 CBSs using our previous ultralow-input MNase-seq (ULI-MNase-seq) data obtained from pronucleus stage embryos (Supplemental Table S2; Wang et al. 2022a). Our analysis revealed more enhanced and evident NDRs around PN-3 CBSs compared with PN-3 nonbinding CTCF motif sites in both pronuclei no later than 6 hpf, demonstrating that the reconstruction of CTCF binding after fertilization is reliable in both pronuclei (Fig. 1D). Immunostaining analysis of CTCF protein was further checked from GV oocytes to two-cell stage embryos to confirm the protein level increase in pronuclei after fertilization (Supplemental Fig. S2B,C). This swift

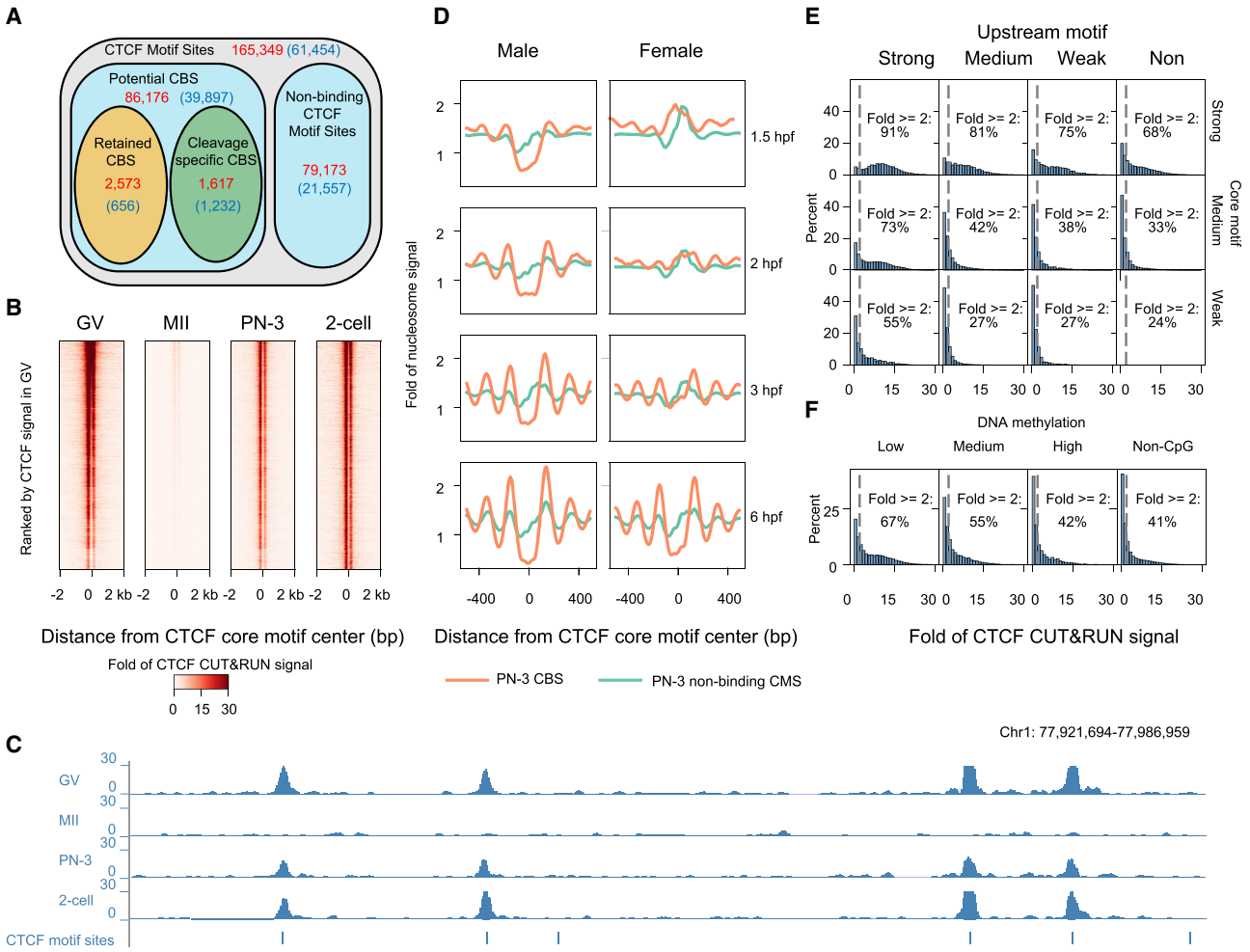

Figure 1. CTCF rebinds to the genome upon fertilization. (*A*) Venn diagrams showing the numbers (original number [red] and filtered with SINE B2 elements [blue]) and relationships between various kinds of CTCF-binding sites based on CTCF binding kinetics. (*B*) Heat maps showing CTCF CUT&RUN signal in GVs, MII oocytes, and PN-3 and two-cell embryos. CTCF binding disappeared at the MII stage and was largely recovered at the PN-3 stage. (*C*) A snapshot of the browser view showing CTCF CUT&RUN signal in the indicated region. (*D*) Nucleosome profiles around CTCF motif sites at each indicated PN stage. Nucleosome depletion regions (NDRs) were established at both pronuclei no later than 3 h postfertilization (hpf). (*E*,*F*) Histograms showing CTCF CUT&RUN signal in PN-3 embryos at CTCF motif sites grouped by motif strength (*E*) or DNA methylation status (*F*). The motif strength score of a CTCF-binding site was partitioned into strong (top 25%), medium, and weak (bottom 25%) according to the motif occurrence within the site. (GV) Germinal vesicle, (MII) metaphase II, (PN) pronucleus, (CMS) CTCF motif site.

resurgence of CBSs following fertilization suggests that the binding of CTCF on chromatin is quite ubiquitous and fundamental in early embryos.

Previous studies have shown that CTCF binding can be influenced by multiple factors (Bell and Felsenfeld 2000; Rhee and Pugh 2011; Nakahashi et al. 2013). To determine the major factors involved in the CTCF rebinding process at the PN-3 stage, we first categorized the potential CBSs into two groups—PN-3 CBSs and PN3 nonbinding CTCF motif sites—according to CTCF binding status in the PN-3 stage and measured various genetic features, including core and upstream motif scores, GC content, and CpG ratio, as well as epigenetic features such as DNA methylation, H3K4me3, H3K27me3, and H3K9me3 for each CBS group. We found that the core motif score shows the highest predictive ability (AUROC = 0.73) (Supplemental Fig. S3A). In addition, 91% of the potential CBSs with the highest scores on both the core and upstream motifs (upper quartile) exhibited a CTCF binding signal (fold change of the CTCF CUT&RUN signal >2) at the PN-3 stage (Fig. 1E). Conversely, the percentage reduced to 24% when the upstream motif was absent and DNA sequences had low core motif scores (lower quartile) (Fig. 1E). Furthermore, lower DNA methylation levels in the core motif led to a higher percentage of CTCF binding signal present at the PN-3 stage (Fig. 1F). The potential CBSs lacking cytosine–phosphate–guanine (CpG) dinucleotides in their core motifs showed the lowest percentage of CTCF binding signal at the PN-3 stage (Fig. 1F). Additionally, we observed a positive correlation between the PN-3 CBSs and H3K4me3, while a negative correlation was found between H3K9me3, H3K27me3, and the PN-3 CBSs (Supplemental

Fig. S3B–D). These results demonstrate that reconstruction of CTCF binding is influenced by both intrinsic sequence and epigenetic features.

CTCF anchor sites preferentially reconstruct after fertilization

After fertilization, 3D genome structures are re-established, and the insulator function of these structures is primarily governed by CTCF (Guo et al. 2015; Du et al. 2017). To investigate the maturation of these insulators, we defined CTCF anchor sites at each stage as the intersection of CBSs at each stage and mESC loop anchors (Supplemental Fig. S4A; Weintraub et al. 2017). We computed the interaction probabilities of embryonic Hi-C data around the CTCF anchor sites in mESCs (see the Materials and Methods for details). The insulator function of CTCF anchor sites was found to increase at the eight-cell stage, as revealed by aggregated Hi-C signal analysis (Fig. 2A). Next, we investigated whether there were any discrepancies between CTCF anchor sites and non-CTCF anchor sites before the eight-cell stage. As expected, we observed that the rebinding of CTCF is established earlier on CTCF anchor sites than on non-CTCF anchor sites (Fig. 2B,C), and the CTCF anchor sites exhibited a greater binding signal at each stage (Supplemental Fig. S4B). To elucidate the underlying reason driving the preference, we first assessed the CTCF motif scores for the above two anchor site groups and found that the core motif score on CTCF anchor sites is significantly higher than on non-CTCF anchor sites (Fig. 2D).

As insulators are characterized by multiple epigenetic features (Heidari et al. 2014), we next analyzed the dynamics of multiple epigenetic profiles at CTCF anchor sites and non-CTCF anchor sites, including DNase-seq; H3K4me3, H3K9me3, and H3K27me3 signal; and DNA methylation level. Our results showed that only the H3K4me3 signal can clearly distinguish CTCF anchor sites and non-CTCF anchor sites before the eight-cell stage (Fig. 2E; Supplemental Fig. S4C–E). In summary, these findings suggest that the core motif is closely related with the preferential reconstruction on CTCF anchor sites, and H3K4me3 may correlate with the reconstruction and maturation of CTCF-mediated insulators.

Cleavage-specific CBSs are derived from SINE B2

Although most CTCF-binding sites remain stable during the two-cell to blastocyst stage (Fig. 3A; Supplemental Fig. S5A), it has been revealed that loss of CTCF affects the efficiency of morula-to-blastocyst transition (Wan et al. 2008; Andreu et al. 2022). We then asked whether unique changes on CBSs occur during the first lineage segregation. We compared CBSs in the eight-cell embryos, inner cell mass (ICM), and trophectoderm at the blastocyst stage and found thousands of CBSs lost in both the ICM and trophectoderm cells (Fig. 3B,C). We defined this specific set of binding sites, which are restricted to the cleavage stages, as cleavage-specific CBSs (cs-CBSs), and defined the consistently maintained CBSs during the eight-cell to blastocyst stage as reserved CBSs (r-CBSs) (Fig. 3B,C). To better validate the chromatin changes around these CBSs, we generated nucleosome occupancy data using ULI-MNase-seq (Wang et al. 2022a) at the corresponding stages to sample-harvest for CTCF CUT&RUN-seq (Supplemental Fig. S5B–D; Supplemental Table S1). Intriguingly, we found a 12-bp shift upstream in the lowest point of NDRs around the cs-CBSs, while the r-CBSs had a subtler 4-bp shift upstream (Fig. 3D), suggesting that the TF binding feature changes upon loss of CTCF binding on cs-CBSs. To verify whether the occupation of CTCF on cs-CBSs uniquely happens in cleavage embryos, we collected 349 publicly available mouse CTCF ChIP-seq samples from the Cistrome data browser (Supplemental Table S2; see the Materials and Methods for details; Mei et al. 2017; Zheng et al. 2019). CTCF binding events were rarely found in these cs-CBSs from all public mouse samples, except a weak but significant recovery in mESCs was consistent with the reported more frequent B2-derived CTCF binding in mESCs (Ichiyanagi et al. 2021). This observation indicates that the binding loss on cs-CBSs was stable in various differentiated tissues but partly reversed upon pluripotency status recovery (Fig. 3E; Supplemental Fig. S5E,F). Next, we reviewed the binding kinetics of these two sets of CBSs. As shown, over half of r-CBSs obtained the CTCF binding in the PN-3 stage, and this percentage was much higher than that of cs-CBSs until the eight-cell stage (Fig. 3F). Taken together, these results provided evidence for the existence of a specific set of cs-CBSs that gain CTCF binding after ZGA but lose binding upon cell differentiation at the blastocyst stage.

We further investigated the influence factors that correlated with the transient CTCF occupancy on cs-CBSs. Upon assessment of the phastCons scores of these CBSs, we observed that the cs-CBSs exhibited lower sequence convergence in comparison with the r-CBSs (Supplemental Fig. S5G). This temporary CTCF binding status echoes the fact that in many species, the CTCF motifs can emerge from TEs, especially SINEs (Bourque et al. 2008; Schmidt et al. 2012), which was restricted by repressive histone modifications or competitive TFs (Kaaij et al. 2019; Gualdrini et al. 2022). Therefore, we examined the enrichment of various kinds of TEs in cs-CBSs and r-CBSs. Interestingly, the majority (76.2%) of cs-CBSs were derived from SINE B2, whereas only a minority (25.5%) of r-CBSs were derived from B2 (Fig. 3C,G; Supplemental Fig. S5H). Specifically, we found that cs-CBSs were enriched with SINE B2 when compared with r-CBSs, especially the mouse-specific B2 subfamilies (B2_Mm1a: 4.5% vs. 3.4%, B2_Mm1t: 15.7% vs. 5.6%, and B2_Mm2: 40.6% vs. 24.5%) (Fig. 3C,G; Supplemental Fig. S5H). These results suggest that cleavage-specific CBSs are mainly derived from SINE B2 elements.

ADNP suppresses SINE B2-derived cleavage-specific CBSs

Previous work showed that nearly 40% of CTCF-binding sites in the mouse genome are derived from transposable elements (Sundaram et al. 2014), and CTCF binding in

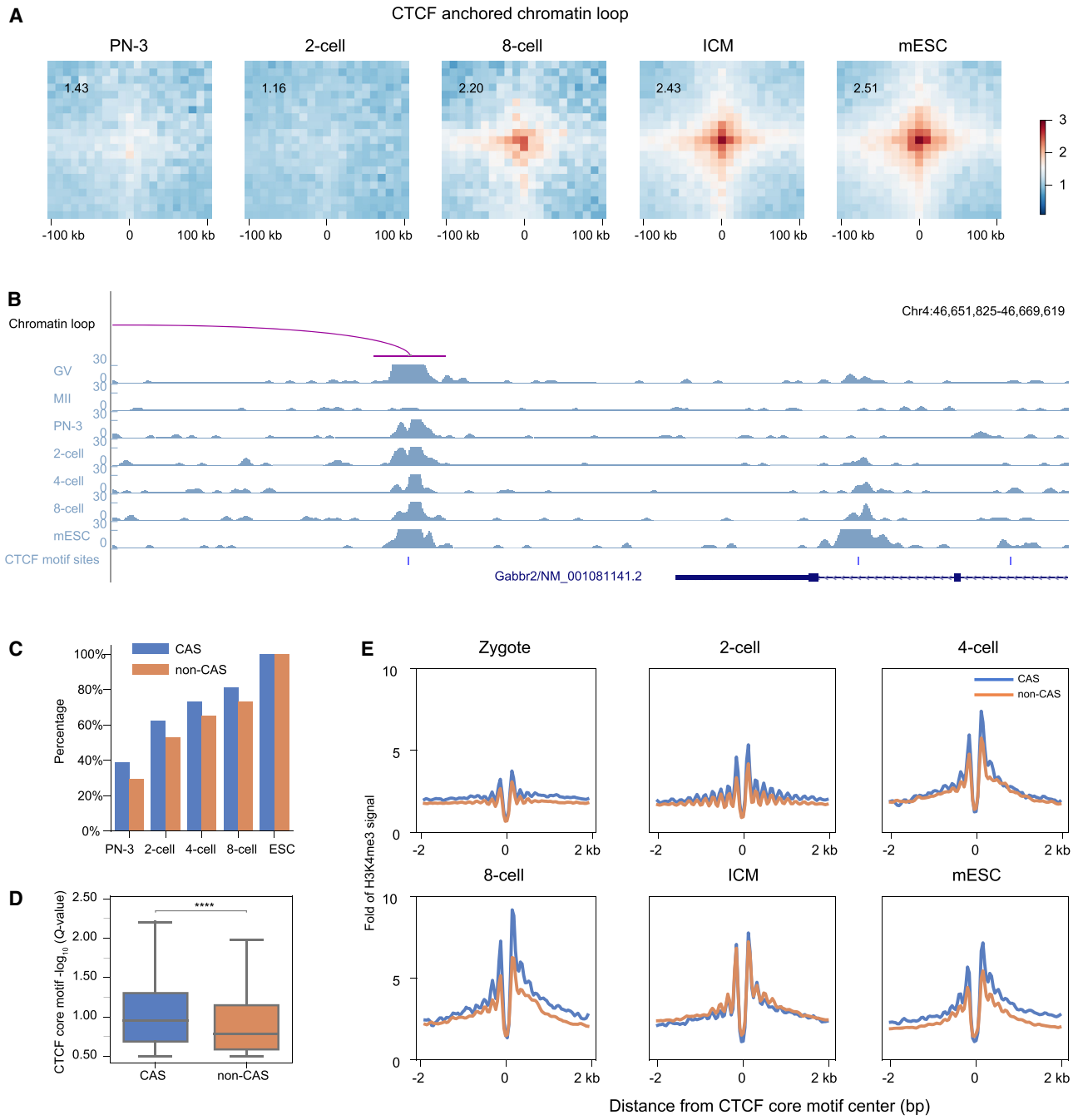

Figure 2. CTCF anchor sites preferentially reconstruct. (*A*) Heat maps showing the normalized average interaction frequencies at each embryonic stage for CTCF anchor sites in each stage. The chromatin interaction was established from the eight-cell stage. (*B*) A snapshot of the browser view showing CUT&RUN signal at the indicated CTCF anchor sites and non-CTCF anchor sites. (*C*) Bar plots showing the binding percentage of distinct CTCF anchor site classes in mESCs at each embryonic stage. CTCF anchor sites were established earlier than non-CTCF anchor sites. (*D*) Box plots showing that CTCF anchor sites exhibit a stronger CTCF core motif than non-CTCF anchor sites. (****) *P*-value < 0.0001 by two-sided Mann–Whitney *U*-test. (*E*) Line plots showing H3K4me3 signal levels around distinct CTCF anchor sites. CTCF anchor sites display higher H3K4me3 signals compared with non-CTCF anchor sites. (mESC) Mouse embryonic stem cell, (ICM) inner cell mass, (CAS) CTCF anchor site, (GV) germinal vesicle, (MII) metaphase II, (PN) pronucleus.

mammals is often associated with species-specific SINE B2 expansions (Schmidt et al. 2012; Thybert et al. 2018). However, the detailed mechanisms enforcing the usage of these novel CTCF-binding sites in early embryos are largely unknown. We identified a total of 61,454 SINE B2-derived CTCF motif sites and classed them into three groups based on their distinct usage in preimplantation embryos. Among them, 656 sites showed strong binding

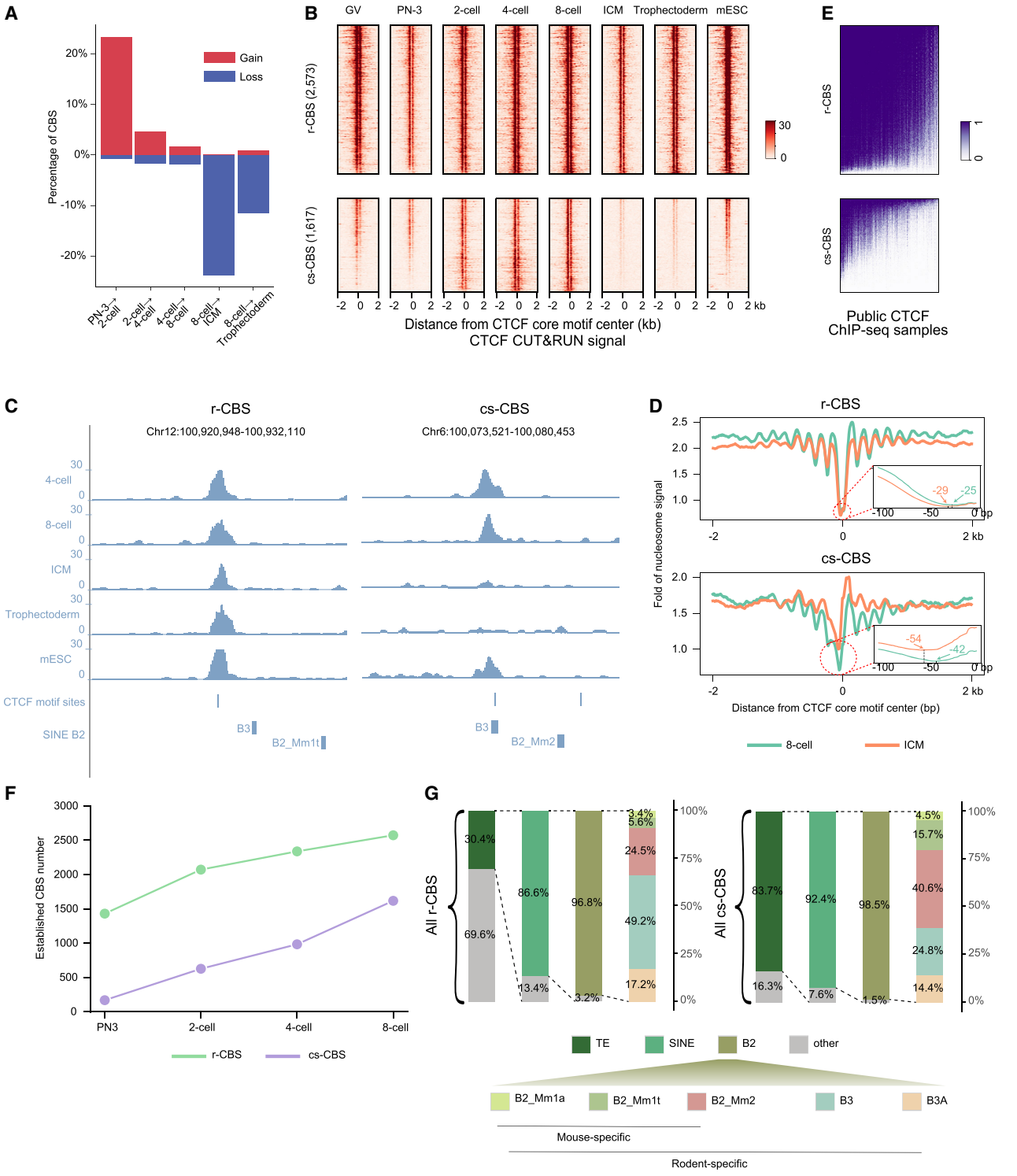

Figure 3. SINE B2 repeats introduce cleavage-specific CBSs. (*A*) Dynamics of CTCF-binding sites (CBSs) during the indicated embryo stages. (*B*) Heat maps showing CTCF CUT&RUN signal at r-CBSs and cs-CBSs. (*C*) A snapshot of the browser view showing the indicated r-CBSs (*left*) and cs-CBSs (*right*). (*D*) Nucleosome profiles around r-CBSs (*top*) and cs-CBSs (*bottom*) in eight-cell embryos and the ICM. The lowest point of NDRs around cs-CBSs shifted to upstream. (*E*) Heat maps showing the existence of CTCF signals in hundreds of publicly available mouse ChIP-seq samples. (*F*) Line plots showing the step-wise establishment of distinct CBS classes. The majority of cs-CBSs were established during cleavage stages. (*G*) Stacked bar plots showing enrichment of each repeat family in r-CBSs (*left*) and cs-CBSs (*right*). We found that 25.5% (30.4% × 86.6% × 96.8%) of r-CBSs and 76.2% (83.7% × 92.4% × 98.5%) of cs-CBSs were derived from B2 elements. (ICM) Inner cell mass, (TE) transposable element, (CMS) CTCF motif site. We used the TE annotation from the UCSC table browser (http://genome.ucsc.edu/cgi-bin/hgTables).

signals throughout the cleavage stage and ICM/trophectoderm and were defined as B2-derived r-CBSs. Additionally, 1232 sites exhibited loss of CTCF binding during blastocyst differentiation and were defined as B2-derived cs-CBSs. The majority of sites (21,260) showed no CTCF binding signals across embryonic stages and all publicly available samples and were defined as B2-derived nonbinding CTCF motif sites.

Subsequently, we explored potential factors contributing to the differential formation of these motif sites. We first assessed the sequence convergence among these three subgroups of B2-derived CTCF motif sites. B2-derived cs-CBSs showed a significantly lower degree of convergence compared with B2-derived r-CBSs, and B2-derived nonbinding CTCF motif sites exhibited the lowest sequence convergence (Fig. 4A). Furthermore, the motif strength of these B2-derived CTCF motif sites exhibited a similar trend, with B2-derived nonbinding CTCF motif sites displaying the weakest motif sequence, and B2-derived r-CBSs displaying the strongest motif sequence (Fig. 4B; Supplemental Fig. S6A). Additionally, more than a quarter of B2-derived nonbinding CTCF motif sites were located in the inactive region B compartments, while 14% of B2-derived cs-CBSs and <5% of B2-derived r-CBSs were found in B compartments (Fig. 4C). These findings indicate that newly derived B2 SINE sequences with poor CTCF motif sequences were more likely to be retained but unused in the genome. Since sequences in A/B compartments contain different histone modifications (Rowley and Corces 2018), and H3K9 trimethylation in active chromatin was reported

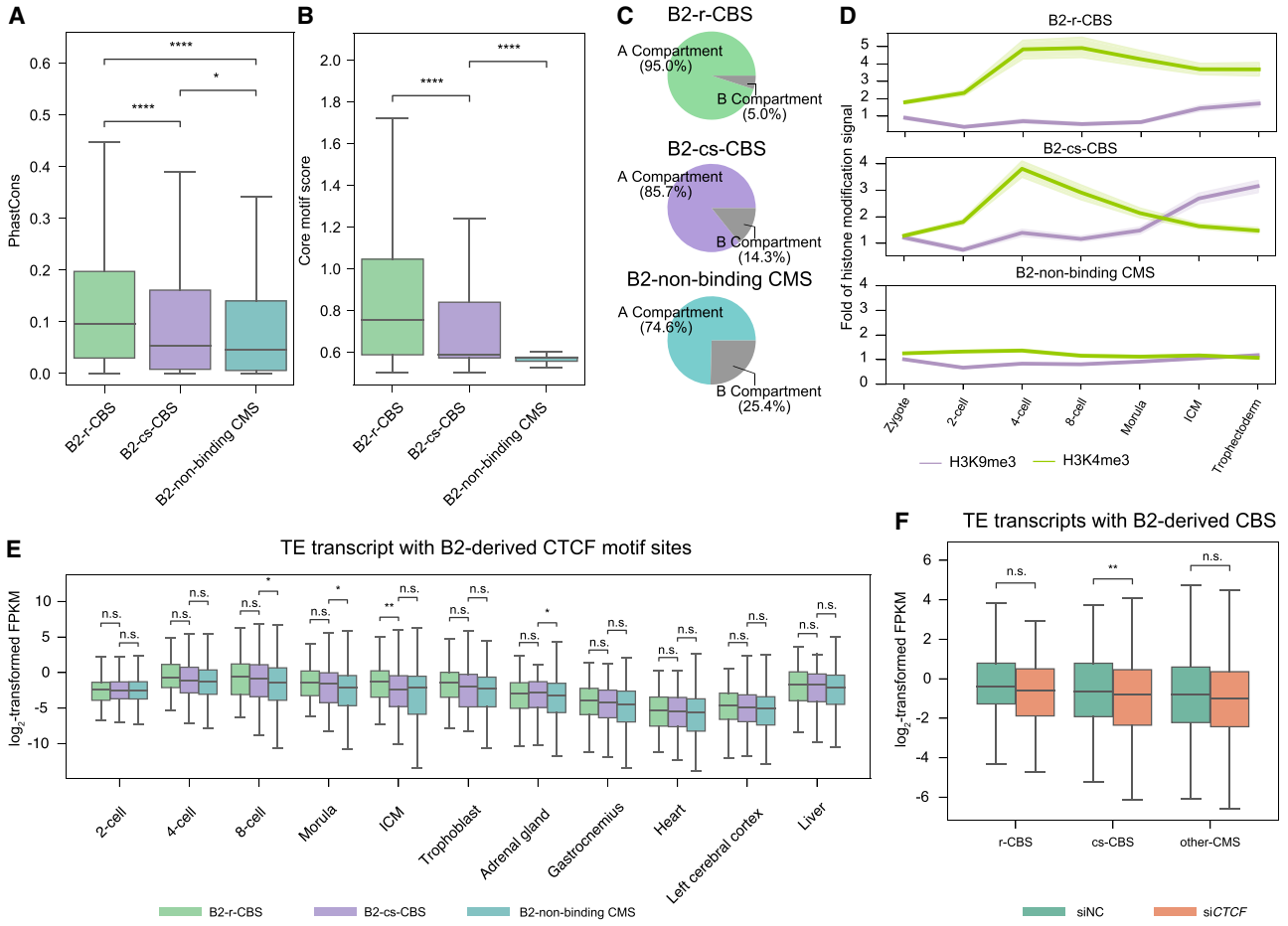

Figure 4. ADNP alters the CTCF binding at B2-derived cs-CSBs in mESCs. (*A*,*B*) Box plots showing the adjacent phastCons score (*A*) and core motif strength (*B*) of three distinct B2-derived CMS classes. B2-derived r-CBSs show the highest sequence convergence and strongest core motif, whereas B2-derived nonbinding CMSs show the lowest sequence convergence and weakest core motif. Motif strength was determined by $-\log_{10}$ (*Q*-value). *P*-values were calculated by two-sided Mann–Whitney *U*-test. (*C*) Pie charts showing the percentage of distinct B2-derived CMSs in the A/B compartment. (*D*) Line plots showing active H3K4me3 and repressive H3K9me3 signals at three distinct B2-derived CMSs across the entire embryogenesis. B2-derived nonbinding CMSs exhibit moderate enrichment of the indicated histone marks. B2-derived r-CBSs display sustained enrichment of active H3K4me3 modifications, whereas B2-derived cs-CBSs show a transition from active H3K4me3 to repressive H3K9me3 enrichment. (*E*) Box plots showing the expression levels of B2 repeats in distinct B2-derived CMSs at each embryonic stage and in adult tissue. *P*-values were calculated by two-sided Mann–Whitney *U*-test. (*F*) Box plots showing the TE transcript expression with B2-derived CBSs in *CTCF* knockdown (si*CTCF*) and control (siNC) conditions. *P*-values were calculated by two-sided paired Student's *t*-test. (r-CBS) Reserved CBS, (cs-CBS) cleavage-specific CBS, (CMS) CTCF motif sites, (TE) transposable element. (*) *P*-value < 0.05, (**) *P*-value < 0.01, (****) *P*-value < 0.0001, (n.s.) not significant.

to restrict the usage of functional CTCF sites in SINE B2 elements, we reanalyzed publicly available histone modification data (Liu et al. 2016; Wang et al. 2018). The B2-derived nonbinding CTCF motif sites generally exhibited moderate enrichment of any specific histone marks, while B2-derived r-CBSs displayed continuous enrichment of active H3K4me3 instead of repressive H3K9me3 modifications. Importantly, B2-derived cs-CBSs showed a transition from active H3K4me3 enrichment to repressive H3K9me3 enrichment, which coincided with the loss of CTCF binding across the blastocyst formation (Fig. 4D; Supplemental Fig. S6B). We also examined DNA methylation levels during the morula-to-blastocyst transition. However, the consistent demethylation observed in all three subgroups indicates that DNA methylation status is not responsible for the loss of CTCF binding in cs-CBSs (Supplemental Fig. S6C). These results indicate that the loss of CTCF binding may correlate with histone modification rather than with DNA methylation changes in B2-derived cs-CBSs during the first wave of embryonic cell differentiation.

We also quantified the transcription levels of TEs contained within three different B2-derived CTCF motif site subgroups (Supplemental Table S4; see the Materials and Methods for details; Wang et al. 2018) and observed a significant up-regulation of SINE B2 in both B2-derived r-CBSs and B2-derived cs-CBSs compared with that in the B2-derived nonbinding CTCF motif sites at both the eight-cell and morula stages (Fig. 4E). However, as embryos differentiated into the ICM and trophectoderm lineages, the expression level of SINE B2 in B2-derived cs-CBSs decreased to a level comparable with that of B2-derived nonbinding CTCF motif sites, which was significantly lower than the expression of B2 in B2-derived r-CBSs (Fig. 4E). In addition, there were no expression differences between those three groups of B2 elements in adult tissues (data downloaded from the ENCODE portal) (Supplemental Table S4) that lost the CTCF binding on cs-CBSs (Luo et al. 2020; Kagda et al. 2023).

To further verify the functional importance of CTCF for TE transcription in preimplantation embryos, we injected siRNA of *CTCF* into zygotes to interrupt the function of CTCF. Consistent with published observations, the depletion of CTCF dramatically inhibited the formation of blastocysts at E3.5 (Andreu et al. 2022) and the proper activation of embryonic development-related genes in eight-cell stage embryos (Supplemental Fig. 6D–H). Additionally, the expression of TEs in cs-CBSs exhibited a notable decrease following *CTCF* depletion (Fig. 4F).

These findings demonstrate that loss of CTCF binding in B2-derived cs-CBSs during embryonic cell differentiation is closely associated with the establishment of H3K9me3 modifications and the decreased transcriptional activity of corresponding TEs but is not correlated with DNA methylation status.

We then investigated the suppression mechanism of B2-derived cs-CBSs in the blastocyst and noticed that the ChAHP complex was reported to counteract chromatin looping at SINE-derived CTCF sites in mouse cell lines (Kaaij et al. 2019). In the ChAHP complex, ADNP is known to compete with CTCF especially at the motif sites in younger SINE elements. We reanalyzed publicly available mESC ADNP and CTCF ChIP-seq data (Kaaij et al. 2019) and found that the three subgroups of B2-derived CBSs displayed different ADNP signal characteristics. Both B2-derived cs-CBSs and B2-derived r-CBSs showed more ADNP signal enrichment compared with B2-derived nonbinding CTCF motif sites. However, only B2-derived cs-CBSs showed an increased CTCF signal upon *Adnp* depletion in mESCs (Supplemental Fig. S6B, I,J). Since ADNP can interact with HP1 and establish heterochromatin nanodomains (HNDs) (Ostapcuk et al. 2018; Thorn et al. 2022), we propose that ADNP may also be related to the increased H3K9me3 levels at both B2-derived cs-CBSs and r-CBSs during the first lineage segregation (Fig. 4D; Supplemental Fig. S6B). Subsequently, we designed functional experiments to demonstrate whether ADNP governs the dynamic binding of CTCF to B2-derived cs-CBSs in early embryos.

Up-regulated ADNP reshapes the CTCF binding landscape

To elucidate whether ADNP is involved in the loss of CTCF binding at B2-derived cs-CBSs in the blastocyst stage, we first checked the expression pattern of ADNP during embryogenesis. Interestingly, we found that the highest RNA level of *Adnp* appears in the eight-cell embryos and ICM/trophectoderm cells, which is consistent with the length change of the broad H3K4me3 domain in the *Adnp* promoter region (Supplemental Fig. S7A–C). However, based on our pervious proteomic data of mouse embryos, the accumulation of ADNP proteins was observed only in blastocysts (Supplemental Fig. S7D; Gao et al. 2017). Our immunostaining analyses against the nucleation timing of ADNP during embryogenesis also confirmed that ADNP is retained in the cytoplasm until the morula stage (Supplemental Fig. S7E). All these results indicate that the ADNP protein is highly expressed during morula-to-blastocyst transition and is potentially involved in nuclear regulation.

To further investigate the role of ADNP in early embryos, we coinjected the multi-sgRNAs for *Adnp* and *Cas9* mRNA into the zygotes to generate *Adnp*-depleted embryos (sgRNA-*Adnp*) (Fig. 5A,B; Supplemental Table S5). qPCR and immunofluorescent staining were performed to confirm the significant removal of *Adnp* mRNA and protein in *Adnp* knockout embryos compared with control embryos (Supplemental Fig. S8A,B). We then found that knocking out *Adnp* does not significantly impact the developmental efficiency and early stage blastocyst formation at E3.5 (Supplemental Fig. S8C). However, deficient *Adnp* can result in poor late blastocyst quality and hatching ability, and the number of $CDX2^+$ trophectoderm cells was reduced significantly after *Adnp* knockout (Fig. 5C,D; Supplemental Fig. S8D).

In the cellular model, ADNP can recruit HP1 and CHD4 to compete with CTCF for putative motif sites, thereby removing SINE-derived chromatin looping at CTCF sites (Ostapcuk et al. 2018; Kaaij et al. 2019). Therefore, an increase in CTCF-binding sites is observed when *Adnp* is

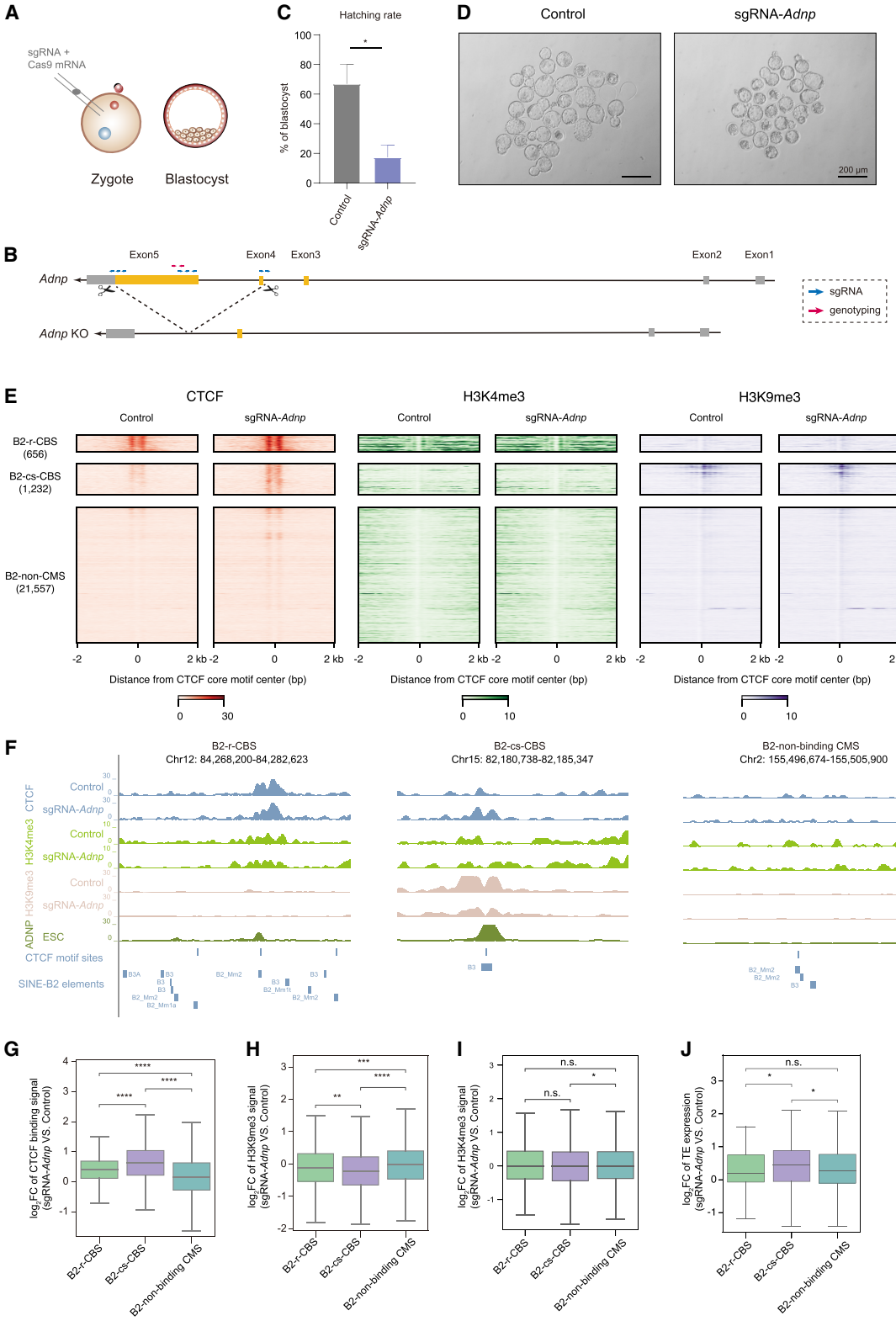

Figure 5. ADNP suppresses CTCF binding at cs-CBSs in blastocysts. (*A*,*B*) Schematic showing the generation of *Adnp* knockout (sgRNA-*Adnp*) mouse embryos by CIRISPR/Cas9-mediated gene editing at the one-cell stage. (*C*) Box plots showing the percentage of sgRNA-*Adnp* embryos that reached the blastocoelic hatching stage. The data are represented as mean ± SD. $n = 3$. (*) *P*-value < 0.01. *P*-value was calculated by Student's *t*-test. (*D*) Representative images of embryos that reached the blastocoelic hatching stage, produced from the indicated treatments related to *C*. Scale bar, 200 μm. (*E*) Heat maps showing the changes of CTCF binding and H3K4me3 or H3K9me3 signals in distinct B2-derived CTCF motif sites between sgRNA-*Adnp* and control blastocyst samples. (*F*) A snapshot of the browser view showing the alternations in CTCF binding and histone modification at three representative distinct B2-derived CTCF motif sites after *Adnp* depletion in embryos. (*G*–*J*) Box plots showing the fold changes of CTCF binding (*G*), normalized H3K9me3 signal (*H*), H3K4me3 signal (*I*), and B2 expression (*J*) in distinct B2-derived CTCF motif sites. *P*-values were calculated by two-sided Mann–Whitney *U*-test. The histone modification signal was normalized according to the local background within ±50 kb. (Control) Water injection samples, (CMS) CTCF motif site, (B2-r-CBS) B2-derived r-CBS, (B2-cs-CBS) B2-derived cs-CBS, (B2-non-CMS) B2-derived nonbinding CTCF motif site. (*) *P*-value < 0.05, (**) *P*-value < 0.01, (***) *P*-value < 0.001, (****) *P*-value < 0.0001, (n.s.) not significant.

knocked out in mouse cell lines (Kaaij et al. 2019). After knocking out *Adnp* in embryos, we compared changes in CTCF binding during the morula-to-blastocyst transition. As expected, we observed enhanced CTCF signals in *Adnp* knockout embryos, and the CTCF binding signals showed a significant increase in B2-derived cs-CBSs compared with those in B2-derived r-CBSs and B2-derived nonbinding CTCF motif sites (Fig. 5E–G; Supplemental Fig. S9A–D). These results provide direct evidence that ADNP-mediated regulation of CTCF binding kinetics also exists in early embryos and can result in the loss of CTCF binding in B2-derived cs-CBSs during blastocyst formation.

Subsequently, we investigated the corresponding histone modification changes in *Adnp* knockout embryos. Given that distinct B2-derived CMS subgroups are located in different genomic regions and that ADNP-associated HNDs are typically marked by narrow H3K9me3 (Fig. 4C; Thorn et al. 2022), we normalized the H3K9me3 signal according to its local background and found that the normalized H3K9me3 signals were significantly reduced in both B2-derived cs-CBSs and r-CBSs, while the H3K4me3 signals remained unchanged following *Adnp* depletion (Fig. 5E,F,H,I; Supplemental Fig. S9A–C,E). Moreover, we performed RNA-seq to uncover the impact of transcriptional activity and found that the expression of SINE B2 elements was up-regulated in B2-derived cs-CBSs of *Adnp* knockout embryos (Fig. 5J; Supplemental Fig. S9A,F). The developmental and molecular defects of *Adnp* knockout embryos were further confirmed by the biological replicate-separated data (sgRNA-*Adnp*-2 vs. sgRNA-*GFP*) when the injection of the Cas9 mRNA + sg-*GFP* group was set as the control (for developmental defects, see Supplemental Fig. S8E–H; for molecular defects, see Supplemental Fig. S9).

In summary, we revealed an ADNP-dependent restriction of functional CTCF usage accompanied by repressed transcriptional activity of corresponding TEs and enrichment of repressive histone modifications at SINE B2-derived cs-CBSs, which further affect cell differentiation and developmental efficiency during embryonic blastocyst formation.

ADNP regulates peri-implantation development

As reported previously, the *Adnp* knockout embryos show lethality at E8.5–E9.5 (Pinhasov et al. 2003). Therefore, we investigated the potential lineage defects of *Adnp* knockout embryos in peri-implantation in addition to reduced trophoblast cell numbers in hatching embryos. We conducted transplantation of the *Adnp* knockout and control embryos (sgRNA-*Adnp* vs. sgRNA-*GFP*) (Supplemental Fig. S10A) into pseudopregnant mice. As observed previously, the implantation ratio was reduced significantly upon *Adnp* depletion at E6.5 and E7.5, and the overall size of E7.5 KO embryos was much smaller than that of control embryos (Fig. 6A,B). Detailed analysis showed that the apoptosis signal of the epiblast (Epi) and the extraembryonic ectoderm (ExE) increased significantly at E6.5 (Supplemental Fig. S10B,C). To identify the major deficiencies in lineage differentiation caused by *Adnp* depletion, we performed RNA-seq of the Epi and ExE at E6.5 and E7.5 stages (Supplemental Table S1) and found that the number of differential expressed transcripts increased significantly in E6.5 ExE and E7.5 Epi samples (168 in E3.5 samples, 1685 in E6.5 ExE samples, 273 in E6.5 Epi samples, 279 in E7.5 ExE samples, and 467 in E7.5 Epi samples). The down-regulated transcripts in E6.5 ExE samples were associated with placenta development, and the down-regulated transcripts in E7.5 Epi samples were associated with neuron development (Fig. 6C; Supplemental Fig. S10D–I). These results indicate that ADNP plays an important regulatory role in peri-implantation embryos.

Discussion

CTCF acts as a core 3D genome structure architecture protein. However, the gradual maturation of 3D genome structures raises questions about the presence of CTCF binding in early embryogenesis (Du et al. 2017; Ke et al. 2017). In this study, we generated the CTCF occupancy data from gametes to blastocysts and revealed the presence of CBSs in both pronuclei no later than the PN-3 stage. Additionally, we observed that although CTCF anchor sites can be established earlier, the formation of chromatin looping, as evident from current Hi-C data (Du et al. 2017; Ke et al. 2017), appears to be significantly delayed. Subsequently, knocking down *CTCF* resulted in impaired genes associated with embryonic development function as well as proper expression of TEs with cs-CBSs. Together with an interesting observation of binding recovery on cs-CBSs in mESCs, we raised an interesting question about whether the CTCF binding status at cs-CBSs is related to the pluripotency status in stem cells. As supporting evidence, the CTCF signal at cs-CBSs is elevated in two-cell-like cells (2CLC) (Supplemental Fig. S11A) based on publicly available data (Supplemental Table S4; Zhu et al. 2021). Moreover, knocking down *Adnp* in mESCs that coexpressed MERVL-tdTomato (Macfarlan et al. 2012) and pZscan4c-EGFP (Dan et al. 2013) fluorescent reporters resulted in elevated expression of *MERVL* but not *Zscan4c* (Fig. 6D–F; Supplemental Fig. S11B). This suggests a multifunctional role of CTCF in early embryos that may extend beyond the induction of 3D structure, and the binding of CTCF to cs-CBSs may potentially be related to the pluripotency status both in vivo and in vitro.

In mammals, transposable elements are major contributors of genetic material: ~50% in mouse and human genomes (Senft and Macfarlan 2021; Fueyo et al. 2022; Lawson et al. 2023). TEs participate in genome structure and transcription regulation through various mechanisms. Earlier studies in humans and mice revealed that a particular family of TEs is often overrepresented in the set of binding sites for a given transcription factor. TE-derived TF binding sites (TFBSs) and *cis*- or *trans*-chromatin regulation expand their regulatory potency and transcriptional activity to a wide range of cell types and developmental stages, which reflects an ongoing coevolution that continues to impact mammalian development.

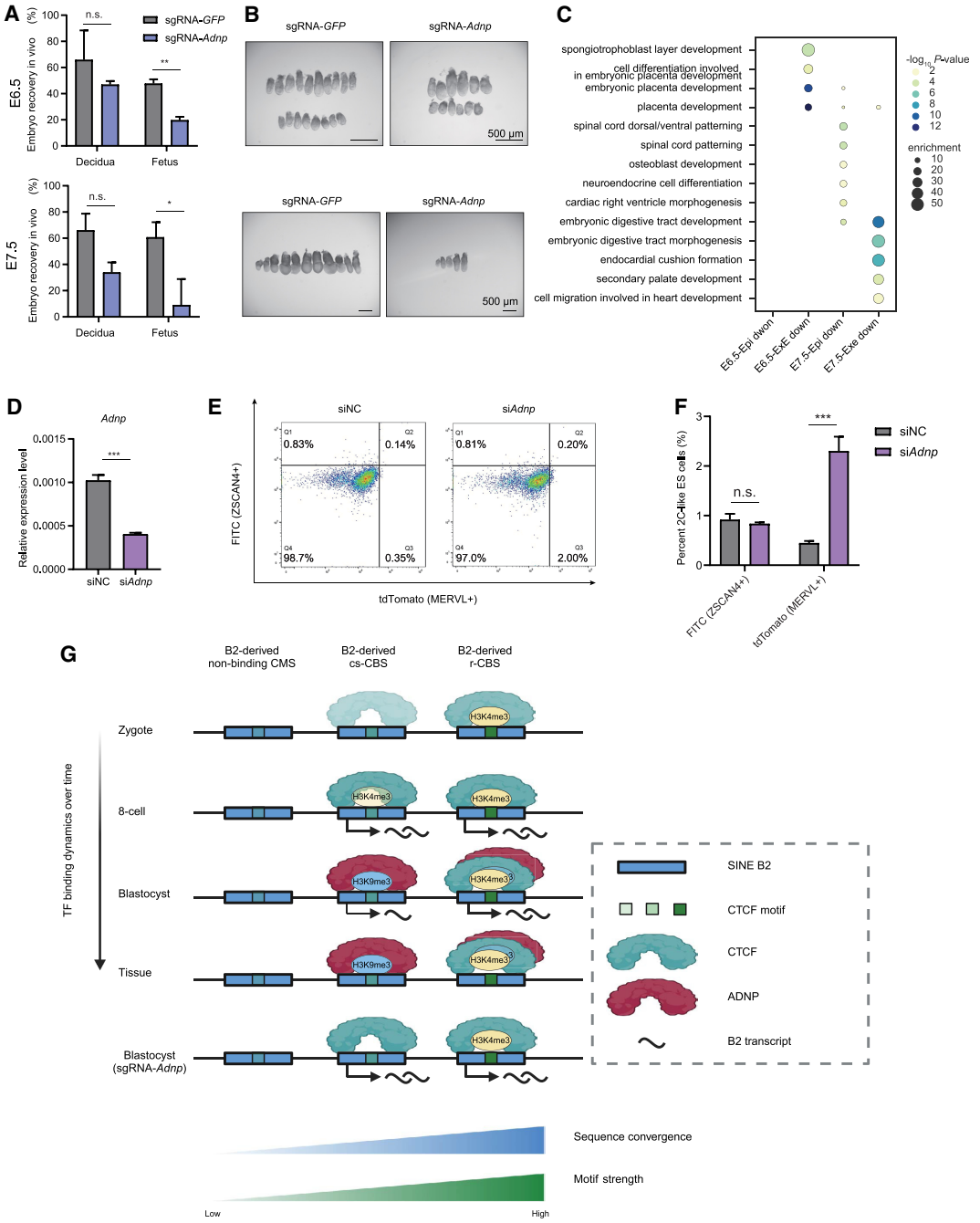

Figure 6. ADNP regulates peri-implantation development and pluripotency status. (*A*) Embryonic development of the sgRNA-*Adnp* and sgRNA-*GFP* groups at E6.5 (*top* panel) and E7.5 (*bottom* panel), respectively. Scale bar, 500 mm. The data are presented as mean ± SD. (*) *P*-value < 0.01, (**) *P*-value < 0.001, (n.s.) not significant. *P*-values were calculated by Student's *t*-test. (*B*) Representative images of post-implantation embryos produced from the indicated treatments related to *A*. (*C*) Dot plots showing the gene ontology (GO) enrichment analysis for down-regulated transcripts upon *Adnp* depletion. (*D*) Validation of Adnp knockdown efficiency by exhibited expression levels in mESCs. The data are represented as mean ± SD. $n = 3$. (***) *P*-value < 0.0001. *P*-value was calculated by Student's *t*-test. (*E*) FACS profiles of MERVL$^+$ (tdTomato) and ZSCAN4$^+$ (EGFP) populations in mESCs using different treatments as indicated. (*F*) FACS quantification of the proportion of MERVL$^+$, ZSCAN4$^+$, and MRVL$^+$ZSCAN4$^+$ cells upon activation by siRNAs targeting *Adnp*. Data are presented as mean ± SD. (***) *P*-value < 0.0001, (n.s.) no significance. *P*-values were calculated by Student's *t*-test. (*G*) Schematics elucidating the fate of CTCF motif sites derived from SINE B2 expansions during embryogenesis (image created with BioRender.com). The youngest class in terms of evolution, inferred as B2-derived nonbinding CTCF motif sites, did not exhibit either CTCF binding or histone marker enrichment. The B2-derived r-CBSs predominantly accumulate CTCF occupancy and H3K4me3 and exhibit high transcriptional activities, marked by the highest sequence convergence and the strongest motif sequence. Alternatively, ADNP substitutes CTCF in the B2-derived cs-CBSs, resulting in transition from H3K4me3 to H3K9me3 and suppression of B2 expression during the morula-to-blastocyst transition. (CMS) CTCF motif site.

Previous research has shown that SINE B2, a rodent-specific TE subfamily, introduces thousands of CTCF motifs into the mouse genome (Gualdrini et al. 2022). CTCF sites introduced by SINE B2 were found to play important roles in chromatin structure in mESCs, and its binding activity is competed by the ChAHP (CHD4, ADNP, and HP1) complex and inhibited by the enrichment of repressive histone modifications (Kaaij et al. 2019). In early embryos, a large number of TEs, including SINEs, lose the repressive DNA or histone modifications and become reactivated, which provides an access for the widespread usage of TE-derived TFBSs. Therefore, investigating the generation and molecular dynamics of TE-derived TFBSs in early embryos is crucial for understanding the impact of TEs on development and evolution, and our study serves as an important example in this context.

Starting with the dynamic binding of CTCF on SINE B2-derived CTCF motif sites, we discovered distinct characteristics during embryogenesis (Fig. 6G), and ADNP may be an important regulatory factor. The largest subgroup (B2-derived nonbinding CTCF motif sites, which exhibit the lowest sequence convergence and the weakest motif sequences) shows little occupation of CTCF or enrichment of H3K4me3 or H3K9me3. On the other hand, the B2-derived cs-CBSs, which mainly acquire CTCF binding at the two-cell stage, possess a highly ordered histone modification switch from H3K4me3 to H3K9me3 during the morula-to-blastocyst transition. Consistent with the elevated CTCF signals at these sites in *Adnp* knockout mESCs, we also successfully observed an increase in CTCF signals, a decrease in H3K9me3 modifications, and the restoration of B2 transcription upon *Adnp* knockout at the blastocyst stage. This suggests the existence of an ADNP-dependent CTCF dissociation mechanism in early embryos that is potentially involved in transcriptional silencing and heterochromatin formation at TE-derived CBSs.

In contrast to these changes, the B2-derived r-CBSs with the highest sequence convergence and strongest motif sequences, although fewer in number, maintain high CTCF binding activity, H3K4me3 modification levels, and transcriptional activities and moderate H3K9me3 deposited in the morula-to-blastocyst transition. Interestingly, all three subgroups of B2-derived CTCF motif sites exhibit increased CTCF signals upon *Adnp* knockout in both mESCs and embryos. The remarkable binding of ADNP at all three subgroups of B2-derived CTCF motif sites in mESCs also supports this competitive relationship. Considering the divergence of CTCF motif sequences among the three subgroups, we proposed that the CTCF motif sequence still serves as the primary determining factor, and weaker CTCF binding allows ADNP to gain an advantage in this competition model. Interestingly, we noticed that the loss of H3K4me3 modifications is not influenced by ADNP, indicating that the erasure of H3K4me3 modifications at the B2-derived cs-CBSs is ADNP-independent and may occur earlier, before the morula stage. The role of dynamic H3K4me3 modifications in this process remains unclear and requires further investigation.

Sequence motifs play a crucial role in determining TF binding sites (Figs. 1D, 2D). Misplaced or misactivated TF binding sites, taking CBSs as an example, can disrupt the 3D structure of the genome and be associated with cancer (Katainen et al. 2015; Choudhary et al. 2020, 2023; Fang et al. 2020; Han et al. 2021). New motif sequences are generated through point mutations, small indels, or insertions. The expansion of TEs in the genome can provide new material for genetic evolution, which may also introduce a mass of motif sequences to the embryonic genome (Gassler et al. 2022). However, introducing new CTCF motifs to the genome is also a challenging process, as it is crucial to avoid mistakes. Previous studies have primarily focused on the B2-derived CBSs and proposed that H3K9me3 and DNA methylation suppress these newly derived CBSs (Choudhary et al. 2020). When broadening our perspective to include all B2-derived CTCF motif sites, we observed that most of the B2-derived CTCF motif sites that lack CTCF binding (referred to as B2-derived nonbinding CTCF motif sites) also lack DNA methylation or H3K9me3 during early embryogenesis. Here, we propose a model to elucidate the evolutionary process of B2-derived CTCF motif sites. The expansion of SINE B2 elements could introduce various types of new CTCF motif sites. B2-derived nonbinding CTCF motif sites with a weak motif recognition represent a type of neutral mutation that remains compatible with the genome, whereas other types of B2-derived CTCF motif sites are mostly eliminated through natural selection. Subsequently, some of these B2-derived CTCF motif sites accumulate mutations and acquire improved motif recognition sufficient for CTCF binding during evolution. These sites may display different fates. First, if they possess the optimal CTCF motif, they become r-CBSs, serving as backup for previous chromatin loop anchors or participating in the formation of new 3D genome structures. Second, if they contain a suboptimal motif, they become cs-CBSs. ADNP counteracts the stable localization of CTCF and suppresses these sites. Third, if they possess the optimal motif but form harmful new 3D genome structures, they would get filtered through natural selection. Further discussion of this hypothesis will enhance our understanding of the evaluation of transcriptional regulatory networks.

The existence of totipotency resulted in the first lineage differentiation (Hackett and Surani 2014), and CTCF binding at cs-CBSs exists only in totipotent cells following suppression by ADNP. The emerged phenotype and gene expression change in *Adnp* knockout embryos reminds us that ADNP is necessary for the totipotency exit process. Moreover, the increased population of 2CLCs in *Adnp* knockdown mESCs suggests that ADNP may be a suppressor of totipotency.

Materials and methods

Animals and mouse oocyte or embryo collection

Specific pathogen-free (SPF) mice were housed in the animal facility at Tongji University, Shanghai, China. All animal maintenance and experimental procedures were

carried out according to the Health Guide for the Care and Use of Laboratory Animals and were approved by the Biological Research Ethics Committee of Tongji University.

To get preimplantation embryos, 7-wk-old C57BL/6N female mice were superovulated by intraperitoneal injection with 6 IU of pregnant mare serum gonadotropin (PMSG) and 6 IU of human chorionic gonadotrophin (hCG; San-Sheng Pharmaceutical). The superovulated female mice were then mated with DBA2 male mice. Zygotes were collected from the oviducts of the female mice at 20 h after hCG injection and further cultured in G-1 Plus medium (Vitrolife) to reach each corresponding developmental stage. PN-3 stage zygotes were distinguished based on the microscopic observation of the size of the two pronuclei and the distance between them (Adenot et al. 1997). Next, we collected preimplantation embryos at the following time points after hCG: two-cell embryos at 36 hpf, four-cell embryos at 58 hpf, eight-cell embryos at 71 hpf, morulas at 85 hpf, and blastocysts at 106 hpf. Germinal vesicle (GV) stage oocytes were collected 48 h after PMSG injection. To collect GV oocytes from 3-wk-old mice, the whole ovaries were clipped mechanically with a razor blade. Fully grown GV oocytes were distinguished mainly based on the microscopic observation of the prominent but structurally homogenous bodies called "nucleolus-like bodies" (Inoue et al. 2007). All embryos were cultured using microdroplet culture methods with G-1 Plus (Vitrolife) medium covered with mineral oil in a humidified incubator at 37°C with 5% CO_2.

Sample harvest for CUT&RUN-seq, total RNA-seq, uliNChIP-seq, and ULI-MNase-seq

Samples of GV oocytes; PN-3 stage, two-cell stage, and eight-cell stage embryos; isolated inner cell mass (ICM) and trophectoderm of day 3.5 blastocysts or day 3.5 whole blastocysts; and an ESC line (R1) were harvested for CUT&RUN-seq, total RNA-seq, and uliNChIP-seq. The zona pellucidae of the GV zygotes and cleavage stage embryos were removed with 0.5% pronase E (Sigma), and the cleavage stage embryos were then incubated in Ca^{2+}-free Chatot–Ziomek–Bavister (CZB) medium for 10–20 min. Polar bodies were removed by gentle pipetting using a fire-polished glass needle. Single blastomeres were separated and manually picked before being subjected to CUT&RUN library preparation and sequencing. For ICM and trophectoderm isolation (Liu et al. 2016), the zona pellucidae of blastocysts were punched, followed by removal with 0.5% pronase E. The embryos were then incubated in Ca^{2+}-free CZB for 20–30 min to disrupt the cell–cell junctions. ICM cells and trophectoderm cells were then distinguished and collected according to their sizes and shapes with the aid of a piezo-driven micromanipulator.

Cell culture

The R1 ESCs were cultured on 0.1% gelatin-precoated dishes in DMEM (Sigma) supplemented with 15% fetal bovine serum (FBS; Gibco), 1000 U/mL leukemia-inhibiting factor (LIF; Millipore), 1 mM L-glutamine (Thermo Fisher Scientific), 100× nonessential amino acids (Millipore), 100× nucleosides (Sigma), 100× penicillin/streptomycin (Gibco), 0.1 mM β-mercaptoethanol (Sigma), 1 μM PD0325901, and 3 μM CHIR9902. The mESC line used in this study was regularly tested and found to be negative for mycoplasma contamination.

CUT&RUN library construction and sequencing

CUT&RUN was conducted following the published protocols (Skene et al. 2018; Patty and Hainer 2021) with a few modifications. Briefly, all isolated fresh blastomeres or cells were washed three times with CUT&RUN wash buffer to avoid possible contamination and then transferred into a 1.5-mL low-binding PCR tube (Eppendorf) containing 100 μL of the wash buffer. The samples were then bound to Concanavalin A-coated magnetic beads (Bangs Laboratories) that had been activated and resuspended in CUT&RUN binding buffer. After cell immobilization, bead-bound samples were successively incubated with the appropriate amount of primary antibodies against the protein of interest (CTCF antibody; Millipore 07-729) in 50 μL of CUT&RUN antibody buffer overnight at 4°C. Cell membranes were permeabilized with 0.01% digitonin to allow the specific antibody to find its target. After unbound antibodies were washed away, 700 ng/mL protein A-MNase (pA-MN; a gift from Steven Henikoff's laboratory) was added and the mixture was incubated for 1 h at 4°C. After washing, $CaCl_2$ was added to a final concentration of 2 mM to activate pA-MN, and the digestion reaction was carried out for 30 min at 0°C and then stopped by adding 100 μL of 2× CUT&RUN stop buffer. The protein–DNA complex fragments were then released by 20-min incubation at 37°C. After the supernatant was transfered to a new tube, 2 μL of 10% SDS and 2.5 μL of Proteinase K (Thermo) were added and the mixture was incubated for 30 min at 55°C. DNA was then precipitated by phenol/chloroform/isoamylalcohol followed by ethanol precipitation with glycogen and then dissolved in nuclease-free water. Sequencing libraries were prepared using KAPA Hyper preparation kit (KAPA Biosystems) following the manufacturer's instructions with slight modifications. Briefly, end repair was conducted for 30 min at 20°C followed by end repair/A-tailing for 30 min at 50°C. After adaptor ligation for 30 min at 20°C, the DNA fragments were purified by 1.2× volume of AMPure beads (Beckman Coulter) followed by 18 cycles of PCR amplification with 2× KAPA HiFi HotStart ready mix. The final libraries were cleaned with 1× vol of AMPure beads, and all CUT&RUN libraries were sequenced on a NovaSeq 6000 (Illumina) platform with paired-end 150-bp reads at Berry Genomics Corporation and Nanjing Jiangbei New Area Biopharmaceutical Public Service Platform Co., Ltd.

ULI-MNase library construction and sequencing

Ten to 15 blastomeres or cells per replicate were isolated and washed before being placed into 0.7 μL of lysis buffer (10 mM Tris-HCl at pH 8.5, 5 mM $MgCl_2$, 0.6% NP-40)

for individual reactions. Next, 2.5 μL of MNase master mix (MNase buffer, 0.125 U/μL MNase [NEB M0247S], 2 mM DTT, 5% PEG 6000) was added into each tube, and the reaction was incubated for 10 min at 25°C for chromatin fragmentation. The reaction was stopped by the addition of 0.32 μL of 100 mM EDTA, and then 0.32 μL of 2% Triton X-100 was added to the reaction to release the fragmented chromatin. Next, 0.2 μL of 20 mg/mL protease was added, and the reaction was incubated for 90 min at 50°C for protein digestion followed by incubation for 30 min at 75°C for protease inactivation. The sequencing libraries were prepared using the KAPA Hyper preparation kit for the Illumina platform following the manufacturer's instructions. After standard procedures including end repair/A-tailing, adapter ligation, postligation cleanup, and library amplification, the resulting products were subjected to a second round of PCR amplification with the same provided primers to generate sufficient DNA material for high-throughput sequencing. Paired-end sequencing with 150-bp reads was performed on a NovaSeq 6000 (Illumina) platform at Berry Genomics Corporation.

Generation of knockout mouse embryos

To ensure the effective deletion of *Adnp* in most embryos, two single guide RNAs (sgRNAs) targeting exon 4 and six sgRNAs targeting exon 5 of the *Adnp* gene were designed (Fig. 5A,B). The Cas9 mRNA and sgRNAs were produced as previously reported (Zhang et al. 2020). The sequence for each sgRNA was cloned into the sgRNA expression vector pUC57, and in vitro transcription was then performed using the MEGAshortscript T7 transcription kit (Invitrogen). To obtain Cas9 mRNA, the Cas9 expression construct pST1374-Cas9-N-NLS-FLAG-linker was linearized and transcribed using mMessage mMachine T7 Ultra transcription kit (Invitrogen). The integrity of the manufactured mRNA was confirmed by electrophoresis. Both Cas9 mRNA and specific sgRNA were purified according to the standard protocol by phenol:chloroform extraction and ethanol precipitation and then dissolved in nuclease-free water (Life Technologies). For microinjection, the Cas9 mRNA was diluted to 100 ng/μL, and each sgRNA mix was diluted to a final concentration of 50 ng/μL. B6D2F1 female mice (~8 wk old) were superovulated and mated with B6D2F1 male mice before zygotes were collected from the oviducts. The harvested zygotes were microinjected with a mixture containing100 ng/μL Cas9 mRNA and 50 ng/μL each sgRNA, and the embryos were then cultured in G-1 Plus medium (Vitrolife). The Cas9 mRNA plus sgRNA targeting *GFP* or water were used as control. Embryos were then observed and summarized from the two-cell stage to the hatching blastocyst stage at E4.5. To test the knockout efficiency, we used 50 injected embryos as the PCR template and amplified targeted exons of *Adnp*. IF was also applied to test the knockout efficiency on protein level. The blastocyst stage embryos were harvested for CUT&RUN-seq, total RNA-seq, and uliNChIP-seq or transferred to pseudopregnant females. Primers and sgRNAs are listed in Supplemental Table S5.

Small interfering RNA (siRNA)-mediated gene knockdown in embryos and mESCs

Five siRNAs targeting *CTCF* (GenePharma) were mixed together and diluted at a total concentration of 20 μM for microinjection; the scramble siRNA (GenePharma) was also diluted at a total concentration of 20 μM. Zygotes obtained from B6D2F1-crossed female mice were injected with ~10 pL of siRNA solution using a piezo-driven micromanipulator. The embryo were then cultured in G-1 Plus medium (Vitrolife) at 37°C in an atmosphere of 5% CO_2 for observation and sample collection. IF and qPCR were used to confirm the KD efficiency in the embryos.

The siRNA mix of *Adnp* and the scramble siRNA (GenePharma) was transferred into mESCs using Lipofectamine RNAiMAX transfection reagent (Thermo Scientific). After 48 h of knockdown, effective knockdown cells were collected for FACS and RNA sampling. qPCR was used to confirm the KD efficiency in mESCs. All siRNAs were synthesized by GenePharma, and the sequence information is listed in Supplemental Table S5.

Blastocyst transfer

To verify the function of ADNP in embryos after implantation, *Adnp* knockout and control embryos (sgRNA-Adnp vs. sgRNA-GFP) were cultured in vitro to the E3.5 blastocyst stage and then transferred into the uteri of pseudopregnant ICR female mice (15 blastoids for each). To obtain postimplantation embryos, female mice at 6.5 or 7.5 d postcoitum (dpc) were executed and the uteri were dissected and transferred to a Petri dish with PBS. Next, each decidua was carefully freed from the uterine muscle layers using properly sharpened forceps. Decidua and embryo numbers were recorded and divided by the number of pseudopregnant females to reflect the implantation rate. Reichart's membrane and the ectoplacental cone were also removed from the embryos, and the Epi and ExE were cut and harvested for total RNA-seq.

TUNEL assay

The *Adnp* knockout and control embryos were checked with TUNEL at E6.5 to detect the apoptosis signal. Embryos were fixed with 4% paraformaldehyde (Sigma) overnight at 4°C and washed twice with DPBS (Life Technologies). Permeabilization was performed with 0.5% Triton X-100 in DPBS for 1 h at room temperature. After two washes with DPBS, the embryos were moved into 100 μL of equilibration buffer for 10 min at room temperature followed by 50 μL of rTdT incubation buffer for 1 h at 37°C in the dark. Next, the embryos were moved into 50 μL of 2× SSC for 15 min at room temperature followed by three washes with DPBS. The TUNEL assay was performed using the DeadEnd fluorometric TUNEL system (Promega). All samples were stained with 1 μg/mL DAPI (Sigma) for 20 min at room temperature. Last, samples were manipulated using a mouse pipette, placed in a 96-well U-type plate, and finally transferred to DPBS drops covered by paraffin oil on a glass-bottom cell culture

dish (NEST) for imaging with a Zeiss LSM 880 microscope. Images were processed and quantified in ImageJ software.

Reverse transcription and quantitative real time-PCR (RT-qPCR)

Mouse embryonic stem cells were disrupted in TRIzol reagent (Invitrogen), and total RNAs were isolated by chloroform extraction coupled with isopropanol precipitation. RNAs of eight-cell embryos were isolated using TRIzol reagent and chloroform, and 1/10 vol of 3 mol/L NaAc and 1 μL of glycogen were added to the aqueous phase of each sample. RNAs were then precipitated by isopropanol and washed twice with 75% ethanol before they were eluted with nuclease-free water. cDNA was then synthesized using All-In-One RT MasterMix (Applied Biological Materials). qPCR was carried out using TB Green Premix Ex Taq II (Takara Bio) and monitored by a 7500 Fast real-time PCR system, and three technical replicates were performed for each sample. The relative expression level of the target gene *Adnp* was normalized to the reference gene *Gapdh* for mESC samples or to *H2afz* for embryo samples. qPCR primers for tested genes are listed in Supplemental Table S5.

Total RNA library construction and sequencing

Embryos with zona pellucidae and polar bodies removed were disrupted in TRIzol reagent (Invitrogen), total RNAs were isolated by chloroform extraction coupled with isopropanol precipitation, and 1/10 vol of 3 M NaAc and 1 μL of glycogen were added to the aqueous phase of each sample. RNAs were washed twice with 75% ethanol before they were eluted with nuclease-free water. Purified RNAs were then subjected to library generation using SMARTer stranded total RNA-seq kit (Takara) following the manufacturer's instructions. Briefly, random primers were used for reverse transcription, and the amplified cDNA was then subjected to ribosomal RNA depletion. Prepared RNA-seq libraries were sequenced on an Illumina NovaSeq 6000 platform with paired-end 150-bp reads at Nanjing Jiangbei New Area Biophamaceutical Public Service Platform Co., Ltd.

uliNChIP library construction and sequencing

uliNChIP-seq was performed as previously described (Brind'Amour et al. 2015) to capture the H3K4me3 and H3K9me3 status in embryos. Briefly, the harvested blastomeres or cells were subjected to nucleus extraction buffer. The MNase master mix was then added and the mixture was incubated for 10 min at 25°C to allow MNase digestion. The reaction was stopped by 10 mM EDTA, and 0.1% Triton X-100 together with 0.1% DOC were added to lyse the nuclear membrane. The released chromatin was then diluted with ChIP buffer. After 1/20 vol of the reaction was saved as the input, the rest was incubated with primary antibody-coated (H3K4me3 antibody [Cell Signaling Technology 9727] and H3K9me3 antibody [Active Motif 39161]) Dynabeads Protein A/G overnight at 4°C. The ChIP samples were washed twice with low-salt wash buffer and twice with high-salt wash buffer. The washed beads were then incubated in hot elution buffer for 2 h at 65°C with shaking. The eluted DNAs were further purified and subjected to sequencing library generation as described above. Paired-end 150-bp sequencing was also performed on ChIP libraries at Nanjing Jiangbei New Area Biophamaceutical Public Service Platform Co., Ltd.

Immunofluorescent staining

Embryos were fixed with 4% paraformaldehyde (Sigma) overnight at 4°C and then permeabilized with 0.5% Triton X-100 for 15 min at room temperature. The samples were blocked with 2.5% bovine serum albumin (BSA; Sigma) for 1 h at 25°C and then incubated with the primary antibodies against CTCF (ABclonal), ADNP (R&D Systems), SOX2 (ABclonal), or CDX2 (Biogenex) overnight at 4°C. After three washes with TBST, the samples were incubated with the appropriate secondary antibodies for 45 min. The nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI). All stained samples were observed using a Zeiss LSM880 confocal microscope. Images were processed and quantified in ImageJ software.

Detection of the 2C-like cell proportion within ESC cultures

To compare the 2C-like cell proportion before and after *Adnp* depletion, mESCs that coexpressed MERVL-tdTomato (Macfarlan et al. 2012) and pZscan4c-EGFP (Dan et al. 2013) fluorescent reporters were obtained for FACS analysis. Cells with credible GFP or tdTomato expression were considered to be in a 2C-like state. The expression levels of *Adnp*, *Zscan4c*, and *MERVL* in *Adnp* knockdown and control mESCs were checked by RT-qPCR to confirm the change of 2C-like cell proportion.

Fluorescence-activated cell sorting (FACS)

For FACS, cells were collected by 0.05% trypsin-EDTA digestion for 4 min and washed with FACS buffer containing PBS supplemented with 2% FBS. After centrifugation, cells were resuspended in FACS buffer. All analyses were performed on a CytoFLEX S (Beckman Coulter). FACS analysis was used to identify the 2C-like cell proportion of *Adnp* knockdown and control mESCs.

Histone modification ChIP-seq data processing

Obtained raw reads were filtered using Trim Galore! (version 0.6.5; https://github.com/FelixKrueger/TrimGalore) with Cutadapt (version 3.3) (Martin 2011) and the parameter "--trim-n." The filtered reads were mapped to the mouse genome (mm10 assembly) using Bowtie 2 (version 2.4.2) (Langmead and Salzberg 2012) with the parameters "--nomixed --no-discordant --no-unal." Mapped reads with a mapping quality (MAPQ) score ≥30 were retained

and converted to BAM format using SAMtools (version 1.6; Danecek et al. 2021). The BAM files were further converted to BED format using BEDTools (version 2.27.1; Quinlan and Hall 2010). Replicates were merged, and the central 73 bp of all unique fragments were used for pile-up analysis. The resulting pile-up data were transformed into bigWig format for visualization and subsequent analysis using custom scripts and BEDTools (version 2.27.1; Quinlan and Hall 2010). Broad H3K4me3 domain detection was performed as previously described (Liu et al. 2016).

TF ChIP-seq data processing

ChIP-seq data for ADNP and CTCF were obtained from the Gene Expression Omnibus (GEO) data sets GSE97945 (Ostapcuk et al. 2018) and GSE125129 (Kaaij et al. 2019), respectively. Data were processed as previously described with mapping against the mouse genome (mm10 assembly; Wang et al. 2022b).

WGBS data processing

Raw reads were obtained from the Gene Expression Omnibus (GEO) data set GSE98151 (Wang et al. 2018) and processed as previously described with little modification (Liu et al. 2018). Raw reads were filtered using Trim Galore! (version 0.6.5; https://github.com/FelixKrueger/TrimGalore) with Cutadapt (version 3.3; Martin 2011) and the "--trim-n" parameter. The processed reads were then aligned to the mouse genome (mm10 assembly) using BSMAP (Xi and Li 2009). Methylation levels were determined using MCALL (Sun et al. 2014). Both BSMAP (Xi and Li 2009) and MCALL (Sun et al. 2014) were modules of MOABS (version 1.3.9.6) (Sun et al. 2014). Subsequently, the methylation level of each CpG site and methylation signal tracks were generated for downstream analysis. Regions with mean DNA methylation levels <0.2 were classified as low DNA methylation, those >0.8 were classified as high DNA methylation, and the remaining sites were classified as medium DNA methylation.

3D chromatin-related data processing

We used publicly available CTCF mESC ChIA-PET loops from GSE99520 (Weintraub et al. 2017) and performed a liftover to the mm10 genome assembly using CrossMap (Zhao et al. 2014). Raw reads of Hi-C data were obtained from GSE82185 (Du et al. 2017) and processed using HiC-Pro (version 3.0.0; Servant et al. 2015) to extract all valid pairs with a minimum MAPQ >30. The replicates were then merged and converted to the mcool format using Cooler (version 0.8.2; Abdennur and Mirny 2020). Aggregation peak analysis was conducted using coolpup.py (version 1.0.0; Flyamer et al. 2020) on the previously mentioned CTCF loop anchors with the parameter "--pad 100000" at a resolution of 10 kb. The mean of the center signal was assigned as the score.

RNA-seq processing

To quantify gene expression, RNA-seq data were filtered using Trim Galore! (version 0.6.5; https://github.com/FelixKrueger/TrimGalore) with Cutadapt (version 3.3; Martin 2011), applying the "--trim-n" parameter. The filtered reads were then aligned to the mouse genome (mm10 assembly) using HISAT2 (version 2.1.0; Kim et al. 2019) with the parameters "--no-mixed --no-discordant." FPKM and read counts for each gene were calculated using StringTie (version 1.3.3b; Kovaka et al. 2019). To quantify the expression of TE transcripts, we followed the methods previously described (Shao and Wang 2021). The RNA-seq data were filtered using fastp (version 0.23.1; Chen et al. 2018) with the parameter "-D" and then mapped to the mouse genome (mm10 assembly) using STAR (version 2.7.2a; Dobin et al. 2013). Transcripts from various developmental stages (two-cell, four-cell, eight-cell, morula, ICM, and trophectoderm) were assembled from the mapping results using StringTie (version 2.2.1; Kovaka et al. 2019) and merged using Taco (version 0.7.3; Niknafs et al. 2017). The expression of transcripts from all samples was determined by counting reads using featureCounts (version 2.0.1; Liao et al. 2014), followed by redistribution using published scripts (Shao and Wang 2021) and normalization to FPKM. Transcripts overlapping with B2 were identified as B2 transcripts.

ULI-MNase-seq data processing

The raw reads were subjected to quality control using Trim Galore! (version 0.6.5; https://github.com/FelixKrueger/TrimGalore) with Cutadapt (version 3.3; Martin 2011) and the parameter "--trim-n." The filtered reads were then aligned to the mm10 genome using Bowtie 2 (version 2.4.2; Langmead and Salzberg 2012) with the parameters "--no-mixed --no-unal." Reads with a MAPQ score <30 were removed, and the resulting alignments were converted to BAM format using SAMtools (version 1.6; Danecek et al. 2021). The BAM files were further converted to BED format using BEDTools (version 2.27.1; Quinlan and Hall 2010). The replicates were then merged, and the central 73 bp of all unique fragments were used for pile-up. The resulting pile-up data were transformed into bigWig format for visualization and subsequent analysis, which was achieved using custom scripts and BEDTools (version 2.27.1; Quinlan and Hall 2010).

CTCF motif site detection

The motif position weight matrix for CTCF (MA0139, MA1929, and MA1930) in MEME format was obtained from JASPAR (Castro-Mondragon et al. 2022). Motif sites were scanned against the mouse genome (mm10 assembly) with MA0139.1 as previously described (Wang et al. 2022b). All CTCF motif sites were merged based on strand and location using BEDTools (version v2.29.2; Quinlan and Hall 2010), and only the center positions were retained for further analysis. To determine the strength of the CTCF upstream motif, MA1929.1 and MA1930.1

were used to scan the 20- to 40-bp region upstream of the CTCF core motif center against the genome sequence. CTCF motif sites overlapping with B2 by ≥10 bp were classified as B2-derived CTCF motif sites.

CTCF CUT&RUN data processing

Raw sequenced read pairs were filtered by Trim Galore! (version 0.6.5; https://github.com/FelixKrueger/TrimGalore) with Cutadapt (version 3.3; Martin 2011) using the parameter "--trim-n." The filtered reads were mapped back to the genome (mm10) using Bowtie 2 (version 2.4.2; Langmead and Salzberg 2012) with the parameters "--no-mixed --no-discordant --no-unal." Mapped read pairs (fragments) with MAPQ of ≥30 were retained and converted to BAM format using SAMtools (version 1.6; Danecek et al. 2021). BAM files were converted to BEDPE files using BEDTools (version 2.27.1; Quinlan and Hall 2010). Replicates were merged and sampled down to 6 M, and peak calling was performed by MACS (version 2.1.3; Zhang et al. 2008) with the parameters "-f BEDPE -g mm -q 0.05." Peaks with a fold change of ≥5 and Q-value $\leq 1 \times 10^{-5}$ were maintained for the downstream analysis. The middle halves of fragments were piled up and transformed into bigWig format using custom scripts and BEDTools (version 2.27.1; Quinlan and Hall 2010).

CTCF-binding site groups definition

We obtained 460 publicly available CTCF ChIP-seq data sets from the Cistrome data browser (Mei et al. 2017; Zheng et al. 2019). Only 349 data sets were kept for the following analysis with at least 10,000 valid peaks (fold change ≥ 5 and Q-value $\leq 1 \times 10^{-5}$), and more than half of those peaks contained CTCF motifs (Supplemental Table S4). We merged and deduplicated peaks from CTCF ChIP-seq peaks and embryonic CTCF CUT&RUN peaks. We found that 86,176 peaks overlapped with CTCF motif sites and denoted them as potential CTCF-binding sites (CBSs) (Supplemental Table S3). In addition, 79,173 CTCF motif sites without any CTCF binding were termed as other CTCF motif sites (Supplemental Fig. S1A,B).

To determine the binding status of potential CBSs in each period, we assessed their overlap with peaks from the corresponding period samples (e.g., PN-3 CBSs and PN-3 nonbinding CTCF motif sites) (as shown in Fig. 1). CTCF anchor sites and non-CTCF anchor sites were CBSs in each period, depending on their overlapping situation with CTCF mESC ChIA-PET loop anchors (Weintraub et al. 2017).

The dynamics of potential CBSs were analyzed considering both the peak and the signal. Potential CBSs were classified as loss CBSs if they overlapped with a peak and had a fold change of >10 in the previous stage but showed an absence of peaks and had a fold change of <5 in the later stage. Conversely, potential CBSs were considered gain CBSs if they overlapped with a peak and had a fold change of >10 in the later stage but showed an absence of peaks and had a fold change of <5 in the previous stage. Specifically, the intersection of lost potential CBSs during the eight-cell stage to the ICM and the eight-cell stage to TEs was identified as cleavage-specific CBSs (cs-CBSs), while CBSs with consistent binding in eight-cell embryos, the ICM, and TEs with a signal >10 were determined to be r-CBSs. Next, based on their overlap with B2, the B2-r-CBSs, B2-cs-CBSs, and B2 nonbinding CTCF motif sites were defined.

UCSC genome browser

The genome browser view was obtained using the UCSC genome browser (Kent et al. 2002) with Track data hubs (Raney et al. 2014) and visualized with smoothing with a mean of pixels. The signal was normalized by the average of the whole genome for visualization if not stated otherwise.

Statistical analysis

P-values were calculated by two-sided Mann–Whitney U-test if not stated otherwise in the figure legends.

Data and code access

The raw sequence data reported here have been deposited in the Genome Sequence Archive (Chen et al. 2021) of the National Genomics Data Center (sCNCB-NGDC Members and Partners 2023), China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA; CRA011730), which is publicly accessible at https://ngdc.cncb.ac.cn/gsa.

The processed results (bigWig files for the CUT&RUN, ChIP-seq, and MNase-seq data sets and gene expression matrix for the total RNA-seq data set) for high-throughput data generated in this study were deposited in Zenodo (https://zenodo.org/records/10518964).

The processing code for CTCF CUT&RUN/Tag data was deposited in GitHub (https://github.com/TongjiZhanglab/ADNP_Modulates_SINE_B2_CTCF).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Dr. S. Henikoff for kindly sharing pA-MNase and advice on the use of CUT&RUN, and Hui Yang and Chenfei Wang for suggestions on computational analysis. We also thank the ENCODE Consortium and related laboratories for generating mouse adult tissue total RNA-seq data used in this study. This work was primarily supported by the National Key R&D Program of China (2021YFC2700300 and 2022YFC2702200 to S.G., 2020YFA0509800 and 2019YFA0110100 to R.G., 2020YFA0113200 to Y.G., 2021YFA1100300 to J.C., and 2021YFA1302500 to Y. Zhang), the National Natural Science Foundation of China (31721003 to S.G., 31922022 and 32370869 to

Y.G., 32270855 to R.G., and 32030022 and 32325012 to Y. Zhang), the Natural Science Foundation of Shanghai Municipality (21ZR1465500 to R.G.), the Shanghai Pilot Program for Basic Research (to Y.G.), and the Peak Disciplines (type IV) of Institutions of Higher Learning in Shanghai (to S.G.).

Author contributions: R.G., W.W., Y.G., Y. Zhang, and S.G. conceived and designed the experiments. R.G., M.M., R.Z., and X.W. performed CUT&RUN experiments. R.G. and M.M. performed ULI-NChIP-seq and total RNA-seq experiments. W.W. and D.Y. performed all the computational analyses. Chuang Chen performed the ULI-MNase-seq. R.G. and M.M. performed the embryo- and mESC-related experiments, assisted by X.K., Y. Zhao, J.C., X.L., J. Lu, B.X., and J. Liu. W.W., D.Y., R.Z., J.C., X.K., Y. Zhao, X.L., and H.W. assisted with the sample and library preparation. Y.H. and Chaoqun Chen assisted with the manuscript preparation. W.W., R.G., D.Y., Y.G., Y. Zhang, and S.G. wrote the manuscript.

References

Abdennur N, Mirny LA. 2020. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36:** 311–316. doi:10.1093/bioinformatics/btz540

Adenot PG, Mercier Y, Renard JP, Thompson EM. 1997. Differential H4 acetylation of paternal and maternal chromatin precedes DNA replication and differential transcriptional activity in pronuclei of 1-cell mouse embryos. *Development* **124:** 4615–4625. doi:10.1242/dev.124.22.4615

Andreu MJ, Alvarez-Franco A, Portela M, Gimenez-Llorente D, Cuadrado A, Badia-Careaga C, Tiana M, Losada A, Manzanares M. 2022. Establishment of 3D chromatin structure after fertilization and the metabolic switch at the morula-to-blastocyst transition require CTCF. *Cell Rep* **41:** 111501. doi:10.1016/j.celrep.2022.111501

Bailey TL, Williams N, Misleh C, Li WW. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* **34:** W369–W373. doi:10.1093/nar/gkl198

Bell AC, Felsenfeld G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405:** 482–485. doi:10.1038/35013100

Bell AC, West AG, Felsenfeld G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98:** 387–396. doi:10.1016/S0092-8674(00)81967-4

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, et al. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18:** 1752–1762. doi:10.1101/gr.080663.108

Brind'Amour J, Liu S, Hudson M, Chen C, Karimi MM, Lorincz MC. 2015. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat Commun* **6:** 6033. doi:10.1038/ncomms7033

Carmona-Aldana F, Zampedri C, Suaste-Olmos F, Murillo-de-Ozores A, Guerrero G, Arzate-Mejía R, Maldonado E, Navarro RE, Chimal-Monroy J, Recillas-Targa F. 2018. CTCF knockout reveals an essential role for this protein during the zebrafish development. *Mech Develop* **154:** 51–59. doi:10.1016/j.mod.2018.04.006

Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I, Lemma RB, Turchi L, Blanc-Mathieu R, Lucas J, Boddie P, Khan A, Manosalva Perez N, et al. 2022. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50:** D165–D173. doi:10.1093/nar/gkab1113

Chao W, Huynh KD, Spencer RJ, Davidow LS, Lee JT. 2002. CTCF, a candidate *trans*-acting factor for X-inactivation choice. *Science* **295:** 345–347. doi:10.1126/science.1065982

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34:** i884–i890. doi:10.1093/bioinformatics/bty560

Chen X, Ke Y, Wu K, Zhao H, Sun Y, Gao L, Liu Z, Zhang J, Tao W, Hou Z, et al. 2019. Key role for CTCF in establishing chromatin structure in human embryos. *Nature* **576:** 306–310. doi:10.1038/s41586-019-1812-0

Chen TT, Chen X, Zhang SS, Zhu JW, Tang BX, Wang AK, Dong LL, Zhang ZW, Yu CX, Sun YL, et al. 2021. The Genome Sequence Archive family: toward explosive data growth and diverse data types. *Genomics Proteomics Bioinformatics* **19:** 578–583. doi:10.1016/j.gpb.2021.08.001

Choudhary MN, Friedman RZ, Wang JT, Jang HS, Zhuo X, Wang T. 2020. Co-opted transposons help perpetuate conserved higher-order chromosomal structures. *Genome Biol* **21:** 16. doi:10.1186/s13059-019-1916-8

Choudhary MNK, Quaid K, Xing X, Schmidt H, Wang T. 2023. Widespread contribution of transposable elements to the rewiring of mammalian 3D genomes. *Nat Commun* **14:** 634. doi:10.1038/s41467-023-36364-9

CNCB-NGDC Members and Partners. 2023. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2023. *Nucleic Acids Res* **51:** D18–D28. doi:10.1093/nar/gkac1073

Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, 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. doi:10.1101/gr.082800.108

Dan J, Li M, Yang J, Li J, Okuka M, Ye X, Liu L. 2013. Roles for Tbx3 in regulation of two-cell state and telomere elongation in mouse ES cells. *Sci Rep* **3:** 3492. doi:10.1038/srep03492

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10:** giab008. doi:10.1093/gigascience/giab008

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485:** 376–380. doi:10.1038/nature11082

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29:** 15–21. doi:10.1093/bioinformatics/bts635

Du Z, Zheng H, Huang B, Ma R, Wu J, Zhang X, He J, Xiang Y, Wang Q, Li Y, et al. 2017. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547:** 232–235. doi:10.1038/nature23263

Fang C, Wang Z, Han C, Safgren SL, Helmin KA, Adelman ER, Serafin V, Basso G, Eagen KP, Gaspar-Maia A, et al. 2020. Cancer-specific CTCF binding facilitates oncogenic transcriptional dysregulation. *Genome Biol* **21:** 247. doi:10.1186/s13059-020-02152-7

Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7:** 1728–1740. doi:10.1038/nprot.2012.101

Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenkov VV. 1996. An

exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-*myc* oncogenes. *Mol Cell Biol* **16:** 2802–2813. doi:10.1128/MCB.16.6.2802

Flyamer IM, Illingworth RS, Bickmore WA. 2020. Coolpup.py: versatile pile-up analysis of Hi-C data. *Bioinformatics* **36:** 2980–2985. doi:10.1093/bioinformatics/btaa073

Fueyo R, Judd J, Feschotte C, Wysocka J. 2022. Roles of transposable elements in the regulation of mammalian transcription. *Nat Rev Mol Cell Biol* **23:** 481–497. doi:10.1038/s41580-022-00457-y

Gao Y, Liu X, Tang B, Li C, Kou Z, Li L, Liu W, Wu Y, Kou X, Li J, et al. 2017. Protein expression landscape of mouse embryos during pre-implantation development. *Cell Rep* **21:** 3957–3969. doi:10.1016/j.celrep.2017.11.111

Gassler J, Kobayashi W, Gáspár I, Ruangroengkulrith S, Mohanan A, Gomez Hernandez L, Kravchenko P, Kummecke M, Lalic A, Rifel N, et al. 2022. Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **378:** 1305–1315. doi:10.1126/science.abn7478

Gualdrini F, Polletti S, Simonatto M, Prosperini E, Pileri F, Natoli G. 2022. H3k9 trimethylation in active chromatin restricts the usage of functional CTCF sites in SINE B2 repeats. *Genes Dev* **36:** 414–432. doi:10.1101/gad.349282.121

Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, 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. doi:10.1016/j.cell.2015.07.038

Hackett JA, Surani MA. 2014. Regulatory principles of pluripotency: from the ground state up. *Cell Stem Cell* **15:** 416–430. doi:10.1016/j.stem.2014.09.015

Han S, Lee H, Lee AJ, Kim SK, Jung I, Koh GY, Kim TK, Lee D. 2021. CHD4 conceals aberrant CTCF-binding sites at TAD interiors by regulating chromatin accessibility in mouse embryonic stem cells. *Mol Cells* **44:** 805–829. doi:10.14348/molcells.2021.0224

Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, et al. 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* **43:** 630–638. doi:10.1038/ng.857

Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405:** 486–489. doi:10.1038/35013106

Harmston N, Ing-Simmons E, Tan G, Perry M, Merkenschlager M, Lenhard B. 2017. Topologically associating domains are ancient features that coincide with metazoan clusters of extreme noncoding conservation. *Nat Commun* **8:** 441. doi:10.1038/s41467-017-00524-5

Hashimoto H, Wang D, Horton JR, Zhang X, Corces VG, Cheng X. 2017. Structural basis for the versatile and methylation-dependent binding of CTCF to DNA. *Mol Cell* **66:** 711–720.e3. doi:10.1016/j.molcel.2017.05.004

Heidari N, Phanstiel DH, He C, Grubert F, Jahanbani F, Kasowski M, Zhang MQ, Snyder MP. 2014. Genome-wide map of regulatory interactions in the human genome. *Genome Res* **24:** 1905–1917. doi:10.1101/gr.176586.114

Hermant C, Torres-Padilla ME. 2021. TFs for TEs: the transcription factor repertoire of mammalian transposable elements. *Genes Dev* **35:** 22–39. doi:10.1101/gad.344473.120

Ichiyanagi T, Katoh H, Mori Y, Hirafuku K, Boyboy BA, Kawase M, Ichiyanagi K. 2021. B2 SINE copies serve as a transposable boundary of DNA methylation and histone modifications in the mouse. *Mol Biol Evol* **38:** 2380–2395. doi:10.1093/molbev/msab033

Inoue A, Akiyama T, Nagata M, Aoki F. 2007. The perivitelline space-forming capacity of mouse oocytes is associated with meiotic competence. *J Reprod Develop* **53:** 1043–1052. doi:10.1262/jrd.19064

Kaaij LJT, Mohn F, van der Weide RH, de Wit E, Bühler M. 2019. The ChAHP complex counteracts chromatin looping at CTCF sites that emerged from SINE expansions in mouse. *Cell* **178:** 1437–1451.e14. doi:10.1016/j.cell.2019.08.007

Kagda MS, Lam B, Litton C, Small C, Sloan CA, Spragins E, Tanaka F, Whaling I, Gabdank I, Youngworth I. 2023. Data navigation on the ENCODE portal. arXiv doi:10.48550/arXiv.2305.00006

Katainen R, Dave K, Pitkänen E, Palin K, Kivioja T, Välimäki N, Gylfe AE, Ristolainen H, Hänninen UA, Cajuso T, et al. 2015. CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat Genet* **47:** 818–821. doi:10.1038/ng.3335

Ke Y, Xu Y, Chen X, Feng S, Liu Z, Sun Y, Yao X, Li F, Zhu W, Gao L, et al. 2017. 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. *Cell* **170:** 367–381.e20. doi:10.1016/j.cell.2017.06.029

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12:** 996–1006. doi:10.1101/gr.229102

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37:** 907–915. doi:10.1038/s41587-019-0201-4

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20:** 278. doi:10.1186/s13059-019-1910-1

Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* **42:** 631–634. doi:10.1038/ng.600

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359. doi:10.1038/nmeth.1923

Lawson HA, Liang Y, Wang T. 2023. Transposable elements in mammalian chromatin organization. *Nat Rev Genet* **24:** 712–723. doi:10.1038/s41576-023-00609-6

Lee MT, Bonneau AR, Giraldez AJ. 2014. Zygotic genome activation during the maternal-to-zygotic transition. *Annu Rev Cell Dev Biol* **30:** 581–613. doi:10.1146/annurev-cellbio-100913-013027

Lee R, Kang MK, Kim YJ, Yang B, Shim H, Kim S, Kim K, Yang CM, Min BG, Jung WJ, et al. 2022. CTCF-mediated chromatin looping provides a topological framework for the formation of phase-separated transcriptional condensates. *Nucleic Acids Res* **50:** 207–226. doi:10.1093/nar/gkab1242

Liao Y, Smyth GK, Shi W. 2014. Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923–930. doi:10.1093/bioinformatics/btt656

Ling JQ, Li T, Hu JF, Vu TH, Chen HL, Qiu XW, Cherry AM, Hoffman AR. 2006. CTCF mediates interchromosomal colocalization between *Igf2/H19* and *Wsb1/Nf1*. *Science* **312:** 269–272. doi:10.1126/science.1123191

Liu X, Wang C, Liu W, Li J, Li C, Kou X, Chen J, Zhao Y, Gao H, Wang H, et al. 2016. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537:** 558–562. doi:10.1038/nature19362

Liu G, Wang W, Hu S, Wang X, Zhang Y. 2018. Inherited DNA methylation primes the establishment of accessible

chromatin during genome activation. *Genome Res* **28:** 998–1007. doi:10.1101/gr.228833.117

Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, Myers Z, Sud P, Jou J, Lin K, et al. 2020. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res* **48:** D882–D889. doi:10.1093/nar/gkz1062

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487:** 57–63. doi:10.1038/nature11244

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17:** 10–12. doi:10.14806/ej.17.1.200

Mei S, Qin Q, Wu Q, Sun H, Zheng R, Zang C, Zhu M, Wu J, Shi X, Taing L, et al. 2017. Cistrome data browser: a data portal for ChIP-seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res* **45:** D658–D662. doi:10.1093/nar/gkw983

Nakahashi H, Kieffer Kwon KR, Resch W, Vian L, Dose M, Stavreva D, Hakim O, Pruett N, Nelson S, Yamane A, et al. 2013. A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep* **3:** 1678–1689. doi:10.1016/j.celrep.2013.04.024

Niknafs YS, Pandian B, Iyer HK, Chinnaiyan AM, Iyer MK. 2017. TACO produces robust multisample transcriptome assemblies from RNA-seq. *Nat Methods* **14:** 68–70. doi:10.1038/nmeth.4078

Oomen ME, Hansen AS, Liu Y, Darzacq X, Dekker J. 2019. CTCF sites display cell cycle-dependent dynamics in factor binding and nucleosome positioning. *Genome Res* **29:** 236–249. doi:10.1101/gr.241547.118

Ostapcuk V, Mohn F, Carl SH, Basters A, Hess D, Iesmantavicius V, Lampersberger L, Flemr M, Pandey A, Thomä NH, et al. 2018. Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. *Nature* **557:** 739–743. doi:10.1038/s41586-018-0153-8

Patty BJ, Hainer SJ. 2021. Transcription factor chromatin profiling genome-wide using uliCUT&RUN in single cells and individual blastocysts. *Nat Protoc* **16:** 2633–2666. doi:10.1038/s41596-021-00516-2

Pinhasov A, Mandel S, Torchinsky A, Giladi E, Pittel Z, Goldsweig AM, Servoss SJ, Brenneman DE, Gozes I. 2003. Activity-dependent neuroprotective protein: a novel gene essential for brain formation. *Dev Brain Res* **144:** 83–90. doi:10.1016/S0165-3806(03)00162-7

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842. doi:10.1093/bioinformatics/btq033

Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, Nguyen N, Paten B, Zweig AS, Karolchik D, et al. 2014. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC genome browser. *Bioinformatics* **30:** 1003–1005. doi:10.1093/bioinformatics/btt637

Rhee Ho S, Pugh BF. 2011. Comprehensive genome-wide protein–DNA interactions detected at single-nucleotide resolution. *Cell* **147:** 1408–1419. doi:10.1016/j.cell.2011.11.013

Rowley MJ, Corces VG. 2018. Organizational principles of 3D genome architecture. *Nat Rev Genet* **19:** 789–800. doi:10.1038/s41576-018-0060-8

Schmidt D, Schwalie PC, Wilson MD, Ballester B, Gonçalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. 2012. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148:** 335–348. doi:10.1016/j.cell.2011.11.058

Schulz KN, Harrison MM. 2019. Mechanisms regulating zygotic genome activation. *Nat Rev Genet* **20:** 221–234. doi:10.1038/s41576-018-0087-x

Senft AD, Macfarlan TS. 2021. Transposable elements shape the evolution of mammalian development. *Nat Rev Genet* **22:** 691–711. doi:10.1038/s41576-021-00385-1

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16:** 259. doi:10.1186/s13059-015-0831-x

Shao W, Wang T. 2021. Transcript assembly improves expression quantification of transposable elements in single-cell RNA-seq data. *Genome Res* **31:** 88–100. doi:10.1101/gr.265173.120

Skene PJ, Henikoff S. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6:** e21856. doi:10.7554/eLife.21856

Skene PJ, Henikoff JG, Henikoff S. 2018. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13:** 1006–1019. doi:10.1038/nprot.2018.015

Sun D, Xi Y, Rodriguez B, Park HJ, Tong P, Meong M, Goodell MA, Li W. 2014. MOABS: model based analysis of bisulfite sequencing data. *Genome Biol* **15:** R38. doi:10.1186/gb-2014-15-2-r38

Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res* **24:** 1963–1976. doi:10.1101/gr.168872.113

Thorn GJ, Clarkson CT, Rademacher A, Mamayusupova H, Schotta G, Rippe K, Teif VB. 2022. DNA sequence-dependent formation of heterochromatin nanodomains. *Nat Commun* **13:** 1861. doi:10.1038/s41467-022-29360-y

Thybert D, Roller M, Navarro FCP, Fiddes I, Streeter I, Feig C, Martin-Galvez D, Kolmogorov M, Janoušek V, Akanni W, et al. 2018. Repeat associated mechanisms of genome evolution and function revealed by the *Mus caroli* and *Mus pahari* genomes. *Genome Res* **28:** 448–459. doi:10.1101/gr.234096.117

Trizzino M, Park Y, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, Perry GH, Lynch VJ, Brown CD. 2017. Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res* **27:** 1623–1633. doi:10.1101/gr.218149.116

Wan LB, Pan H, Hannenhalli S, Cheng Y, Ma J, Fedoriw A, Lobanenkov V, Latham KE, Schultz RM, Bartolomei MS. 2008. Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development. *Development* **135:** 2729–2738. doi:10.1242/dev.024539

Wang C, Liu X, Gao Y, Yang L, Li C, Liu W, Chen C, Kou X, Zhao Y, Chen J, et al. 2018. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat Cell Biol* **20:** 620–631. doi:10.1038/s41556-018-0093-4

Wang C, Chen C, Liu X, Li C, Wu Q, Chen X, Yang L, Kou X, Zhao Y, Wang H, et al. 2022a. Dynamic nucleosome organization after fertilization reveals regulatory factors for mouse zygotic genome activation. *Cell Res* **32:** 801–813. doi:10.1038/s41422-022-00652-8

Wang X, Wang W, Wang Y, Chen J, Liu G, Zhang Y. 2022b. Antibody-free profiling of transcription factor occupancy during early embryogenesis by FitCUT&RUN. *Genome Res* **32:** 378–388. doi:10.1101/gr.275837.121

Wei C, Jia L, Huang X, Tan J, Wang M, Niu J, Hou Y, Sun J, Zeng P, Wang J, et al. 2022. CTCF organizes inter-A compartment interactions through RYBP-dependent phase separation. *Cell Res* **32:** 744–760. doi:10.1038/s41422-022-00676-0

Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, Abraham BJ, Cohen MA, Nabet B, Buckley DL, et al. 2017. YY1 is a structural regulator of enhancer-promoter loops. *Cell* **171:** 1573–1588.e28. doi:10.1016/j.cell.2017.11.008

Xi Y, Li W. 2009. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10:** 232. doi:10.1186/1471-2105-10-232

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-seq (MACS). *Genome Biol* **9:** R137. doi:10.1186/gb-2008-9-9-r137

Zhang W, Chen Y, Yang J, Zhang J, Yu J, Wang M, Zhao X, Wei K, Wan X, Xu X, et al. 2020. A high-throughput small molecule screen identifies farrerol as a potentiator of CRISPR/Cas9-mediated genome editing. *Elife* **9:** e56008. doi:10.7554/eLife.56008

Zhao H, Sun Z, Wang J, Huang H, Kocher JP, Wang L. 2014. Crossmap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30:** 1006–1007. doi:10.1093/bioinformatics/btt730

Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H, Chen CH, Brown M, Zhang X, Meyer CA, et al. 2019. Cistrome data browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res* **47:** D729–D735. doi:10.1093/nar/gky1094

Zhu Y, Yu J, Gu J, Xue C, Zhang L, Chen J, Shen L. 2021. Relaxed 3D genome conformation facilitates the pluripotent to totipotent-like state transition in embryonic stem cells. *Nucleic Acids Res* **49:** 12167–12177. doi:10.1093/nar/gkab1069