

# Relaxed 3D genome conformation facilitates the pluripotent to totipotent-like state transition in embryonic stem cells

Yezhang Zhu<sup>1,†</sup>, Jiali Yu<sup>1,\*†</sup>, Jiahui Gu<sup>1,†</sup>, Chaoran Xue<sup>1</sup>, Long Zhang<sup>1</sup>, Jiekai Chen<sup>3</sup> and Li Shen<sup>1,2,4,\*</sup>

<sup>1</sup>The MOE Key Laboratory of Biosystems Homeostasis & Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou 310058, China, <sup>2</sup>Department of Orthopedics Surgery, the Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310058, China, <sup>3</sup>CAS Key Laboratory of Regenerative Biology, Joint School of Life Sciences, Guangzhou Institutes of Biomedicine and Health, Guangzhou Medical University, Chinese Academy of Sciences, Guangzhou 511436, China and <sup>4</sup>Hangzhou Innovation Center, Zhejiang University, Hangzhou 310058, China

Received November 21, 2020; Revised October 12, 2021; Editorial Decision October 14, 2021; Accepted November 15, 2021

## ABSTRACT

**The 3D genome organization is crucial for gene regulation. Although recent studies have revealed a uniquely relaxed genome conformation in totipotent early blastomeres of both fertilized and cloned embryos, how weakened higher-order chromatin structure is functionally linked to totipotency acquisition remains elusive. Using low-input Hi-C, ATAC-seq and ChIP-seq, we systematically examined the dynamics of 3D genome and epigenome during pluripotent to totipotent-like state transition in mouse embryonic stem cells (ESCs). The spontaneously converted 2-cell-embryo-like cells (2CLCs) exhibited more relaxed chromatin architecture compared to ESCs, including global weakening of both enhancer-promoter interactions and TAD insulation. While the former correlated with inactivation of ESC enhancers and down-regulation of pluripotent genes, the latter might facilitate contacts between the putative new enhancers arising in 2CLCs and neighboring 2C genes. Importantly, disruption of chromatin organization by depleting CTCF or the cohesin complex promoted the ESC to 2CLC transition. Our results thus establish a critical role of 3D genome organization in totipotency acquisition.**

## INTRODUCTION

In mammals, intense chromatin remodeling occurs shortly after fertilization, which involves both reprogramming of

epigenetic modifications and reorganization of chromatin architecture (1,2). This process prepares both maternal and paternal genomes for the zygotic genome activation (ZGA), which takes place at the late 1-cell and 2-cell (2C) stages in mouse embryos and is characterized by transient expression of a group of 2C-specific genes and repeats, such as *Zscan4* cluster genes and murine endogenous retrovirus with leucine tRNA (MERVL) repeats (3–5). It is noteworthy that ZGA is coupled with the acquisition of totipotency, the ability of a cell to generate all cell types in an organism, including both the embryo proper and extraembryonic tissues. In contrast, embryonic stem cells (ESCs), which are derived from the inner cell mass (ICM) of the blastocyst-stage embryos, are considered pluripotent, as they can differentiate into all three germ layers and germ cells of the embryo but rarely into extraembryonic tissues. Intriguingly, a rare dynamic subset of cells in mouse ESC cultures has been discovered to possess the ability to contribute to both embryonic and extraembryonic tissues as well, showing a totipotent-like developmental potency similar to the 2C-stage blastomeres and highly expressing 2C-specific transcripts (6). These cells are thus known as 2C-like cells (2CLCs) and have been demonstrated to be a unique model for understanding totipotency and ZGA. Nevertheless, it remains elusive what defines the chromatin state of a totipotent cell.

Proper 3D genome architecture is crucial for gene regulation. It has been reported that the higher-order chromatin structures of mouse zygotes and 2C embryos are remarkably relaxed, with weakened topologically associating domains (TADs) compared to those of later-stage embryos and somatic cells (7,8). Similar relaxed chromatin ar-

\*To whom correspondence should be addressed. Tel: +86 571 88981751; Email: li.shen@zju.edu.cn

Correspondence may also be addressed to Jiali Yu. Tel: +86 571 88208959; Email: jl.fish@zju.edu.cn

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

chitectures of early embryos were also noted in other organisms including fly, fish, and human (9–11). Moreover, when a somatic nucleus is reprogrammed to the totipotent state during somatic cell nuclear transfer (SCNT), its chromatin architecture becomes markedly relaxed as well (12). It appears that the relaxed genome architecture is a unique feature of totipotency. Indeed, the totipotent 2CLCs have also been reported to exhibit increased histone mobility and chromatin accessibility as compared to ESCs (13,14). However, the higher-order chromatin structure in 2CLCs has not been studied genome wide, and how relaxed genome architecture is functionally linked to totipotency remains elusive. In this study, we profiled the dynamics of both 3D genome and epigenome during the spontaneous pluripotent to totipotent-like state transition in mouse ESCs and defined the role of chromatin architecture relaxation in establishing totipotency.

## MATERIALS AND METHODS

### Cell lines and cell culture

The stable mouse ESC line with the MERVL::tdTomato reporter (2C-E14) were generated as previously described (6). Briefly, the MERVL::tdTomato reporter (a kind gift from Dr Samuel Pfaff) was transfected into E14 cells using Lipofectamine 2000 (11668030, Thermo Scientific), and single colonies were isolated. The 2C-CTCF-AID cell line was similarly generated by transfecting the MERVL::tdTomato reporter into the CTCF-AID cell line (15) (a kind gift from Dr Benoit Bruneau). Mouse ESCs were cultured on 0.1% gelatin-coated plates with standard LIF/serum medium containing 15% FBS (VS500T, Ausbian), 1000 U/ml mouse LIF (ESG1107, Millipore), 0.1 mM non-essential amino acids (11140, Gibco), 0.055 mM  $\beta$ -mercaptoethanol (21985023, Gibco), 2 mM GlutaMAX (35050, Gibco), 1 mM sodium pyruvate (11360, Gibco), and 100 U/ml penicillin/streptomycin (15140, Gibco). For 2C-CTCF-AID cells, 500  $\mu$ M indole-3-acetic acid (IAA, chemical analog of auxin, C3290, APExBIO) was used to induce CTCF degradation.

### Analysis of mitotic cells and cell viability

Analysis of mitotic cells was performed as previously described (16). Briefly, cells were stained with the anti-H3S10ph antibody (a kind gift from Dr Fangwei Wang) and Hoechst 33342, and measured by flow cytometry. Cell viability was measured by 7-AAD and Annexin V staining followed by flow cytometry analysis.

### sisHi-C

ESCs and 2CLCs were purified by fluorescence-activated cell sorting (FACS), and 100 000 cells were used for each sisHi-C library preparation. The sisHi-C libraries were prepared as previously reported with minor modifications (8). Briefly, the sorted cells were cross-linked with 1% formaldehyde for 10 min and quenched with 0.2 M glycine for 10 min, and then re-suspended in lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 0.5% NP-40, 0.1 mM EDTA

and 1 $\times$  proteinase inhibitor) for nucleus extraction. The extracted nuclei were further lysed in 10  $\mu$ l of 0.5% SDS solution for 10 min at 62°C, followed by adding 28  $\mu$ l of 2% Triton X-100 solution and 15-min incubation at 37°C. Genomic DNA was then digested overnight at 37°C by adding 100 U of MboI and 5  $\mu$ l of 10 $\times$  NEB buffer 2. After inactivation of MboI enzyme, cohesive ends were filled in by adding 4.5  $\mu$ l of dCTP/dGTP/dTTP mix (0.33 mM each), 3.75  $\mu$ l of biotin-14-dATP (0.4 mM) and 10 U of Klenow fragment, and incubation at 37°C for 90 min. Subsequently, 39  $\mu$ l of water, 12  $\mu$ l of 10 $\times$  T4 ligase reaction buffer, 7  $\mu$ l of 10% Triton X-100 solution, 1.2  $\mu$ l of 10 mg/ml BSA and 400 U of T4 DNA ligase were added for proximity ligation (5.5 h at room temperature). After crosslinking reversal, genomic DNA was purified and sheared to a length of  $\sim$ 400 bp, and point ligation junctions were pulled down with Dynabeads MyOne Streptavidin C1 (65001, Thermo Fisher) for sequencing library preparation. The final libraries were sequenced on an Illumina HiSeq X Ten platform in the paired-end mode.

### ULI-NChIP-seq and CUT&Tag

For ULI-NChIP-seq, 10 000 or 40 000 cells were used per reaction. The ULI-NChIP procedure was performed as previously described (17,18). Briefly, 2  $\mu$ g of H3K27ac antibody (39133, Active Motif), H3K4me1 antibody (39297, Active Motif), H3K4me3 antibody (9727, Cell Signaling Technology), H3K27me3 antibody (9733, Cell Signaling Technology) or H3K9me3 antibody (39161, Active Motif) was used for each immunoprecipitation reaction. The sequencing libraries were generated using the NEBNext Ultra II DNA Library Prep Kit for Illumina (E7645, New England Biolabs) following the manufacturer's instructions. The CUT&Tag assays for CTCF were performed using 50 000 sorted ESCs and 2CLCs as previously described (19). Two or three replicates were performed. Barcoded libraries were pooled and sequenced on an Illumina HiSeq X Ten platform in the paired-end mode.

### ATAC-seq

The ATAC-seq libraries were prepared as previously described with minor modifications (20). Briefly, samples were lysed in lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl<sub>2</sub> and 0.1% Igepal CA-630) for 10 min on ice to prepare the nuclei. Immediately after lysis, nuclei were spun at 500 g for 5 min to remove the supernatant. Nuclei were then incubated with the Tn5 transposome and tagmentation buffer at 37°C for 30 min (TD501, Vazyme Biotech). After the tagmentation, the products were purified with 2.0 $\times$  VAHTS DNA Clean Beads (N411, Vazyme Biotech). PCR was performed to amplify the library for 14 cycles using the following PCR conditions: 72°C for 3 min; 98°C for 30 s; and thermocycling at 98°C for 15 s, 60°C for 30 s and 72°C for 3 min; following by 72°C 5 min. After the PCR reaction, libraries were purified with the 1.2 $\times$  VAHTS DNA Clean Beads. Barcoded libraries were pooled and sequenced on an Illumina HiSeq X Ten platform in the paired-end mode.

### ChIP-seq, CUT&Tag and ATAC-seq data analysis

Raw reads were trimmed to 50 bp and aligned to the mouse genome (mm9) using bowtie2 (version 2.3.4.1) with default parameters. Unmapped reads and PCR duplicates were removed. Peaks were called using MACS2 (version 2.1.1.20160309) with parameters '-q 0.05 -nomodel -nolambda -broad -extsize 300 -B -SPMR -g mm'. Average intensity profiles were generated using deepTools (version 2.5.4).

### RNA-seq and data analysis

RNA-seq libraries were prepared using the Smart-seq2 method as previously described (21). Barcoded libraries were pooled and sequenced on an Illumina HiSeq X Ten platform in the paired-end mode. Raw reads were trimmed to 50 bp and mapped to the mouse genome (mm9) using TopHat (version 2.1.1) with default parameters. Only uniquely mapped reads were kept for downstream analysis. Gene counts were generated using HTSeq-count (version 0.9.1). For each sample, the gene count matrices were merged together and then the 'Trimmed Mean of M values' normalization (TMM) method (edgeR package in Bioconductor, version 3.5.0) was used to calculate the normalized expression. *P* values were generated in edgeR (exactTest) and then adjusted to the false discovery rate (FDR). The significantly up-regulated and down-regulated genes were identified as genes with  $FDR < 0.05$ ,  $\log_2(\text{fold change})$  greater than 1.5 and less than -1.5, respectively.

### Hi-C data analysis

Raw reads were first processed with HiC-Pro (version 2.11.1). Briefly, raw reads were trimmed to 50 bp and then aligned to the mouse genome (mm9) using the Bowtie 2 end-to-end algorithm and '-very-sensitive' option. To rescue the chimeric fragments spanning ligation junctions, the ligation site was detected and the 5' fraction of the reads was aligned back to the reference genome. Unmapped reads, multiple mapped reads and singletons were then discarded. Pairs of aligned reads were then assigned to MboI restriction fragments. Read pairs from uncut DNA, self-circle ligation and PCR artefacts were filtered out, and the valid read pairs involving two different restriction fragments were used to build the contact matrix. Valid read pairs were then binned at a specific resolution by dividing the genome into bins of equal size. The binned interaction matrices were then normalized using the iterative correction method in HiC-Pro with default parameters to correct for biases such as GC content, mappability and effective fragment length. For each chromosome, we obtained the expected Hi-C contact values by calculating the average contact intensity for all loci at a certain distance. We then transformed the normalized Hi-C matrix into an observed/expected (O/E) matrix by dividing each normalized observed value by its corresponding expected value. We identified A and B compartments and TADs from Hi-C data using cworld (Dekker lab, <https://github.com/dekkerlab/cworld-dekker>). To calculate PC1 values, we used the "matrix2compartment.pl" script using contact maps binned at 500 kb resolution as input.

Bins with positive or negative PC1 values were defined as A or B compartment, respectively. The compartment strength was then calculated as  $(AA + BB)/2AB$  as previously described (22). To calculate insulation scores, we used the 'matrix2insulation.pl' script with default parameters at 10 kb resolution. Insulation scores were further normalized by dividing each bin's insulation score by the average scores in the nearest 2 Mb region,  $\log_2$ -transformed, and multiplied by -1. The 'insulation2tads.pl' script was used with default parameters to identify TADs. Stable TADs between two samples were identified as previously described (7). Directionality index (DI) scores were calculated using a previously described pipeline (23). The relative frequency of inter-TAD interactions was calculated from the number of interactions between two TADs divided by the total number of cis-interactions in the corresponding chromosome. Aggregate Hi-C contact maps were generated by extracting subsets from the OE matrix (these can be either single regions along the diagonal, or region pairs corresponding the matrix segments off the diagonal) and averaging all resulting sub-matrices. If the sub-matrices were of different size, they were interpolated to a fixed size. TAD strength was calculated as the mean of values in the OE matrix in the TAD region. Loop strength was calculated as the mean of values in the OE matrix in the nearest 10 regions of both anchors. Global contact decay curves were plotted as previously described (24). In order to better visualize differences between conditions, we calculated the distribution of the Hi-C contacts as a simple contact probability (sum of the observed counts per  $\log_2$  bin, divided by total observed contacts, without normalizing for the bin size).

### siRNA transfection

ESCs were transfected with siRNAs (Gene Pharma) using the DharmaFECT 1 (T-2001, Dharmacon) reagent following the manufacturer's instructions. The control siRNA (siNC) was derived from sequences in the *Caenorhabditis elegans* genome and does not target any mammalian genes. Cells were collected for RNA extraction 2 days after transfection.

### RNA isolation, and RT-qPCR

Total RNA was isolated using RNeasy Plus Mini Kit (74134, Qiagen), cDNA was generated using HiScript II Q RT Super Mix with gDNA wiper (R223, Vazyme Biotech), and qPCR was performed using the AceQ qPCR SYBR Green Master Mix (Q111, Vazyme Biotech). Relative quantification was performed using the comparative Ct method with normalization to *Actin*. Primer sequences are available in Supplementary Table S2.

### Statistical analysis

Statistical analyses were performed with the R software. Statistics were calculated using the Wilcoxon signed rank test unless indicated otherwise.

## RESULTS

### Global weakening of the 3D genome conformation during ESC to 2CLC transition

To investigate the 3D genome architecture in 2CLCs, we generated a mouse ESC line containing a tdTomato transgene driven by the MERVL promoter (MERVL::tdTomato), which has been established as a faithful indicator of the 2C-like state (6). Consistent with previous reports, about 0.4% of tdTomato+ cells (i.e. 2CLCs) were observed in this ESC line by flow cytometry (Supplementary Figure S1A). These 2CLCs were viable and showed a similar percentage of mitotic cells as compared to ESCs (Supplementary Figure S1B and C). We then purified tdTomato+ 2CLCs and tdTomato- ESCs and performed RT-qPCR and RNA-seq analyses to validate the fidelity of the purified 2CLCs. As expected, 2C-specific transcripts such as *Zscan4* family genes, *Dux*, *Dub1*, *Zfp352* and MERVL repeats were highly expressed in 2CLCs (Supplementary Figure S1D–G), and RNA-seq revealed more upregulated genes than down regulated genes (3–5). Importantly, the 1044 up-regulated genes and 212 down-regulated genes respectively showed higher and lower expression in 2C embryos than in the ICM of blastocyst-stage embryos (Supplementary Figure S1H). Pathway analyses also revealed an enrichment of pluripotent genes in the down-regulated genes (Supplementary Figure S1I and J), consistent with a previous report that down-regulation of pluripotent genes occurred in the first step of 2C-like transition (25).

After validating the purified cells, we carried out small-scale in situ Hi-C (sisHi-C) (8) to determine the chromatin conformation landscapes in 2CLCs and ESCs. About 200 million valid read pairs were obtained for both 2CLCs and ESCs, and the biological replicates were highly correlated (Supplementary Table S1). We first examined large-scale chromosome folding by examining active (A) and inactive (B) compartmentalization of the genome. As shown by contact maps and principal component 1 (PC1) values (26), A/B compartments are largely maintained during ESC to 2CLC transition (Figure 1A and B). Nevertheless, we observed significant decrease in the strength of compartmentalization in 2CLCs compared to ESCs (Figure 1C). Contacts between B compartments decreased and interaction frequency between A/B compartments increased (Supplementary Figure S2A and B). In addition, we identified 7.9% of genomic intervals with A-to-B (1.0%) or B-to-A (6.9%) compartment switching during ESC to 2CLC transition (Supplementary Figure S2C). As expected, total transcription within A-to-B switching compartments was down-regulated, while B-to-A switching compartments showed increased transcription during ESC to 2CLC transition (Supplementary Figure S2D). Although we found a few 2C-specific genes such as the *Tdpoz* family genes in the B-to-A switching compartments (Supplementary Figure S2E), it is worth noting that only a small fraction (85/1044, 8.1%) of 2C-specific genes manifested B-to-A switching (Supplementary Figure S2F), indicating that compartment switching is not a general requirement for the activation of 2C-specific genes in 2CLCs. Interestingly, genomic intervals showing A-to-B switching are slightly enriched for previously identified

ESC enhancers (27) (Supplementary Figure S2G), implying inactivation of ESC enhancers in 2CLCs.

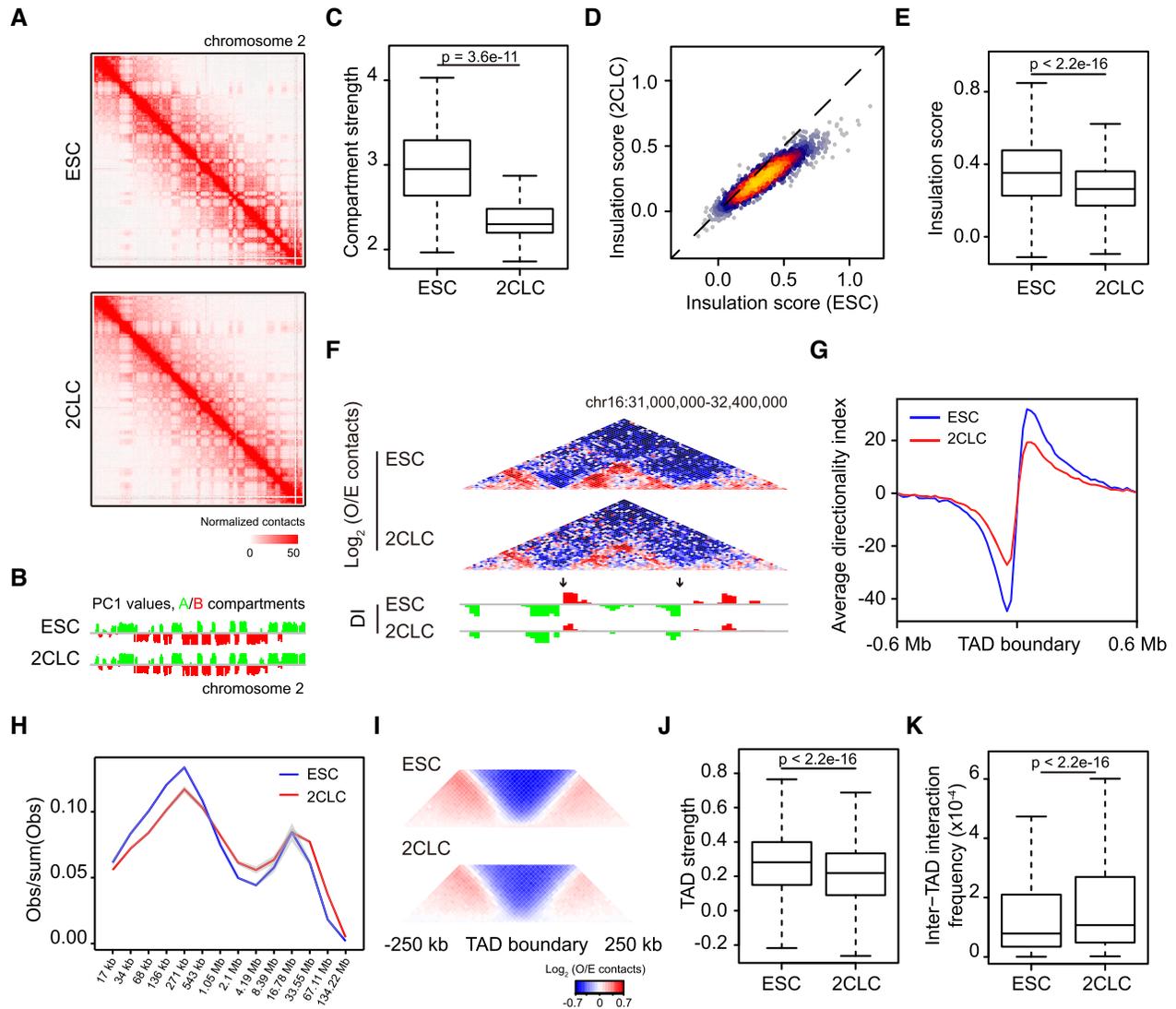
We next asked whether the local insulation of TADs was reprogrammed during ESC to 2CLC transition. Examination of insulation scores (28) at the previously reported ESC TAD boundaries (23) revealed significant attenuation of TAD insulation in 2CLCs as compared to ESCs (Figure 1D and E; Supplementary Figure S2H), which is further supported by genome-wide decrease of absolute directionality index (23) in 2CLCs (Figure 1F and G). In line with these observations, global contact decay curves showed reduced frequencies of short- to intermediate-range interactions (<500 kb) but increased frequencies of long-range interactions (>1 Mb) in 2CLCs as compared to ESCs (Figure 1H). Indeed, 2CLCs exhibited weaker TAD strength and higher frequencies of inter-TAD interactions (Figure 1I–K). Despite the global reduction of TAD insulation, the numbers of TADs called in 2CLCs ( $n = 3674$ ) and ESCs ( $n = 3625$ ) were comparable, of which the vast majority ( $n = 2957$ ) were stable (Supplementary Figure S2I), supporting the overall conserved TADs among different cell types (23,29). Taken together, our observations suggested a genome-wide weakening of TAD insulation in 2CLCs instead of specific gain or loss of TADs.

### Transcription of 2C-specific genes and repeats did not establish local chromatin insulation in 2CLCs

Active transcription was reported to be frequently correlated with chromatin insulation (30,31). Given that 2CLCs highly express a set of 2C-specific genes and the MERVL repeats, we asked whether active transcription at these loci could establish novel insulation in 2CLCs. However, we did not observe any increase in insulation scores at the promoters of up-regulated genes in 2CLCs, indicating no significant changes in chromatin insulation around 2C genes (Supplementary Figure S2J). Similarly, the highly expressed MERVL repeats also did not cause increased local chromatin insulation in 2CLCs (Supplementary Figure S2K). Thus, our results support previous observations that transcription is not sufficient to cause chromatin insulation (30).

### Loss of enhancer-promoter interactions and down-regulation of pluripotent genes during ESC to 2CLC transition

Reduced TAD insulation indicates loss of chromatin loops. Indeed, we observed in 2CLCs a mild reduction in the strength of previously identified chromatin loops, most of which are anchored at TAD boundaries (29) (Figure 2A and B). Given that chromatin loops not only link TAD boundaries but also tether enhancers and promoters, we specifically analyzed interactions between reported ESC enhancer-promoter pairs (32). This analysis revealed that enhancer-promoter loops were also weakened in 2CLCs (Figure 2C and D). Furthermore, reanalysis of published Hi-C datasets showed reduced strength of TADs and both types of chromatin loops in 2C embryos as compared to ICM (Supplementary Figure S2L–N). These observations suggested that weakened chromatin loops, both between TAD boundaries and between enhancer-promoter pairs, is a unique feature of totipotency. Interestingly, the weak-



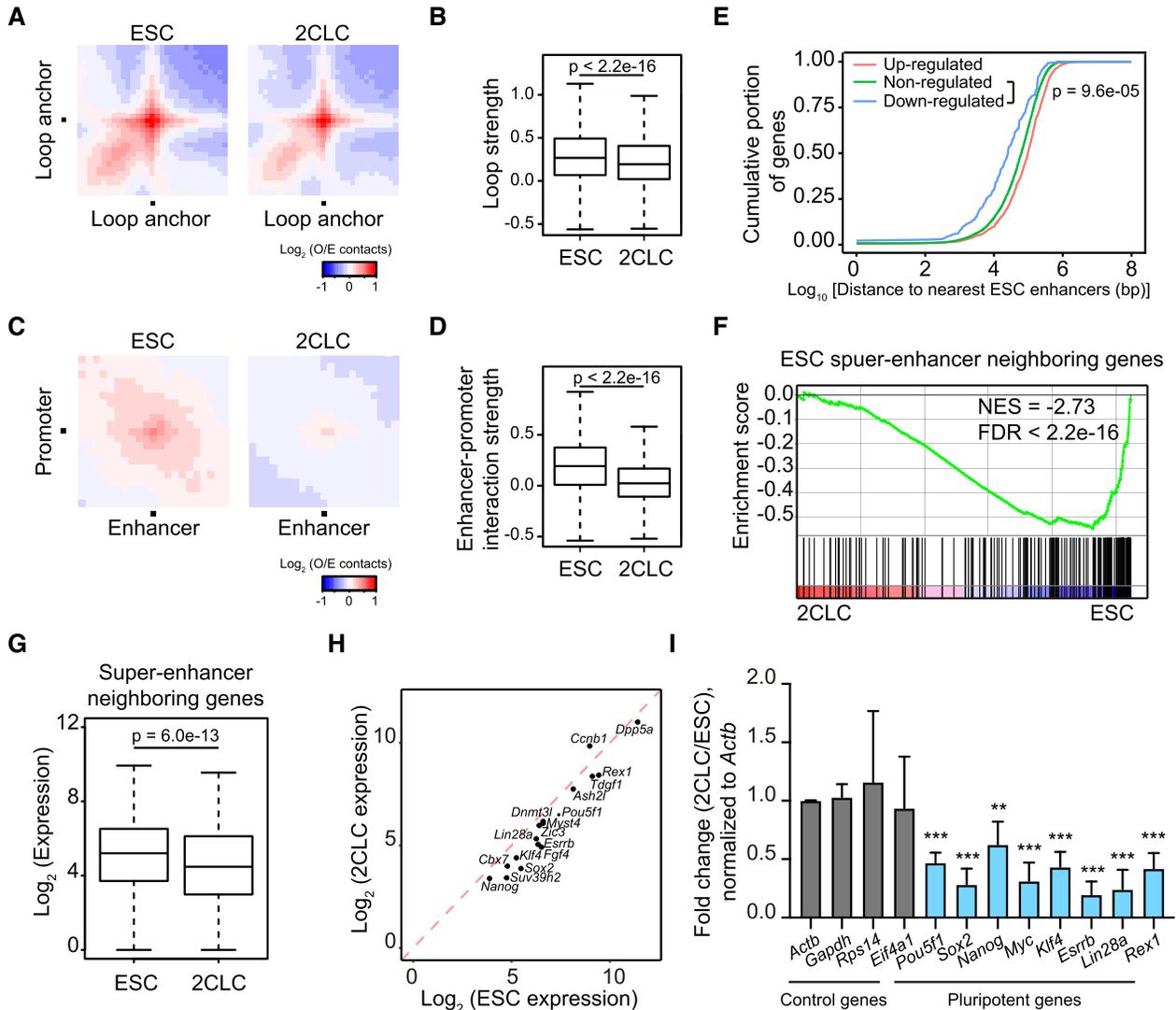
**Figure 1.** Global weakening of the 3D genome conformation during ESC to 2CLC transition. (A) Hi-C contact maps at 250-kb resolution across the entire chromosome 2. (B) PC1 values at 500-kb resolution across the entire chromosome 2. (C) Box plot showing compartment strength in ESCs and 2CLCs. (D) Scatter plot comparing insulation scores at ESC TAD boundaries (23) between ESCs and 2CLCs. (E) Box plot showing the insulation score at TAD boundaries. Note that higher score denotes higher insulation potential. (F) Genome browser shot showing Hi-C contacts (top) and directionality index (DI) (bottom) at 20-kb resolution. Arrows indicate TAD boundaries. O/E, observed/expected. (G) Average DI in a 0.6 Mb region centered on TAD boundaries. (H) Contact probability as a function of the genomic distance in logarithmic bins. Lines represent means of biological replicates; edges of semi-transparent ribbons represent individual data points of the two biological replicates. (I) Aggregate Hi-C contact maps around TAD boundaries. (J) Box plot comparing TAD strengths in ESCs and 2CLCs. (K) Box plot showing inter-TAD interaction frequencies in ESCs and 2CLCs.

ened enhancer-promoter interactions appeared to correlate with a reduction in the ESC enhancer activity. Indeed, we found that down-regulated genes in 2CLCs were located closer to ESC enhancers (Figure 2E) and were enriched for ESC super-enhancer neighboring genes (Figure 2F). In turn, super-enhancer neighboring genes also displayed significantly lower expression in 2CLCs (Figure 2G). In line with the critical role of ESC super-enhancers in regulating pluripotent genes (33), down-regulated genes in 2CLCs were enriched for pluripotency (Supplementary Figure S1J). Almost all pluripotent genes we examined were down-regulated in 2CLCs, including *Pou5f1*, *Sox2*, *Nanog*, *Myc*, *Klf4*, *Esrrb*, *Lin28a* and *Rex1* (Figure 2H and I), cor-

relating with weakened enhancer-promoter interactions in 2CLCs (Supplementary Figure S3A and B).

### Inactivation of ESC enhancers and formation of new enhancers in 2CLCs

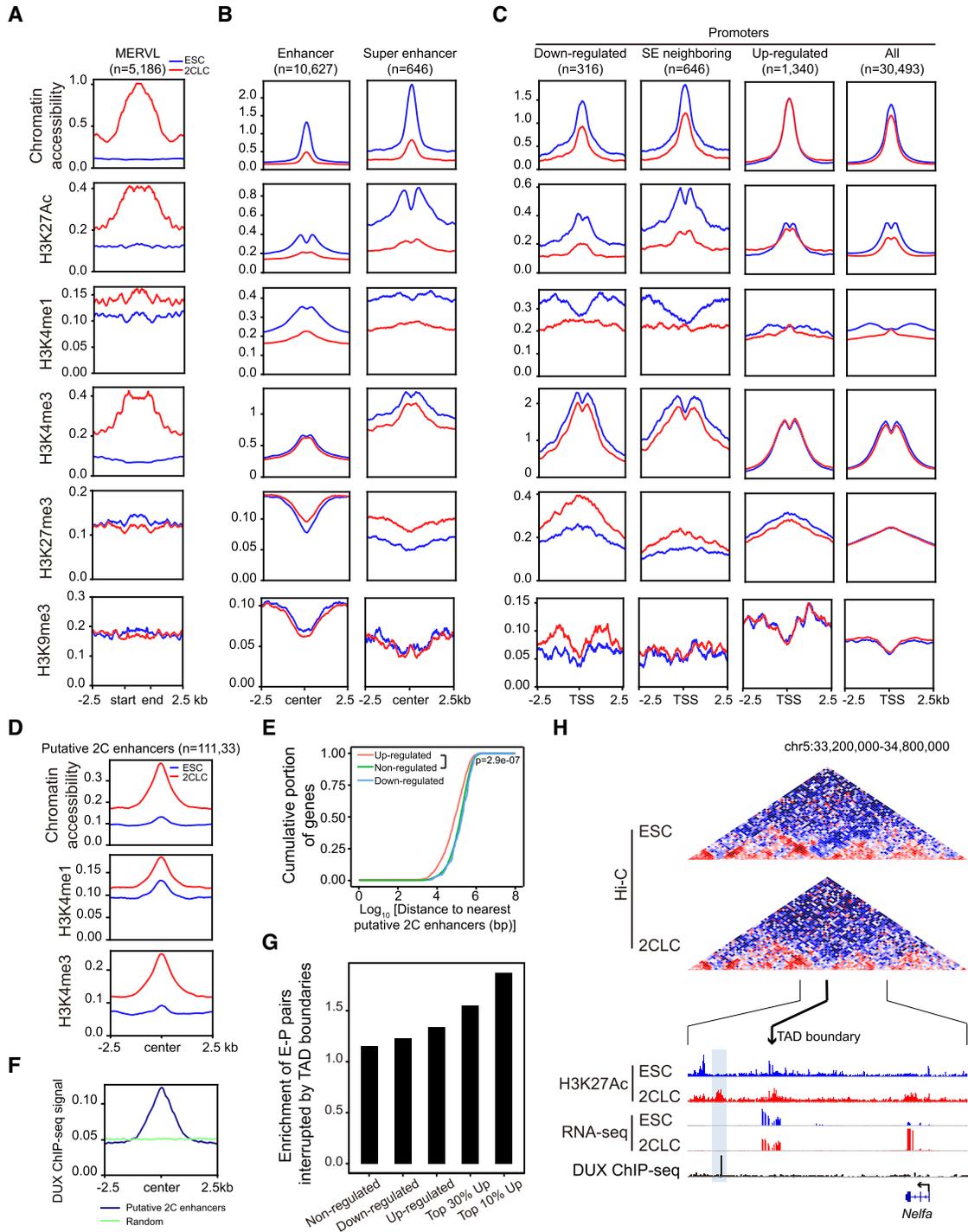
Chromatin accessibility is a strong indicator of the regulatory activity of enhancers (34). To directly examine whether weakened enhancer-promoter interactions in 2CLCs are associated with a reduction in the ESC enhancer activity, we carried out ATAC-seq analysis (20) for both ESCs and 2CLCs. As expected, chromatin accessibility around MERVL repeats was strongly induced in 2CLCs (Fig-



**Figure 2.** Loss of enhancer-promoter interactions and down-regulation of pluripotent genes in 2CLCs. (A) Aggregate Hi-C contact maps between pairs of loop anchors (29). (B) Box plot showing loop strengths in ESCs and 2CLCs. (C) Aggregate Hi-C contact maps between ESC enhancer-promoter pairs (32). (D) Box plot showing enhancer-promoter interaction strengths in ESCs and 2CLCs. (E) Down-regulated genes in 2CLCs tend to locate closer to ESC enhancers. Shown are cumulative frequency curves comparing the distances between transcription start sites (TSS) of non-/up-/down-regulated genes and their nearest ESC enhancers.  $P$  value was calculated using two-sided Kolmogorov-Smirnov test. (F) Gene set enrichment analysis indicating that down-regulated genes in 2CLCs were enriched for ESC super-enhancer neighboring genes. Red, up-regulated genes in 2CLCs; blue, down-regulated genes in 2CLCs. (G) Box plot comparing the expression of super-enhancer neighboring genes in ESCs and 2CLCs. (H) Scatter plot comparing expression of pluripotent genes in ESCs and 2CLCs. (I) Relative expression levels of pluripotent genes in 2CLCs versus ESCs by RT-qPCR. Data are normalized to *Actin* and are presented as mean  $\pm$  SD, \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (multiple  $t$  tests).

ure 3A), consistent with the activation of MERV1 during ESC to 2CLC transition. In contrast, meta-analysis at ESC enhancers and super-enhancers revealed a substantial decrease of chromatin accessibility in 2CLCs (Figure 3B), indicating inactivation of ESC enhancers, particularly super-enhancers, during the pluripotent to totipotent state transition. In agreement with this observation, promoters of ESC super-enhancer neighboring genes and down-regulated genes also displayed reduced chromatin accessibility (Figure 3C). Interestingly, promoters of up-regulated genes did not show increased accessibility (Figure 3C), suggesting that these genes may have already existed in a primed state in ESCs. To further confirm that ESC en-

hancers were silenced during ESC to 2CLC transition, we performed ultra-low-input native ChIP-seq (ULI-NChIP-seq) (17,18) and compared genome-wide histone modifications between ESCs and 2CLCs, including H3K27Ac, H3K4me1, H3K4me3, H3K27me3 and H3K9me3 (Figure 3A-C; Supplementary Figure S3C and D; Supplementary Table S1). MERV1 repeats exhibited remarkable increase in H3K27Ac, H3K4me1, and H3K4me3 in 2CLCs (Figure 3A), consistent with their activation in 2CLCs. However, ESC enhancers and super-enhancers displayed a substantial decrease in both H3K27Ac and H3K4me1 (Figure 3B), confirming the inactivation of ESC enhancers during ESC to 2CLC transition. It is worth noting that ESC enhancers



**Figure 3.** Inactivation of ESC enhancers and formation of new enhancers in 2CLCs. (A–C) Average chromatin accessibility, H3K27Ac, H3K4me1, H3K4me3, H3K27me3 and H3K9me3 signals around MERVL (A), ESC enhancers/super-enhancers (B), and promoters of the indicated genes (C) in ESCs and 2CLCs. SE, super-enhancer. (D) Average chromatin accessibility, H3K4me1 and H3K4me3 signals around putative 2C enhancers. (E) Up-regulated genes in 2CLCs tend to locate closer to putative 2C enhancers. Shown are cumulative frequency curves comparing the distances between TSS of non-/up-/down-regulated genes and their nearest putative 2C enhancers. *P* value was generated using two-sided Kolmogorov–Smirnov test. (F) Average DUX ChIP-seq signals around putative 2C enhancers. Equal number of random regions were used as the control. (G) Bar plot showing that up-regulated genes tend to have ESC TAD boundaries separating them from neighboring putative 2C enhancers. Values represent enrichment of boundary presence between the promoters of indicated genes and their nearest putative 2C enhancers. Boundary presence between the same promoters and the symmetrical sites of their nearest putative 2C enhancers serves as the control. (H) Genome browser snapshot showing a TAD boundary between the promoter of *Nelfa* and a putative 2C enhancer (highlighted). Shown are Hi-C contact maps at 20-kb resolution (top) and genome browser tracks of H3K27Ac ChIP-seq, RNA-seq and DUX ChIP-seq signals (bottom). The thicker black line with an arrowhead indicates the weakened TAD boundary, and the two thinner black lines show the region that was zoomed in.

are also inactive in 2C embryos and are established gradually during preimplantation development as revealed by our reanalysis of published ATAC-seq and ChIP-seq data of early embryos (Supplementary Figure S4A).

Among the five histone modifications we profiled, H3K27Ac peaks displayed the most notable difference in genome-wide distribution between ESC and 2CLCs, with more peaks enriched in distal intergenic regions in 2CLCs (Supplementary Figure S3D). Examination of these 2CLC-specific distal H3K27Ac peaks ( $n = 11,133$ ) not only revealed increases of chromatin accessibility, H3K27Ac, H3K4me1 and H3K4me3 during ESC to 2CLC transition (Figure 3D), but also high levels of chromatin accessibility and H3K27Ac in 2C embryos (Supplementary Figure S4A), suggesting the formation of new enhancers (i.e. putative 2C enhancers) at these regions during totipotency acquisition. Indeed, up-regulated genes in 2CLCs were located closer to these putative 2C enhancers in the genome (Figure 3E). By comparing with published ChIP-seq data of DUX, an important transcription factor in 2CLCs (35–37), we found strong DUX signals at these putative 2C enhancers, suggesting that they may arise from DUX binding (Figure 3F). Interestingly, the most up-regulated genes during the 2C-like transition tend to have ESC TAD boundaries separating them from putative 2C enhancers nearby (Figure 3G), suggesting that weakened TAD insulation may potentially facilitate contacts between the putative 2C enhancers activated by sporadically expressed DUX in ESCs and the promoters of their neighboring 2C genes, as exemplified by *Nelfa*, an early driver of the 2C-like state (38) (Figure 3H). However, the interaction between *Nelfa* and the DUX-occupied putative enhancer nearby could not be readily detected in our analysis, which may require Hi-C data of much higher resolution. Taken together, inactivation of ESC enhancers, formation of putative 2C enhancers, and weakening of TAD boundaries may function together to fully activate the totipotent-like transcriptional program in 2CLCs.

### Depletion of CTCF or cohesin complex facilitates the ESC to 2CLC transition

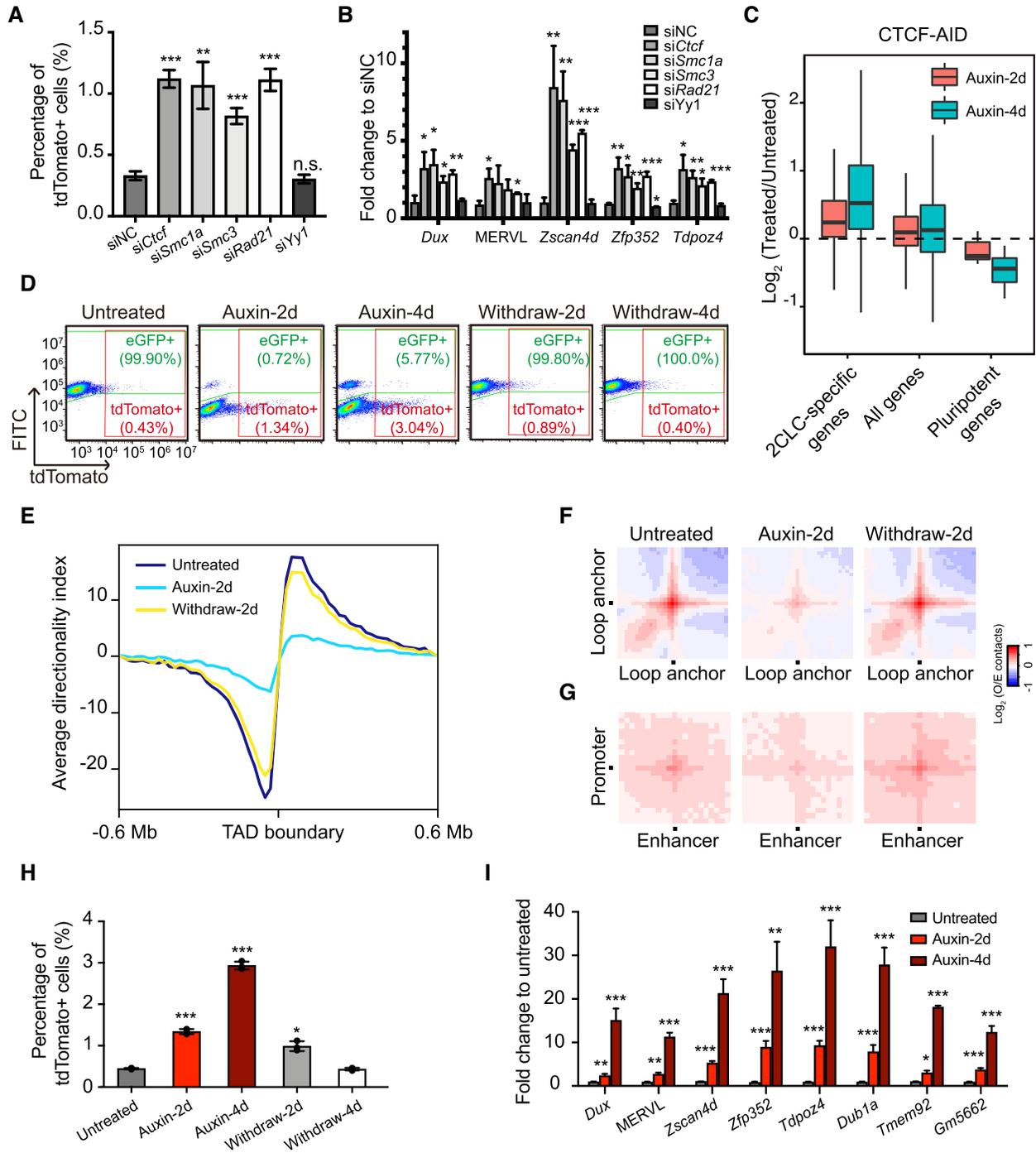
Our results have demonstrated that 2CLCs possess a relaxed chromatin architecture, including weakened TAD boundary loops and enhancer-promoter loops. While reduced enhancer-promoter loops correlates with down-regulation of ESC enhancers and pluripotent genes, weakened TAD insulation may facilitate the activation of 2C genes by sporadically activated 2C enhancers (Supplementary Figure S4B). We thus hypothesized that disruption of chromatin loops by depletion of CTCF or the cohesin complex would facilitate the ESC to 2CLC transition. To directly test this, we knocked down *Ctcf* and core cohesin complex genes including *Smc1a*, *Smc3* and *Rad21* individually with small interfering RNAs (siRNAs) in the MERVL::tdTomato ESCs (Supplementary Figure S5A). As a control, we also knocked down *Yy1*, a structural regulator that was reported to mediate only enhancer-promoter interactions but not TAD boundary loops (39). Supporting our hypothesis, knock-down of *Ctcf*, *Smc1a*, *Smc3* and *Rad21*, but not *Yy1*, significantly increased the fraction of tdTomato+ cells (Fig-

ure 4A, Supplementary Figure S5B), as well as the expression of MERVL and 2C-specific genes including *Dux*, *Zscan4d*, *Zfp352* and *Tdpoz4* (Figure 4B), demonstrating an enhanced ESC to 2CLC transition upon disruption of all chromatin loops rather than enhancer-promoter loops alone. We further examined the effect of CTCF and cohesin removal on all 2C genes by reanalyzing published RNA-seq datasets of ESCs with acute protein degradation of CTCF or RAD21 (12,15). This analysis also showed that depletion of either CTCF or the cohesin complex could result in activation of 2C genes and repression of pluripotent genes (Figure 4C, Supplementary Figure S5C).

We further constructed an ESC line (2C-CTCF-AID) containing both the MERVL::tdTomato reporter and an auxin-inducible degron (AID) system targeting CTCF. Endogenous CTCF in this cell line is also fused with an eGFP tag to indicate the protein level of CTCF (15). As shown by flow cytometry analysis, 2 days of auxin treatment led to a near complete depletion of the CTCF protein, which can be fully restored 2 days after auxin withdrawal (Figure 4D). Reanalyzing published Hi-C data (15) confirmed weakened TAD insulations together with reduced chromatin loops upon 2 days of auxin treatment (Figure 4E–G). Similar to *Ctcf* knockdown (Figure 4A and B), the percentage of tdTomato+ 2CLCs in the culture (Figure 4D and H), as well as the expression of MERVL and 2C-specific genes (Figure 4I), significantly increased upon CTCF degradation. It is noteworthy that CTCF-depletion-induced increase in ESC to 2CLC transition is reversible, as the percentage of tdTomato+ 2CLCs returned to the initial state after auxin withdrawal and CTCF recovery (Figure 4D and H). We next purified ESCs and 2CLCs in the cells treated with auxin for 2 days and performed RNA-seq analysis. These cells are both viable (Supplementary Figure S5D), comparable to untreated ESCs and 2CLCs (Supplementary Figure S1B). Our analysis showed similar transcriptomes of 2CLCs in the presence and absence of CTCF, both with up-regulation of 2C-specific genes and down-regulated pluripotent genes as compared to ESCs (Supplementary Figure S5E), demonstrating that 2CLCs converted in the absence of CTCF is indeed in a 2C-like state. Taken together, our results strongly suggested that the higher-order chromatin structure in ESCs is an impediment to the 2C-like state transition.

## DISCUSSION

In addition to epigenetic modifications, 3D chromatin architecture has been increasingly recognized as another regulatory layer of gene expression, which is tightly linked to cell identity. ESC differentiation has been shown to coincide with formation or strengthening of TAD boundary loops, indicating a relatively relaxed chromatin architecture in ESCs compared to differentiated cells (30,40). Furthermore, mouse zygotes and 2C embryos also display more relaxed chromatin architectures compared to later-stage embryos, with much weaker TAD boundaries (7,8). Similar relaxed chromatin architectures of early embryos were also observed in other organisms including fly, fish and human (9–11). Therefore, high developmental potency appears to be associated with relaxed higher-order chromatin struc-



**Figure 4.** Depletion of CTCF or cohesin facilitates ESC to 2CLC transition. (A) The percentage of 2CLCs upon knockdown of *Ctcf*, *Smc1a*, *Smc3*, *Rad21*, or *Yy1*. Data are presented as mean  $\pm$  SD,  $n = 3$ . \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (multiple  $t$  tests). (B) Relative expression levels of 2C-specific transcripts after knocking down *Ctcf*, *Smc1a*, *Smc3*, *Rad21*, or *Yy1*. Data are presented as mean  $\pm$  SD,  $n = 3$ . \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (multiple  $t$  tests). (C) Box plot showing log<sub>2</sub> fold change of indicated groups of genes upon acute depletion of the CTCF protein in ESCs by auxin-inducible degron (AID). Analyses were performed using a published RNA-seq dataset (15). (D) Representative FACS analysis of 2C-CTCF-AID cells treated with or without auxin. Note that endogenous CTCF in this cell line is fused with an eGFP tag to indicate CTCF protein levels. (E) Average DI in a 0.6 Mb region centered at TAD boundaries. (F) Aggregate Hi-C contact maps between pairs of loop anchors. (G) Aggregate Hi-C contact maps between ESC enhancer-promoter pairs. (H) The percentages of 2CLCs at indicated time points after auxin treatment or withdrawal. Data are presented as mean  $\pm$  SD,  $n = 3$ . \* $P < 0.05$ , \*\*\* $P < 0.001$  (multiple  $t$  tests). (I) Relative expression levels of 2C-specific transcripts upon auxin-induced CTCF degradation. Data are presented as mean  $\pm$  SD,  $n = 3$ . \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (multiple  $t$  tests).

ture. In this study, using the pluripotent to totipotent-like transition in mouse ESCs as a model, we defined the role of genome architecture relaxation in establishing totipotency. This spontaneous transition represents a simplified reprogramming process in totipotency acquisition, which is valuable to investigate the functional contribution of one single factor among the complex developmental events. But the scarcity of spontaneous 2C-like transition in the ESC culture limits such investigations. By using low-input technologies, we presented genome-wide chromatin contact maps of totipotent-like 2CLCs and pluripotent ESCs, together with their transcriptome, chromatin accessibility, and histone modification profiles. This rich dataset has allowed us to investigate the dynamics of chromatin state and gene expression during the pluripotent to totipotent-like state transition.

We revealed that chromatin architecture of ESCs became more relaxed during the spontaneous 2C-like transition, with globally weakened chromatin loops. This included not only TAD boundary loops but also chromatin loops between ESC enhancers particularly super-enhancers and the promoters of their neighboring genes, correlating with ESC enhancer inactivation and transcriptional down-regulation of many pluripotent genes controlled by ESC super-enhancers. The down-regulation of pluripotent genes in 2CLCs was previously reported to take place only at the protein level but not the mRNA level (6,14). Nevertheless, a recent single-cell analysis of DUX-induced ESC to 2CLC transition showed a two-step reprogramming process with transcriptional inactivation of pluripotent genes as the first step (25), in agreement with our findings. The down-regulation of pluripotent genes at both mRNA and protein levels suggested a complex regulation of gene expression during ESC to 2CLC transition. Of note, our 20 kb resolution data did not allow us to perform robust *de novo* calling of loops, preventing detailed analysis of ESC- and 2CLC-specific loops. Also, it remains to be elucidated how chromatin relaxation occurs in spontaneously converted 2CLCs. As shown in Supplementary Figure S5F, none of the genes involved in genome organization that we examined were down-regulated in 2CLCs, at least at the transcriptional level. We further examined genome-wide CTCF binding for both ESCs and 2CLCs using CUT&Tag (19). This analysis showed that CTCF similarly bound to known ESC CTCF binding sites and insulator elements in both ESCs and 2CLCs, with 2CLCs showing a slightly decreased enrichment (Supplementary Figure S5G). While this observation may partially account for the weakened 3D genome conformation in 2CLCs, further studies are needed to understand the regulation and function of CTCF in the spontaneous ESC to 2CLC transition.

We further demonstrated a causal relationship between relaxed chromatin architecture and totipotency acquisition, by showing that disruption of 3D organization of chromatin in ESCs could facilitate 2C-like transition. We also observed formation of putative 2C enhancers during ESC to 2CLC transition. Importantly, up-regulated genes during the 2C-like transition tend to have ESC TAD boundaries separating them from putative 2C enhancers nearby. Therefore, disruption of TAD boundary loops in ESCs may facilitate the contacts between putative 2C enhancers that are likely spo-

radically activated in ESCs and the promoters of nearby 2C genes, thus promoting their expression, which in turn fully activates the 2C-like transcriptional program. Consistently, a recent study showed that cohesin pre-depletion in donor cells could enhance SCNT efficiency at least partially through up-regulation of the minor ZGA genes in donor cells, many of which are also 2C-specific genes (12).

In sum, we demonstrated that the spontaneous pluripotent to totipotent-like state transition in mouse ESCs not only coincides with global weakening of 3D genome conformation, but also can be promoted by disruption of chromatin loops. Further investigations are still required to elucidate the molecular mechanisms underlying the global weakening of higher-order chromatin structure during totipotency acquisition, as well as to fully understand how 2C-specific genes are activated upon disruption of chromatin loops.

## DATA AVAILABILITY

The RNA-seq, ATAC-seq, ChIP-seq, CUT&Tag, and Hi-C sequencing data reported in this paper have been deposited with the NCBI GEO under accession number GSE159623.

Scripts for data analysis are available on GitHub (<https://github.com/shenlab423/2CLC-Project-Code>).

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Dr Samuel Pfaff for the MERVL::tdTomato reporter plasmid, Dr Benoit Bruneau for the CTCF-AID cell line and Dr Fangwei Wang for the anti-H3S10ph antibody. We also appreciate Dr Yangming Wang and all members of the Shen laboratory for providing advice and comments.

## FUNDING

National Natural Science Foundation of China [31871478, 32022023]; National Key Research and Development Programs of China [2017YFC1001500]; Zhejiang Provincial Natural Science Foundation of China [LR18C060001]. Funding for open access charge: National Natural Science Foundation of China [31871478].

*Conflict of interest statement.* None declared.

## REFERENCES

- Xu, Q. and Xie, W. (2018) Epigenome in early mammalian development: inheritance, reprogramming and establishment. *Trends Cell Biol.*, **28**, 237–253.
- Xu, R., Li, C., Liu, X. and Gao, S. (2021) Insights into epigenetic patterns in mammalian early embryos. *Protein Cell*, **12**, 7–28.
- Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D. and Knowles, B.B. (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell*, **7**, 597–606.
- Falco, G., Lee, S.L., Stanghellini, I., Bassey, U.C., Hamatani, T. and Ko, M.S. (2007) Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev. Biol.*, **307**, 539–550.

5. Hamatani, T., Carter, M.G., Sharov, A.A. and Ko, M.S. (2004) Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell*, **6**, 117–131.
6. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D. and Pfaff, S.L. (2012) Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, **487**, 57–63.
7. 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.
8. 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.
9. Hug, C.B., Grimaldi, A.G., Kruse, K. and Vaquerizas, J.M. (2017) Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell*, **169**, 216–228.
10. 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.
11. Kaaij, L.J.T., van der Weide, R.H., Ketting, R.F. and de Wit, E. (2018) Systemic loss and gain of chromatin architecture throughout zebrafish development. *Cell Rep.*, **24**, 1–10.
12. Zhang, K., Wu, D.Y., Zheng, H., Wang, Y., Sun, Q.R., Liu, X., Wang, L.Y., Xiong, W.J., Wang, Q., Rhodes, J.D.P. *et al.* (2020) Analysis of genome architecture during SCNT reveals a role of cohesin in impeding minor ZGA. *Mol. Cell*, **79**, 234–250.
13. Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Boskovic, A., Ziegler-Birling, C., Rodriguez-Terrones, D., Wakayama, T., Vaquerizas, J.M. and Torres-Padilla, M.E. (2015) Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.*, **22**, 662–671.
14. Eckersley-Maslin, M.A., Svensson, V., Krueger, C., Stubbs, T.M., Giehr, P., Krueger, F., Miragaia, R.J., Kyriakopoulos, C., Berrens, R.V., Milagre, I. *et al.* (2016) MERVL/Zscan4 network activation results in transient genome-wide DNA demethylation of mESCs. *Cell Rep.*, **17**, 179–192.
15. Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A. and Bruneau, B.G. (2017) Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, **169**, 930–944.
16. Chen, C.C.L., Goyal, P., Karimi, M.M., Abildgaard, M.H., Kimura, H. and Lorincz, M.C. (2018) H3S10ph broadly marks early-replicating domains in interphase ESCs and shows reciprocal antagonism with H3K9me2. *Genome Res.*, **28**, 37–51.
17. Brind'Amour, J., Liu, S., Hudson, M., Chen, C., Karimi, M.M. and Lorincz, M.C. (2015) An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat. Commun.*, **6**, 6033.
18. Zhu, Y., Yu, J., Rong, Y., Wu, Y.-W., Li, Y., Zhang, L., Pan, Y., Fan, H.-Y. and Shen, L. (2021) Genomewide decoupling of H2AK119ub1 and H3K27me3 in early mouse development. *Sci. Bull.*, **66**, 2489–2497.
19. Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K. and Henikoff, S. (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.*, **10**, 1930.
20. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. and Greenleaf, W.J. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods*, **10**, 1213–1218.
21. Picelli, S., Bjorklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G. and Sandberg, R. (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods*, **10**, 1096–1098.
22. Flyamer, I.M., Gassler, J., Imakaev, M., Brandao, H.B., Ulianov, S.V., Abdennur, N., Razin, S.V., Mirny, L.A. and Tachibana-Konwalski, K. (2017) Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*, **544**, 110–114.
23. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
24. Olivares-Chauvet, P., Mukamel, Z., Lifshitz, A., Schwartzman, O., Elkayam, N.O., Lubling, Y., Deikus, G., Sebra, R.P. and Tanay, A. (2016) Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature*, **540**, 296–300.
25. Fu, X., Wu, X., Djekidel, M.N. and Zhang, Y. (2019) Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells. *Nat. Cell Biol.*, **21**, 835–844.
26. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
27. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A. and Young, R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934–947.
28. Crane, E., Bian, Q., McCord, R.P., Lajoie, B.R., Wheeler, B.S., Ralston, E.J., Uzawa, S., Dekker, J. and Meyer, B.J. (2015) Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*, **523**, 240–244.
29. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.
30. Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A. *et al.* (2017) Multiscale 3D genome rewiring during mouse neural development. *Cell*, **171**, 557–572.
31. You, Q., Cheng, A.Y., Gu, X., Harada, B.T., Yu, M., Wu, T., Ren, B., Ouyang, Z. and He, C. (2021) Direct DNA crosslinking with CAP-C uncovers transcription-dependent chromatin organization at high resolution. *Nat. Biotechnol.*, **39**, 225–235.
32. Kieffer-Kwon, K.R., Tang, Z., Mathe, E., Qian, J., Sung, M.H., Li, G., Resch, W., Baek, S., Pruett, N., Grontved, L. *et al.* (2013) Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell*, **155**, 1507–1520.
33. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I. and Young, R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
34. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B. *et al.* (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75–82.
35. De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J. and Trono, D. (2017) DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.*, **49**, 941–945.
36. Hendrickson, P.G., Dorais, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L. *et al.* (2017) Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.*, **49**, 925–934.
37. Whiddon, J.L., Langford, A.T., Wong, C.J., Zhong, J.W. and Tapscott, S.J. (2017) Conservation and innovation in the DUX4-family gene network. *Nat. Genet.*, **49**, 935–940.
38. Hu, Z., Tan, D.E.K., Chia, G., Tan, H., Leong, H.F., Chen, B.J., Lau, M.S., Tan, K.Y.S., Bi, X., Yang, D. *et al.* (2020) Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells. *Nat. Cell Biol.*, **22**, 175–186.
39. Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L. *et al.* (2017) YY1 is a structural regulator of enhancer-promoter loops. *Cell*, **171**, 1573–1588.
40. Pekowska, A., Klaus, B., Xiang, W., Severino, J., Daigle, N., Klein, F.A., Oles, M., Casellas, R., Ellenberg, J., Steinmetz, L.M. *et al.* (2018) Gain of CTCF-anchored chromatin loops marks the exit from naive pluripotency. *Cell Syst.*, **7**, 482–495.