rRNA biogenesis regulates mouse 2C-like state by 3D structure reorganization of peri-nucleolar heterochromatin

Hua Yu1,2,3,4,10, Zhen Sun1,2,3,4,10, Tianyu Tan1,2,3,4,10, Hongru Pan1,2,3,4,10, Jing Zhao1,2,3,4, Ling Zhang1,2,3,4, Jiayu Chen5, Anhua Lei1,2,3,4, Yuqing Zhu1,2,3,4, Lang Chen6, Yuyan Xu1,2,3,4, Yaxin Liu1, Ming Chen6, Jinghao Sheng7, Zhengping Xu7, Pengxu Qian1,2,3, Cheng Li8, Shaorong Gao5, George Q. Daley9 & Jin Zhang1,2,3,4✉

The nucleolus is the organelle for ribosome biogenesis and sensing various types of stress. However, its role in regulating stem cell fate remains unclear. Here, we present evidence that nucleolar stress induced by interfering rRNA biogenesis can drive the 2-cell stage embryo-like (2C-like) program and induce an expanded 2C-like cell population in mouse embryonic stem (mES) cells. Mechanistically, nucleolar integrity maintains normal liquid-liquid phase separation (LLPS) of the nucleolus and the formation of peri-nucleolar heterochromatin (PNH). Upon defects in rRNA biogenesis, the natural state of nucleolus LLPS is disrupted, causing dissociation of the NCL/TRIM28 complex from PNH and changes in epigenetic state and reorganization of the 3D structure of PNH, which leads to release of Dux, a 2C program transcription factor, from PNH to activate a 2C-like program. Correspondingly, embryos with rRNA biogenesis defect are unable to develop from 2-cell (2C) to 4-cell embryos, with delayed repression of 2C/ERV genes and a transcriptome skewed toward earlier cleavage embryo signatures. Our results highlight that rRNA-mediated nucleolar integrity and 3D structure reshaping of the PNH compartment regulates the fate transition of mES cells to 2C-like cells, and that rRNA biogenesis is a critical regulator during the 2-cell to 4-cell transition of murine pre-implantation embryo development.

1 Center of Stem Cell and Regenerative Medicine, Department of Basic Medical Sciences, and Bone Marrow Transplantation Center of the First Affiliated Hospital, Zhejiang University School of Medicine, 310003 Hangzhou, China. 2 Zhejiang Laboratory for Systems & Precision Medicine, Zhejiang University Medical Center, 1369 West Wenyi Road, 311121 Hangzhou, China. 3 Institute of Hematology, Zhejiang University, 310058 Hangzhou, China. 4 Center of Gene/Cell Engineering and Genome Medicine, 310058 Hangzhou, Zhejiang, China. 5 College of Life Sciences, Zhejiang University, 310058 Hangzhou, China. 6 College of Life Sciences, Zhejiang University, 310058 Hangzhou, China. 7 Institute of Environmental Medicine, and Cancer Center of the First Affiliated Hospital, Zhejiang University School of Medicine, 310058 Hangzhou, China. 8 Center for Bioinformatics, School of Life Sciences, Center for Statistical Science, Peking University, 100871 Beijing, China. 9 Stem Cell Transplantation Program, Division of Pediatric Hematology Oncology, Boston Children’s Hospital, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA. 10 These authors contributed equally: Hua Yu, Zhen Sun, Tianyu Tan, Hongru Pan. ✉email: zhgene@zju.edu.cn
Two-cell (2C) stage embryonic cells are totipotent cells in an earlier stage of embryo development and can generate all cell types of embryonic and extraembryonic tissues. In the culture of mouse embryonic stem (mES) cells, a rare population of mES cells sporadically transit into a 2C stage embryo-like (2C-like) cells with similar molecular features of totipotent 2C-stage embryos. Recent works have demonstrated that the conversion of mES cells to 2C-like cells is regulated by a variety of factors related to epigenetic modification, including histone methylation and acetylation and DNA methylation. In addition, it was also found that RNA hydroxymethylation and protein sumoylation affect the epigenetic state of chromatin to regulate the activation of ERV genes. The structure of chromatin is an important epigenetic factor and is closely related to the regulation of gene expression and cell fate transition. Interestingly, chromatin structure appears to emerge as an important factor in 2C gene regulation and transition of mES cells to 2C-like cells. For instance, 2C/ERV gene activation is regulated by a pioneer transcription factor DUX which increases chromatin accessibility and two recent studies reported that the pluripotency factors DAPP2 and DAPP4 and the maternal factor NELFA can bind to the promoter region of Dux and directly trans-activate its expression. Moreover, 2C-like cells can be induced by downregulation of chromatin remodeling factor CAF. However, the molecular players of chromatin structure in mES cell to 2C-like cell transition have yet to be fully understood.

With the emergence and development of high-throughput chromatin conformation capture technology (Hi-C), the dynamic changes of higher-order chromatin structures during early embryonic development and stem cell differentiation have been elucidated. Two-cell embryo or 2C-like cells show contrasting differences from inner cell mass or ES cells, suggesting that the 3D chromatin structure is a key factor mediating the transition of mES cells to 2C-like cells. Importantly, one of the mechanisms of Dux expression is dependent on a complex of nucleolin NCL and heterochromatin factor TRIM28 in the Peri-Nucleolar Heterochromatin (PNH) region. However, it is not completely known how nucleolus integrity influences higher-order chromatin structure and how the chromatin structure determines Dux expression.

Here, we find that inhibition of nucleolar rRNA biogenesis triggered nucleolar stress, which activated 2C-like transcriptional program and induced an expanded 2C-like cell population in mES cells with a mechanism involving 3D structure reorganization of the PNH and the Dux expression. Consistently, the 3D structure of the PNH reorganizes after early 2-cell during murine early embryo development, which coincides with rRNA biogenesis and Dux repression. Moreover, in mouse early embryos, rRNA biogenesis and matured nucleolus are indispensable for the 2-cell to 4-cell transition. Taken together, our findings provide a mechanistic perspective of rRNA biogenesis in regulating the homeostasis between 2C-like and mES cells, and highlight that rRNA biogenesis in the nucleolus is a critical molecular switch from ZGA gene expressing 2-cell stage to nucleolus-matured blastocyst stage embryos.

Results

Inhibition of rRNA biogenesis activated the 2C-like transcriptional program and induced an expanded 2C-like cell population in mES cells. We first explored whether nucleolar stress produced by inhibiting rRNA biogenesis could induce cell fate reprogramming to 2C-like cells (2CLCs) by performing RNA-seq analysis of mES cells treated by three inducers of cellular stress, including CX-5461, an RNA polymerase I (Pol I) inhibitor; rotenone, an electron transport chain complex 1 inhibitor, and rapamycin, a mTOR pathway inhibitor (CX-5461 treatment dosage: 2 μM, CX-5461 treatment time: 12 h; rotenone treatment dosage: 1 μM, rotenone treatment time: 12 h; rapamycin treatment dosage: 2 μM, rapamycin treatment time: 12 h). We found that nucleolar stress induced by CX-5461 and rotenone, activated the 2-cell marker genes Zscan4d, Dux, and Gmi12794 and repressed pluripotency marker gene Pou5f1. However, the other two cellular stresses upon rotenone or rapamycin treatment did not influence the expression of these genes (Fig. 1a, b and Supplementary Fig. 1a, b). The 2C-like transcriptional program was characterized by activation of transposable elements (TEs), particularly major satellite repeats (GSAT_MM) and ERVL subclasses MERVL-int and MT2_Mm. We systematically examined ERV genes and found global up-regulation of GSAT-MM and each LTR class (Fig. 1c and Supplementary Fig. 1e), particularly GSAT_MM, MERVL-int and MT2_Mm sub-clases, in CX-5461-treated mES cells (Fig. 1c, and Supplementary Fig. 1c, d). However, the other two types of stress, did not activate these repeat elements (Fig. 1c, and Supplementary Fig. 1c, d). Using unsupervised K-means clustering analysis, we identified four major gene clusters specifically expressed in different stages during mouse pre-implantation embryo development (Supplementary Fig. S1d). We found that CX-5461 treatment upregulated 2-cell expressing cluster 1 (C1) and 2-cell/4-cell expressing cluster 2 (C2) genes, and decreased genes expressed in other two stages (C3 and C4) (Fig. 1d and Supplementary Fig. S1f). Yet, other two cellular stresses did not induce this expression pattern (Fig. 1d and Supplementary Fig. S1f).

Moreover, unsupervised hierarchical clustering of transcriptomes of pre-implantation embryos and mES cells from published studies of 2C-like cells confirmed that mES cells treated with CX-5461 were most like the sorted 2CLCs from mES cells, or genetically modified mES cells with 2LC signatures from other studies as well as the 2C embryos. (Fig. 1e). As expected, we observed the abundance of rRNA is significantly reduced under CX-5461 treatment (Supplementary Fig. S1g). Together, these results demonstrate that nucleolar stress induced by rRNA biogenesis defect activated 2C-like transcriptional program in mES cells.

We next asked how the population homeostasis of 2CLCs and ES cells altered in response to rRNA biogenesis defect at the single cell level. In line with bulk RNA-seq data, we observed that the 2C marker genes are up-regulated and pluripotent genes were downregulated in CX-5461-treated mES cells (Supplementary Fig. S2a, b). As expected, we found a marked expansion of the population with MERVL expression in mES cells treated with CX-5461 (Fig. 2a). Strikingly, we observed that the expression level of MERVL genes showed significant negative correlation with ribosomal protein genes (Fig. 2b–d and Supplementary Fig. S2c). Using a 2C::tdTomato reporter in which a tdTomato gene is under control of a MERVL promoter, we examined 2C status of individual cells by fluorescence activated cell sorting (FACS) analysis. Consistent with single-cell RNA-seq data, we observed that CX-5461-treatment induced a significant increase in the number of tdTomato positive (2C::tdTomato+) mES cells in a dose-dependent manner (Fig. 2e and Supplementary Fig. S2d). Moreover, the percentage of tdTomato positive cells was largely maintained even at 24 h after CX-5461 withdrawal (Fig. 2f). Importantly, although CX-5461, mostly at a high concentration, induced mild cell apoptosis, the majority of tdTomato positive cells were negative for the apoptosis markers Annexin-V and DAPI (Supplementary Fig. S2e–g), suggesting that the emergence of 2C::tdTomato+ cells under CX-5461 treatment was not due to the activation of apoptosis pathway. Collectively, these results demonstrate that inhibiting rRNA biogenesis induced a shift of the ES cell homeostasis toward the MERVL-expressing and ribosomal gene repressed 2CLCs.
Deficiency of rRNA biogenesis disrupted the normal state of nucleolar liquid–liquid phase separation (LLPS) and the epigenetic state of the PNH. As it has been reported that ribosomal RNA plays a critical role in maintaining phase separation of nucleolus and phase transition is involved in nucleolar stress, we examined whether nucleolar stress induced by rRNA biogenesis defect leads to changes of nucleolar phase separation. Using electron microscopy, we observed that CX-5461-treated mES cells displayed abnormal nucleolar structure, missing the outer layer usually associated with dense electron intensity (Fig. 3a). Immunofluorescence of key granular compartment (GC) and nucleolar dense fibrillar component (DFC) marker proteins NPM1 and NCL revealed that CX-5461 treatment led to disappearance of the NPM1-marked and
Hierarchical clustering of transcriptomes from our study, published 2C-like cell models and pre-implantation mouse embryos; Control, CX-5461, Roterone and Rapamycin: GEO accession code GSE66640; 2C::tgTomato and 2C::ctgTomato: GEO accession code GSE33923; Zscan4_Em: GEO accession code GSE66640; LoxP::tTA: GEO accession code GSE85632; Dux_GFPpos and Dux_GFPneg: GEO accession code GSE120953 and Dppa4::GFP: GEO accession code GSE100939; Dppa4::GFPpos and Dppa4::GFPneg: GEO accession code GSE113671; Lin28a_KO and Lin28a_WT: GEO accession code GSE66639; Gene expression levels of two biological replicates in same developmental stage were averaged for clustering. The center line is the median, the bottom of the box is the 25th percentile boundary, the top of the box is the 75th, and the top and bottom of vertical line define the bounds of the data that are not considered outliers, with outliers defined as greater/less than ±1.5× IQR, where IQR inter-quartile range.

As it has been observed that phase separation can regulate the epigenetic state of chromatin39–41, we examined the epigenetic changes at the loci of the peri-nucleolar heterochromatin (PNH) region in CX-5461-treated mES cells. It has been reported that transcriptionally inactive genomic regions organize into Inactive Hubs around the nucleolus42. In addition, the nucleolar associated domains (NAD) or LINE1/L1 repeat sequence regions are also defined as repressive chromosomal segments enriched with peri-nucleolar heterochromatin43–49. We thus examined the epigenetic changes on these regions and found decreased H3K9me3 and H3K27me3 levels in CX-5461 treated cells (Fig. 4a, b and Supplementary Fig. S3a–c). Moreover, we observed increased H3K4me3 and H3K27ac levels and improved chromatin accessibility at Inactive Hub42, NAD43,45, and L1 regions (downloaded from UCSC Table Browser) in CX-5461 treated cells (Fig. 4c–g and Supplementary Fig. S3d–f). Previous work has reported that nucleolar protein nucleolin NCL and its interacting partner the heterochromatin protein TRIM28 repress 2C-like program by maintaining the PNH region in mES cells44. The disappeared NCL-marked “ring” structure suggested that RNA biogenesis defect promoted the dissociation of NCL/TRIM28 complex from PNH region. To validate this, we conducted ChiP-seq experiment to investigate the binding changes of NCL/TRIM28 complex on the loci of the PNH region in CX-5461-treated mES cells and found decreased binding of NCL and TRIM28 proteins on Inactive Hub, NAD, and L1 regions (Fig. 4h, i and Supplementary Fig. S3g–i). Altogether, these results demonstrated the RNA biogenesis defect affected the normal state of nucleolar phase separation and changed the epigenetic state of the heterochromatic regions at the periphery of nucleoli by breaking up the binding of NCL/TRIM28 complex on the PNH region.

2C/ERV genes were activated through Dux. Recent studies have reported that a pioneer transcription factor, the DUX protein, directly binds to promoters and LTR elements on 2C genes and repetitive elements and activates their transcription10–12. As Dux expression has been reported to be influenced by nucleolar protein NCL14, we speculated that Dux is the key molecular regulator for nucleolar stress-mediated activation of 2C-like transcriptional program. To this end, we performed the binding motif sequence enrichment analysis of transcription factors on 5258 genes induced by CX-5461. We found that the significantly enriched motifs include both P53 binding sites and DUX binding sites (Fig. 5a). In line with this, we found that 1229 CX-5461-induced genes are P53 direct target genes (Fig. 5b, hypergeometric test, p-value = 0.36), consistent with the fact that p53 signaling is usually activated under nucleolar stress51–57. We further analyzed the overlap between CX-5461 treatment-induced genes and DUX over-expression-induced genes using published RNA-seq data11, and found 621 genes are overlapped (Fig. 5b, hypergeometric test, p-value = 2.96e–195). Using P53 and DUX ChiP-seq data in mES cells11,50, we further observed that both P53 and DUX showed a pattern of binding to the transcriptional start site (TSS) of CX-5461-induced genes (Fig. 5c). The binding pattern of DUX on these genes is weaker when comparing with the strong binding of P53 on CX-5461-induced genes. This was expected as CX-5461 induces many genes that are P53 target, but not DUX target. Interestingly, we found that P53 favors to bind specifically to CX-5461-induced genes, while DUX exclusively binds to the commonly induced genes between CX-5461-treated and Dux-overexpressed cells (Fig. 5d), and to ERV genes induced by CX-5461 (Fig. 5e, f). Consistently, a significant increase of chromatin accessibility in the promoter region is not observed for 4637 CX-5461-induced genes but is observed for 621 commonly induced genes or 10,173 CX-5461-induced ERV genes in Dux-overexpressed mES cells11 (Supplementary Fig. S4a). In addition, we observed the decreased H3K9me3 and H3K27me3 levels and increased H3K4me3 and H3K27ac levels around Dux locus (Supplementary Fig. S4b). These analyses suggested that nucleolar stress-induced 2C activation is through Dux. To validate this hypothesis, we silenced Dux expression in CX-5461-treated mES cells and found that it reversed the 2C/ERV gene induction (Fig. 5g). We further performed ChiP-qPCR experiments to assess the changes of H3K9me3 & H3K27me3 levels and Dux binding of CX-5461-induced 2C marker genes after Dux silencing. We observed that, after Dux silencing, both H3K9me3 and H3K27me3 were increased. In contrast, Dux binding is decreased for CX-5461-induced 2C genes (Fig. 5h–j). When the Dux was silenced in CX-5461-treated mES cells, we observed that H3K9me3 and H3K27me3 levels, and DUX binding on CX-5461-induced 2C genes were reversed (Fig. 5h–j). We also performed
ATAC-seq experiment to assess the changes of chromatin accessibility after Dux silencing. In well support with these results, we observed that chromatin accessibility of CX-5461-induced 2C genes was reversed by Dux silencing (Fig. 5k). These results demonstrated that nucleolar stress induced 2C/ERV gene activation through Dux. In line with DUX’s role in chromatin accessibility, we observed that the 621 commonly induced genes and 10173 CX-5461 induced ERV genes have increased chromatin accessibility, and decreased H3K9me3 and H3K27me3 levels when comparing CX-5461 treated mES cells with control mES cells or comparing 2-cell embryos with ICM stage embryos28,29,58 (Supplementary Fig. S4c–h). In addition, we also found increased H3K4me3 and H3K27ac levels of 621 commonly induced genes and 10,173 CX-5461 induced ERV genes in
CX-5461 treated mES cells (Supplementary Fig. S4i, j). Collectively, these results demonstrated that nucleolar stress induced a 2C-like transcriptional and epigenetic program in mES cells through Dux.

rRNA biogenesis defect drove 3D chromatin structure reorganization of PNH and MERVL regions towards the 2C-like state. The disassembly of PNH region suggested that the 3D chromatin structure within PNH region might have reshaped under CX-5461 treatment. To explore the reorganization of 3D chromatin conformation landscape of CX-5461-treated mES cells relative to control mES cells, we performed in situ Hi-C with randomly selected genomic regions, Mann–Whitney U-test, the replicates of experiment N = 10, averaged p = 0.0, 3.42E−06 for Inactive Hub, NAD, and L1, respectively. Moreover, the Dux locus is significantly further away from the PNH region as characterized by the largely decreased Hi-C contacts (Fig. 6a–c, g–i) compared with randomly selected genomic regions, Mann–Whitney U-test, the replicates of experiment N = 10, averaged p = 1.65E−05, 7.82E−05, 2.68E−03 for Inactive Hub, NAD, and L1, respectively). We further analyzed the 3D chromatin structural correlation within PNH region, and between the Dux locus and PNH region by comparing Hi-C pearson correlation coefficient (PCC) matrix of control and CX-5461-treated mES cells. We observed the obviously decreased 3D chromatin structural correlation within PNH region (compared with randomly selected genomic regions, Mann–Whitney U-test, the replicates of experiment N = 10, averaged p = 0.0 for all Inactive Hub, NAD, and L1) and between the Dux locus and PNH region (compared with randomly selected genomic regions, Mann–Whitney U-test, the replicates of experiment n = 10, averaged p = 1.86E−6, 5.03E−6, 4.02E−5 for Inactive Hub, NAD, and L1, respectively) (Fig. 6d–f, j–l). We further validated the above findings using DNA fluorescence in situ hybridization (FISH) and found that the Dux locus and a locus within PNH region located further from the peri-nucleolar region indicated by NCL staining in the CX-5461-treated cells (Fig. 6m). Interestingly, by performing the same analysis as above on public Hi-C data of mouse pre-implantation embryo development21, we observed that a similar trend of 3D chromatin structure reorganization of PNH and Dux regions when comparing early two-cell embryos with ICM stage embryos, namely, the two-cell embryos showed less organized PNH and less contacts between the Dux Locus and the PNH (Supplementary Fig. S5a–d), suggesting a process of maturation of PNH 3D structure organization and of the Dux release from the PNH during embryo development. Interestingly, we observed that 3D structure reorganization of PNH was initiated during early two-cell to late two-cell transition, which coincides with the beginning of shutting down Dux gene expression during murine early embryo development (Supplementary Figs. S5a–d and S9d).

As it has been reported that 2CLCs display increased three-dimensional structural plasticity relative to mES cells23, we next asked whether the global 3D chromatin architecture is changed in CX-5461-treated mES cells. We compared control and CX-5461-treated mES cells Hi-C maps, with lymphoblastoid cells as a reference for fully differentiated cells39. A global analysis of A(active)/B(inactive) compartment strength showed a slight decrease of contacts within the B compartments in CX-5461-treated mES cells compared with control mES cells (Supplementary Fig. S6a–d). However, at topologically associating domain (TAD) or chromatin loop level, we found a mild increase in their strength in CX-5461-treated mES cells (Supplementary Fig. S6e–i). To specifically investigate 2C-related genes, we further performed an analysis of local architectural differences around MERVL loci. We found that the insulation score60 of chromatin around MERVL genes activated by CX-5461 treatment is markedly increased both globally (Supplementary Fig. S7a) and at local MERVL sites (Fig. 7a and Supplementary Fig. S7b), similar as observed in two-cell embryos compared with ICM23 (Fig. 7a, b and Supplementary Fig. S7b), and the topological associated domain (TAD) structure around MERVL gene loci is more obvious (Fig. 7a, b and Supplementary Fig. S7b), showing a more similar pattern as that in two-cell embryos21 (Fig. 7a, b and Supplementary Fig. S7b). The chromatin structure reorganization around MERVLs is accompanied with more open chromatin states around MERVLs both globally (Supplementary Fig. S7a) and locally (Fig. 7a and Supplementary Fig. S7b) and with their increased expression (Fig. 7a and Supplementary Fig. S7b). These results together demonstrate that nucleolar stress promoted the transformation of mES cells to 2C-like cells with reshaped 3D chromatin structure and its associated epigenetic status to facilitate gene expression, particularly at the PNH and MERVL regions.

Genetic perturbation of rRNA biogenesis recapitulated CX-5461-induced 2C-like molecular phenotypes. To further investigate the critical role of rRNA biogenesis in regulating the 2C program and the homeostasis between mES cells and 2C-like cells, we generated two rRNA biogenesis-inhibited mES cells lines: 1) a line with degraded Pol I protein (PRA1) by an auxin-inducible degron system61 (Supplementary Fig. S8a), and 2) a snoRNA knockout line (SNORD113–114 gene cluster, a gift from Pengxu Qian) (Supplementary Fig. S8b), as snoRNAs are required for rRNA modification and biogenesis. Using these two cell lines, we performed DNA FISH experiments and found the similar molecular phenotypes of CX-5461-treated mES cells, i.e., the Dux locus and a representative locus within PNH region located further from the peri-nucleolar region (Fig. 8a–d). We then carried out FRAP experiments and consistently observed significantly increased mobility of NCL and NPM1 proteins in these two cell lines compared with their wild-type controls (Fig. 8e–h and Supplementary Fig. S8c–f). In line with these, we found that the 2C marker genes,
including MERVL, Dux, Zscan4d, Gm12794, and Gm4340, were significantly activated in these two cell lines (Fig. 8i, j). Moreover, using FACS analysis, we observed a significant increase of the percentage of tdTomato positive cells in these two cell lines compared with control cells (Fig. 8k, l and Supplementary Fig. S8g, h). Collectively, these results further confirmed that repressing rRNA biogenesis can activate 2C-like program and induce the transition of mES cells to 2C-like cells.
Fig. 3 Deficiency of rRNA biogenesis disrupted the normal state of nucleolar LLPS. a CX-5461 treatment causes abnormal nucleolar structure with electron microscopy; Experiment was repeated independently three times with similar results. b Immunofluorescence staining of NCL in control mES cells and CX-5461-treated mES cells; Experiment was repeated independently three times with similar results. c Immunofluorescence staining of NPM1 in control mES cells and CX-5461-treated mES cells; Experiment was repeated independently three times with similar results. d Immunofluorescence staining of FBL in control mES cells and CX-5461 treated mES cells; Experiment was repeated independently three times with similar results. e Immunofluorescence staining of RPA194 in control and CX-5461 treated mES cells; Experiment was repeated independently three times with similar results. f FRAP analysis showing CX-5461 treatment causes accelerated recovery after photobleaching of NCL; N = 3 biologically independent experiments; Data are presented as mean values +/- SEM. SEM standard error of mean. Shown images are representative of three times of biologically independent experiments. g FRAP analysis showing CX-5461 treatment causes accelerated recovery after photobleaching of NPM1; N = 3 biologically independent ChIP-seq experiments; Data are presented as mean values +/- SEM. SEM standard error of mean. Shown images are representative of three times of biologically independent experiments. h FRAP analysis showing CX-5461 treatment causes accelerated recovery after photobleaching of FBL; N = 3 biologically independent ChIP-seq experiments; Data are presented as mean values +/- SEM. SEM standard error of mean. Shown images are representative of three times of biologically independent experiments. Source data are provided as a Source Data file.
rRNA biogenesis is critically required at the two-cell-to-four-cell stage transition during pre-implantation embryo development. To better understand whether the physiological function of rRNA biogenesis is to facilitate embryo development during and after the two-cell exit, we first inspected rRNA expression levels in all stages of pre-implantation mouse embryos. The precursor and matured rRNA levels were low in MII-oocyte, pronuclear zygote, early two-cell and middle two-cell stages but increased sharply from the late two-cell stage to the blastocyst stage (Fig. 9a). We then analyzed the expression levels of different subunits of gene polymerase I (Pol I) in pre-implantation embryos. Different from rRNA expression, Pol I gene mRNA levels increased significantly from the late two-cell stage to four-cell stage but decreased markedly as embryos progressed through eight-cell, morula stages, and reached to blastocysts which still had a higher level than those stages before the late two-cell (Fig. 9b). Consistent with Pol I genes, we observed the similar pattern of increased expression of ribosome biogenesis genes (Supplementary Fig. S9a). This indicated that while the levels of rRNA were gradually accumulated during pre-implantation embryo development, the rRNA biogenesis rate reached to a peak during the late two-cell to the four-cell stage. In contrast, the ERV and 2C marker genes, such as Dux, Zscan4d, and Gmb12794 were significantly decreased during the late two-cell-to-four-cell stage (Supplementary Fig. S9b–d). This reciprocal expression pattern between 2C marker genes and rRNA biogenesis genes suggested that rRNA biogenesis may play a key role in shutting down the 2C program, as revealed by our 2CLC emergence analysis in cultured mES cells, and in promoting the transition from the two-cell to the four-cell stage. We next applied CX-5461 (an embryo tolerable concentration) to mouse early embryos as they progress through pronuclear zygotes to blastocysts. When compared with the control embryos, we found that CX-5461-treated embryos were indeed blocked before the four-cell stage (Supplementary Fig. S9e). We further divided mouse embryos into four groups according to the different stages of CX-5461 treatment, including transitions of zygote-to-two-cell, two-cell-to-four-cell, and morula-to-blastocyst, respectively (Fig. 9c).

Compared with the untreated control, we found that blastocyst formation rates of all three CX-5461-treated groups were decreased, and the two-cell-to-four-cell-treated group showed the strongest decrease of the blastocyst formation rate at both early and late blastocyst stages (Fig. 9d). This is consistent with the pattern of Pol I gene RNA expression and the pattern of PNH reshaping after early two-cell stage during pre-implantation embryo development (Supplementary Fig. S5a–d), indicating that rRNA biogenesis is most critically required during the two-cell-to-four-cell transition compared with other stages. In well support with this, the knockout of Pol1a gene led to mouse embryos arrested at two-cell [82]. Moreover, upon successful inhibition of rRNA biogenesis at the morula/blasto phase (Fig. 9e), we found the disappearance of the NPM1-marked and NCL-marked “ring” structure and abnormal localization and reduced signal density of FBL in the nucleolus (Supplementary Fig. S9f). Importantly, we observed increased expression of 2C genes such as Zscan4d, Gmb4340, Dux, and Mervl-pol (Fig. 9f). RNA-seq also demonstrated upregulated C1 and C2 clusters of 2C genes (Fig. 9g) defined before (Supplementary Fig. S1f) as well as ERV genes (Fig. 6h–j) in CX-5461-treated embryos compared with controls, consistent with the results from mES cells described above (Fig. 1a–d and Supplementary Fig. S1a–e). Altogether, these data demonstrated rRNA biogenesis and nucleolar integrity is a molecular switch for the transition from the two-cell to the four-cell embryos.

### Discussion

Starting from zygotic genome activation (ZGA) at the early two-cell stage, an embryo undergoes multiple developmental stages to be ready for itself for implantation. Along with the pre-implantation embryo development, the activity of anabolic metabolism and translation is progressively increased. Nucleoli, a nuclear sub-compartment responsible for the biogenesis of ribosome, originally derived from nucleolar precursor bodies (NPB), become mature structurally and functionally during this process [65–65]. Interestingly, a recent study reported that TRIM28/Nucleolin/LINE1 complex can regulate both Dux repression and
Fig. 5 2C/ERV genes were activated through the Dux. a Enriched binding motifs of 5258 genes induced by CX-5461 treatment. b Venn diagrams showing the overlap among CX-5461 treatment induced genes, p53 activated direct target genes and Dux-overexpression induced genes. c Heatmap plots demonstrate the binding of p53 and Dux proteins on within 5 kb region around transcription start sites of 5258 CX-5461 induced genes using published P53 ChIP-seq data and DUX ChIP-seq data. GEO accession code GSE26360 for P53 and GEO accession code GSE85632 for DUX. d Line plots demonstrate the meta-analysis results of P53 and DUX on within 5 kb region around transcription start sites of 621 commonly induced genes between CX-5461 treatment and Dux overexpression and 4637 specifically induced genes by CX-5461 using published P53 ChIP-seq data and DUX ChIP-seq data. GEO accession code GSE26360 for P53 and GEO accession code GSE85632 for DUX. e Line plots demonstrate the meta-analysis results of P53 and DUX protein on within 5kb region around transcription start and end sites of 10,173 CX-5461 induced ERV genes using published P53 ChIP-seq data and DUX ChIP-seq data. The regions of different lengths of gene body were fitted to 5 kb. GEO accession code GSE26360 for P53 and GEO accession code GSE85632 for DUX. f Heatmap plots demonstrate the binding of P53 and DUX proteins on within 5 kb region around transcription start and end sites of 5258 CX-5461 induced MERVL-int and MT2_Mm genes; The regions of different lengths of gene body were fitted to 5 kb. GEO accession code GSE85632 for DUX. g qRT-PCR showing the expression of Dux or 2C-related genes in Dux silenced mES cells; GMI2794: p = 1.21E−13(***), Zscan4: p = 3.00E−09(***), MERVL-pol: p = 4.55E−03(**), Dux: p = 9.60E−14(***), two-way ANOVA; N = 3 biologically independent ChIP-qPCR experiment; Data are presented as mean values +/- SEM. SEM standard error of mean. h ChIP-PCR showing H3K9me3 levels of Dux or 2C-related genes in Dux silenced mES cells; MERVL-LTR: p = 1.04E−11(***), MERVL-Pol: p = 9.68E−12(***), Zscan4-1: p = 2.51E−10(***), Zscan4-2: p = 1.67E−11(***), GMI2794-1: p = 1.90E−09(***), GM12794-2: p = 2.66E−11(***), Dux-TSS: p = 1.35E−04(*), Dux-coding: p = 4.92E−11(***), two-way ANOVA; N = 3 biologically independent ChIP-qPCR experiment; Data are presented as mean values +/- SEM. SEM standard error of mean. i ChIP-PCR showing H3K27me3 levels of Dux or 2C-related genes in Dux silenced mES cells; MERVL-LTR: p = 9.87E−12(***), MERVL-Pol: p = 2.05E−08(***), Zscan4-1: p = 1.12E−11(***), Zscan4-2: p = 5.81E−10(***), GM12794-1: p = 9.56E−12(***), GM12794-2: p = 9.55E−12(***), Dux-TSS: p = 9.55E−12(***), Dux-coding: p = 1.02E−11(***), two-way ANOVA; N = 3 biologically independent ChIP-qPCR experiment; Data are presented as mean values +/- SEM. SEM standard error of mean. j ChIP-PCR showing DUX protein binding levels on 2C-related genes in Dux silenced mES cells; MERVL-LTR: p < 1.00E−15(***), MERVL-Pol: p < 1.00E−15(***), Zscan4-1: p < 1.00E−15(***), Zscan4-2: p < 1.00E−15(***), GM12794-1: p < 1.00E−15(***), GM12794-2: p < 1.00E−15(***), two-way ANOVA; N = 3 biologically independent ChIP-qPCR experiment; Data are presented as mean values +/- SEM. SEM standard error of mean. k Line plots demonstrate the meta-analysis results of chromatin accessibility in Dux silenced mES cells within 5 kb region around transcription start sites or transcription start and end sites of 621 commonly induced genes between CX-5461 treatment and Dux overexpression and 10,173 CX-5461 induced ERV genes. The regions of different lengths of ERV genes were fitted to 5 kb. GEO accession code GSE166041. p53-ctrl: untreated mES cells, p53-adr: mES cells treated with adriamycin, a DNA damage agent widely used to activate p53, Dux (12 h): mES induced with doxycycline for 12 h, Dux (18 h): mES induced with doxycycline for 18 h. Source data are provided as a Source Data file.
rRNA biogenesis. This finding suggests a tight interconnection between turning off ZGA and initiating nucleolar formation. However, we did not completely know how the switch between ZGA and nucleolar formation occurs in the nucleus.

Here, we reported that nucleolar rRNA biogenesis and higher-order 3D chromatin structure remodeling of PNH might coordinate to develop during the two-cell to later stage transition, and we found that mES cells cultured in vitro can transform into 2C-like cells upon nucleolar stress caused by repressing rRNA biogenesis. We propose a mechanistic model for the role of rRNA biogenesis in regulating mouse 2C-like state (Fig. 10). In the unperturbed mouse embryonic stem cells, nucleolar integrity mediated by rRNA biogenesis maintains the normal liquid–liquid phase separation (LLPS) of nucleolus and the establishment of peri-nucleolar heterochromatin (PNH), and this normal nucleolar LLPS facilitates NCL/TRIM28 complex occupancy on the Dux locus and repression of Dux expression. In contrast, in the rRNA biogenesis-repressed mouse embryonic stem cells, the natural liquid-like phase of nucleolus is disrupted, causing the dissociation of the NCL/TRIM28 complex from PNH and alterations of epigenetic state and 3D structure of PNH, which finally leads to Dux released from PNH region, activation of 2C-like program and transition of mES cells to 2C-like cells. Given the dynamic regulation of nucleolus and rRNA gene

Fig. 6 rRNA biogenesis defect drove 3D chromatin structure reorganization of PNH region towards the 2C-like state. a Hi-C contact maps of Inactive Hub and 1.5 Mb genomic regions around Dux at 150 kb resolution. GEO accession code GSE166041. b Hi-C contact maps of NAD and 1.5 Mb genomic regions around Dux at 150 kb resolution. GEO accession code GSE166041. c Hi-C contact maps of L1 and 1.5 Mb genomic regions around Dux at 150 kb resolution. The zoomed-in regions aim to demonstrate the change of Hi-C contacts between Dux and chromosome 10 in control mES cells and CX-5461 treated mES cells. GEO accession code GSE166041. d Hi-C pearson correlation coefficient (PCC) heat maps of Inactive Hub and 1.5 Mb genomic regions around Dux at 150 kb resolution. GEO accession code GSE166041. e Hi-C pearson correlation heatmap of NAD and 1.5 Mb genomic regions around Dux at 150 kb resolution. GEO accession code GSE166041. f Hi-C pearson correlation heat maps of L1 and 1.5 Mb genomic regions around Dux at 150 kb resolution. The zoomed-in regions aim to demonstrate the change of Hi-C PCC between Dux and chromosome 10 in control mES cells and CX-5461 treated mES cells. GEO accession code GSE166041. g Scatter plot demonstrates the log$_2$(fold change) of Hi-C contacts between Inactive Hub and different types of genes in control and CX-5461 treated mES cells. GEO accession code GSE166041. h Scatter plot demonstrates the log$_2$(fold change) of Hi-C contacts between NAD and different types of genes in control and CX-5461 treated mES cells. GEO accession code GSE166041. i Scatter plot demonstrates the log$_2$(fold change) of Hi-C contacts between L1 and different types of genes in control and CX-5461 treated mES cells. GEO accession code GSE166041. j Scatter plot demonstrates the PCC difference between Inactive Hub and different types of genes in control and CX-5461 treated mES cells. GEO accession code GSE166041. k Scatter plot demonstrates the PCC difference between NAD and different types of genes in control and CX-5461 treated mES cells. GEO accession code GSE166041. l Scatter plot demonstrates the PCC difference between L1 and different types of genes in control and CX-5461 treated mES cells. GEO accession code GSE166041. m DNA FISH analysis with a Dux locus probe and Inactive Hub locus probe, and co-immunostained with NCL protein. The percentage of Nucleolus-localized (overlapped with NCL) and Nucleoplasm-localized (nonoverlapped with NCL) of FISH signals is calculated. Dux: $p = 3.67E^{-05}(***), \text{Inactive Hub: } p = 4.13E^{-05}(***), \text{chi-square test; } N$ denotes the number of observed mES cells; $N = 11$ biologically independent observations; Data are presented as mean values $+\bar{\text{SEM}}$. SEM standard error of mean. GEO accession code GSE166041. Source data are provided as a Source Data file.
Fig. 7 rRNA biogenesis defect drove 3D chromatin structure reorganization of MERVL region towards the 2C-like state. a Aggregate Observed(O)/Expected(E) Hi-C matrices centered on CX-5461 induced MERVL genes in control, CX-5461 treated mES cells and mouse embryos throughout mouse pre-implantation embryonic development. GEO accession codes GSE166041, GSE63525, and GSE82185. b Representative 40 kb Hi-C O/E interaction matrices of a MERVL loci located at TAD boundaries (chr7:97,113,503-97,122,376) are shown as heatmaps, along with the insulation score and genome browser tracks of RNA-Seq, H3K9me3, H3K27me3, H3K4me3, and H3K27ac ChIP-Seq signals of the expanded genomic region containing the TAD boundary (arrows) in control and CX-5461 treated mES cells as well as in mouse early embryos. GEO accession code GSE166041 and GSE82185.
chromatin during early embryo development and the sensitivity of embryos to environmental stress at the early stages, it is conceivable that embryos may use the mechanisms elucidated above to ensure its safe development.

Nucleolus, the largest membrane-less condensate in a cell, is a stress-sensitive organelle and ensure quality control of nuclear proteome under stress. Its association with heterochromatin in its periphery confers genetic regulation of key cell fate decision factors such as Dux in pluripotent stem cells. Previous studies on nucleolus in stem cells mainly focused on the role of rRNA and its associated chromatin in the context of ES cell self-renewal and differentiation or exit of pluripotency. In contrast, our work provided a perspective in reprogramming mES cells back to 2C-like cell, and in nucleolar phase separation and 3D chromatin structure remodeling at the PNH. These findings are in line with the emerging notion that phase-separated condensates regulate transcription, epigenetics, and higher-order chromatin structure, and shed light on a previously neglected area of nucleolus-associated condensates in chromatin control during early development. It is worth-mentioning that we do not intend to overstate that the fate transition of mES to 2C-like cell triggered by rRNA biogenesis defect is only explained by LLPS. What we observed is that the integrity of nucleolus mediated by rRNA biogenesis maintains the normal nucleolar LLPS and 3D structure of the PNH. It is possible that 3D structure reshaping of the PNH mediated by nucleolar LLPS is a common thread of RNA and RNA binding protein-mediated Dux silencing and 2C repression (e.g., through rRNA, snoRNA, LINE1 RNA, NCL, TRIM28, or LIN28). The NCL/TRIM28 complex or other proteins localized at the nucleolus and its peripheral heterochromatin are possibly the key factors coordinately connecting nucleolar LLPS and the establishment and maintenance of the PNH. Our study provides initial evidence that the LLPS model is helpful in explaining how the assembly and function of nucleolus participate in the gene regulation process in the nucleolus, and the quantitative models merit future investigations.

Methods

Cell culture. E14 wild-type mES cells were cultured on 0.1% gelatin-coated plates with mouse embryonic fibroblast feeder cells in LIF/2i medium (1:1 mix of DMEM/F12 (11320-033, Gibco) and Neurobasal medium (21103-049, Gibco) containing...
Knockout causes accelerated recovery after photobleaching of NPM1; degradation causes accelerated recovery after photobleaching of NPM1; values $p =$ mean. Source data are provided as a Source Data file.

probe and Inactive Hub locus probe, and co-immunostained with NCL protein in control and snoRNA knockout mES cell lines; Experiment was repeated independently three times with similar results. c The percentage of Nucleolus-localized (overlapped with NCL) and Nucleolus-labeled (nonoverlapped with NCL) of FISH signals in control and snoRNA knockout mES cell lines; Dux: $p =$ 0.04(*), chi-square test; N denotes the number of observed mES cells; N = 6 biologically independent observations; Data are presented as mean values $+/−$ SEM. SEM standard error of mean. d DNA FISH analysis with a Dux locus probe and Inactive Hub locus probe, and co-immunostained with NCL protein in control and snoRNA knockout mES cell lines; Experiment was repeated independently three times with similar results. e FRAP analysis showing Pol I degradation causes accelerated recovery after photobleaching of NCL; N = 3 biologically independent experiments; Data are presented as mean values $+/−$ SEM. SEM standard error of mean. f FRAP analysis showing Pol I degradation causes accelerated recovery after photobleaching of NPM1; N = 3 biologically independent experiments; Data are presented as mean values $+/−$ SEM. SEM standard error of mean. h FRAP analysis showing snoRNA knockout causes accelerated recovery after photobleaching of NPM1; N = 3 biologically independent experiments; Data are presented as mean values $+/−$ SEM. SEM standard error of mean. i The percentage of 2C:tdTomato positive cells was quantified using FACS analysis in control mES cells and Pol I degraded mES cells; $p =$ 0.10–04 (*), two-way ANOVA; N = 4 biologically independent experiments; Data are presented as mean values $+/−$ SEM. SEM standard error of mean. j The percentage of 2C:tdTomato positive cells was quantified using FACS analysis in control mES cells and snoRNA knockout mES cells; $p =$ 1.64E $−$03(**), two-way ANOVA; N = 4 biologically independent experiments; Data are presented as mean values $+/−$ SEM. SEM standard error of mean.

Source data are provided as a Source Data file.
primary probe and secondary probe were freshly mixed into hybridization buffer (2x SSC, 50% formamide, 20% dextran sulfate) at 6 and 1 µM final concentration and dropped on samples. The samples were heated at 85 °C for 20 min and transferred to a 37 °C incubator before hybridization overnight. For the co-immunostaining, the cell samples above were washed three times in PBS and then incubated with primary antibody (1:200, Nucleolin, CST, cat. no. 145745) for 12 h at 4 °C. After three times of washing with PBS, Donkey anti-Rabbit secondary antibody (1:200, Abcam, cat. no. ab150077) was performed and incubated for 4 h. After washing three times with PBS, DAPI was used for nucleus staining.

Transmission electron microscope (TEM). For transmission electron microscopy, massive mouse E14 ES cells cultured in one 10 cm dish with three days, and then were collected to get rid of feeder cells, and subsequently fixed with 2.5% glutaraldehyde. After that, the cell mass was dispersed into several small pieces and fixed at least 6 h at 4 °C. Next, the cell samples were treated with standard...
procedures and cut into ultra-thin sections meeting the imaging requirements. Finally, the obtained samples were imaged on FEI Spirit 120 kV LaB6 Routine Cryo-EM Capable Electron Microscope.

Embryo collection and culture. The animal facility of Zhejiang University was used to house the C57BL/6J mice. The C57BL/6J mice were cultivated on 12 h light/dark cycle. All embryo experiments were carried out according to the Animal Research Committee guidelines of Zhejiang University. To collect pre-implantation mouse embryos RNA-seq data; Gene expression levels of two biological replicates in same developmental stage were averaged for plotting. GEO accession code GSE166041.

Embryo immuno-fluorescence staining. Embryos were first fixed with 1 and 2% paraformaldehyde (PFA) in 1× PBS for 3 min sequentially, followed by treatment with 4% PFA for 30 min at room temperature (RT). Embryos were washed three times with 1× PBS, permeabilized for 15 min in PBS/0.25% Triton X-100 and blocked in blocking buffer (PBS/0.2% BSA/0.01% Tween-20) for 1 h at RT, followed by incubation overnight with primary antibodies (1:200, Nucleolin, CST, 145745; 1:400, NPM1, Sigma, B0556; 1:200, Fibrillarin, Abcam, ab4566; RPA194, 1:50 (***), two-sided, Wilcox signed rank test, N = 2 biologically independent RNA-seq experiments. GEO accession code GSE166041. j Violin plots show expression levels of major ERV gene classes in control blastocyst embryos and CX-5461 treated blastocyst embryos; seq_n denotes the number of annotated GSAT_MM sequences; MERVL-int: p

ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-26576-2 | www.nature.com/naturecommunications

Fig. 9 rRNA biogenesis is critically required at the two-cell-to-four-cell stage transition during pre-implantation embryo development. a Expression of Pre-rRNA and 28S rRNA across different embryo developmental stages; N = 4 biologically independent experiments; Data are presented as mean values +/- SEM. SEM standard error of mean. b Expression of different subunit genes of RNA polymerase I across different embryo developmental stages using public pre-implantation mouse embryos RNA-seq data; Gene expression levels of two biological replicates in same developmental stage were averaged for plotting. GEO accession code GSE66390. c Different schemes of treatment with CX-5461. The 24 h time window for CX-5461 treatment is highlighted in red; Dpc days post-coitum. d Stacked bar plots showing fraction of embryos at different developmental stages with the different CX-5461 treatment schemes in Fig. 6c. The numbers of embryos of group A to group D were all 15 embryos. e qRT-PCR showing rRNA expression level in blastocysts, after CX-5461 treatment of morula embryos followed by in vitro culture of the treated embryos; 18S rRNA: p = 2.68E–06 (**), 28S rRNA: p = 1.98E–05 (**), Pre-rRNA: p = 2.17E–05 (**), two-way ANOVA; N = 3 biologically independent experiments; Data are presented as mean values +/- SEM. SEM standard error of mean. f qRT-PCR showing 2 CM marker gene expression level in blastocysts, after CX-5461 treatment of morula embryos followed by in vitro culture of the treated embryos. Dux: p = 4.09E–03(*), MERVL-pol: p = 6.82E–03(*), Zscan4: p = 2.56E–02(*), Gml2794: p = 2.82E–01, two-way ANOVA; N = 3 biologically independent experiments; Data are presented as mean values +/- SEM. SEM standard error of mean. g Violin plots demonstrating the expression level changes of stage-specific gene clusters of mouse pre-implantation embryos (as defined in Supplementary Fig. S1) in CX-5461-treated and control blastocyst embryos, C1 vs. All genes: p = 4.55E–09, C2 vs. all genes: p = 2.30E–14, two-sided, Mann-Whitney U-test, N = 2 biologically independent RNA-seq experiments. GEO accession code GSE166041. h Violin plots show expression levels of major ERV gene classes in control blastocyst embryos and CX-5461 treated blastocyst embryos; 18S rRNA: p = 5.11E–07(**), MERVL-MaLR: p = 1.65E–07(**), MERVL: p = 3.65E–07(**), MERVL-M: p = 3.67E–07(**), two-sided, Wilcox signed rank test; N = 2 biologically independent RNA-seq experiments. GEO accession code GSE166041. i Violin plots show expression levels of ERV gene sub-classes of GSAT_MM in control blastocyst embryos and CX-5461 treated blastocyst embryos; seq_n denotes the number of annotated MERVL-int and MT2_Mm sequences; MERVL-int: p = 2.65E–18(**), MT2_Mm: p = 8.51E–05(**), two-sided, Wilcox signed rank test, N = 2 biologically independent RNA-seq experiments. GEO accession code GSE166041. g-j The center line is the median, the bottom of the box is the 25th percentile boundary, the top of the box is the 75th, and the top and bottom of vertical line define the bounds of the data that are not considered outliers, with outliers defined as greater/less than ±1.5× IQR, where IQR inter-quartile range. Source data are provided as a Source Data file.

Embryo RNA-seq library preparation and sequencing. Embryos were collected (five embryos per sample) in 0.2 ml PCR tubes with a micro-capillary pipette and processed into cDNA with Superscript II reverse transcriptase. The cDNA is amplified with KAPA HiFi HotStart using 12 cycles. Sequencing libraries were constructed from 1 ng of pre-amplified cDNA using DNA library preparation kit (TruePrep DNA Library Prep Kit V2 for Illumina, Vaizeyme). Libraries were sequenced on a HiSeq-PE150, with pair-end reads of 150 bp length each.

Bulk RNA-seq library preparation and sequencing. A total amount of 2 µg RNA per sample was used as input materials for the RNA sample preparation. mRNA was purified from total RNA samples by poly T oligoattached magnetic beads. Purified mRNA was fragmented at 94 °C for 15 min by using divalent cations under elevated temperature in NEBNext first strand synthesis reaction buffer (5x). First strand cDNA was synthesized using random primer and ProtoScript II reverse transcriptase in a preheated thermal cycler as follows: 10 min at 25 °C, 15 min at 42 °C, 15 min at 70 °C. Immediately finished, second strand synthesis reaction was performed by using second strand synthesis reaction buffer (10x) and enzyme mix at 16 °C for 1 h. The library fragments were purified with QiaQuick PCR kits and elution with EB buffer, then terminal repair, A-tailing and adapter added were implemented. The products were retrieved, and PCR was performed for library enrichment. The libraries were sequenced on an Illumina platform.

10x single-cell mRNA library preparation and sequencing. Single-cell suspensions of control and CX-5461 treated mES cells were resuspended in DPBS-0.04% BSA at 1 × 10⁶ cells/ml. Then scRNA-seq libraries were generated from the 10x Single Cell 3' Solution Reagents V2 according to the manufacturer’s protocol (10x Genomics). After the GE-MRT incubation, barcoded-cDNA was purified with DynaBeads cleanup mix, followed by 10-cycles of PCR amplification (98 °C for 3 min; [98 °C for 15 s, 67 °C for 20 s, 72 °C for 1 min] × 10; 72 °C for 1 min). The total cDNA of single-cell transcriptomes was fragmented, double-size selected with SPRI beads (Beckman), followed by 12 cycles sample index PCR amplification (98 °C for 45 s; [98 °C for 20 s, 54 s for 30 s, 72 °C for 1 min] × 10; 72 °C for 1 min), then another double-size selection with SPRI beads was performed before sequencing. Libraries were sequenced on the Illumina HiSeqX10 platform according to the manufacturer’s instructions (Illumina). Read 1 and Read 2 (paired end) were 150 bp, and the length of index primer was designed as 8 bp.

ChIP-seq library preparation and sequencing: mES cells were cross-linked in 1% formaldehyde for 10 min at 37 °C, followed by adding glycine to a final concentration of 125 mM and incubated for 5 min at room temperature. Spin the cells for 5 min at 4 °C, 111×g, and wash twice in ice-cold PBS. Cell pellet was resuspended with lysis buffer containing 1× Protease Inhibitor Cocktail and incubated on ice for 10 min, then

Green Master (Vaizeyme) in a final volume of 20 µl per reaction as manufacturer’s instructions.

Embryo collection, cDNA synthesis, and qRT-PCR. Ten embryos were rinsed in 0.2% BSA/PBS without Ca²⁺ and Mg²⁺ and placed in 0.2 ml PCR tube, immediately transferred in liquid nitrogen, and stored at ~80 °C. It was hybridized with 0.5 µl oligo-dF30 (10 µM, Takara) and 1 µl random (1 M) and 1 µl dNTP mix (10 mM) in 2 µl cell lysis buffer (2 U RNase inhibitor, 0.01% Triton X-100) at 72 °C for 3 min. Then, the reaction was immediately quench on ice. After the reaction tube was centrifuged, 2 µl was used for reverse transcription with Super Script II Reverse Transcriptase 5× first strand buffer, 0.25 µl RNase inhibitor (40 U), 0.06 µl MgCl₂ (1 M), 2 µl betain (3 M), and 0.5 µl Reverse Transcriptase Superscript II (Takara). Reverse transcription was carried out in the thermocycler at 42 °C for 90 min, 70 °C for 15 min, and then 4 °C for holding. Subsequently, cDNA was diluted 1:10 (v/v) with RNase free water and used for a qPCR amplification in triplicate with SYBR...
vortexed vigorously for 10 s and centrifuged at 825×g for 5 min. The pellet was re-suspended in ChIP lysis buffer and incubated on ice for 10 min and vortexed occasionally. Afterwards, the chromatin lysate was transferred to a 1.5 mL centrifuge tube and chromatin sheared using water bath sonication with the following conditions: shear 15 cycles at 4 °C, 15 s on, 30 s off. Centrifuge and transfer supernatant to a new tube. Taking 5 μL (1 %) from the 500 μL containing sheared chromatin as input. Each chromatin sample was incubated with antibodies for H3K9me3 Rabbit polyclonal antibody (1:100, abcam, cat. no. ab8898), H3K27me3 Rabbit mAb (1:50, CST, cat. no. 9733), H3K4me3 Rabbit mAb (1:50, CST, cat. no. 9751), H3K27ac Rabbit mAb (1:100, CST, cat. no. 8173), Nucleolin (D4C7O) Rabbit (1:100, CST, cat. no. 14574), TRIM28 Mouse monoclonal (20C1) (1:100, abcam, cat. no. ab22553) overnight on a rotating platform at 4 °C. The next day, the sample was incubated with protein A+G magnetic beads (HY-K0202, MCE) for 3 h at 4 °C with rotation. The beads-antibody/chromatin complex was washed three times with low-salt wash buffer and once with high-salt wash buffer and resuspended with elution buffer. The elute DNA was treated with RNase A at 42 °C for 30 min, then treated with protease K at 60 °C for 45 min followed by heat inactivation at 95 °C for 15 min. The purified DNA was subjected to library preparation or analyzed by qPCR. The libraries were sequenced on an Illumina platform. All primers of ChIP-qPCR used are listed in Supplementary Table 3.

In situ Hi-C library preparation and sequencing. 106 cells were cross-linked for 10 min in 10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% NP40, and complete protease inhibitors (Roche). The extracted nuclei were re-suspended with 150 μL 0.1% SDS and incubated at 65 °C for 10 min, then SDS molecules were quenched by adding 120 μL water and 30 μL 10% Triton X-100, and incubated at 37 °C overnight. On the next day, the MboI enzyme was inactivated at 65 °C for 20 min. Next, the cohesive ends were filled in by adding 1 μL of 10 mM dTTP, 1 μL of 10 mM dATP, 1 μL of 10 mM dGTP, 2 μL of 5 mM biotin-14-dCTP, 14 μL water and 4 μL (40 U) Klenow, and incubated at 37 °C for 2 h. Subsequently, 663 μL water, 120 μL 10× blunt-end ligation buffer (300 mM Tris-HCl, 100 mM MgCl2, 100 μg/mL BSA, pH 7.9) and 150U of MboI, and incubated at 37 °C overnight. On the following day, the MboI enzyme was inactivated at 65 °C for 20 min. Next, the cohesive ends were filled in by adding 1 μL of 10 mM dATP, 1 μL of 10 mM dTTP, 1 μL of 10 mM dGTP, 2 μL of 5 mM biotin-14-dCTP, 14 μL water and 4 μL (40 U) Klenow, and incubated at 37 °C for 2 h. The purified DNA was subjected to library preparation or analyzed by qPCR. The libraries were sequenced on an Illumina platform.
Bulk RNA-seq data analysis. All bulk RNA-seq reads were trimmed using Trimmomatic software (Version 0.36) with the following settings: ILLUMINA-CLIP/PE36-PE36 redundantly paired reads trimmed to 150 bp and SE reads trimmed to 150 bp. Reads were quality filtered using FASTQ ToolKit (http://hannonlab.cshl.edu/fastx_toolkit/) with quality settings of minimum quality score 20 and minimum percent of 80% that has a quality score larger than 10. The high-quality reads were mapped to the mm10 reference genome using HISAT2 (v2.1.0), a fast-sensitive alignment tool for mapping RNA-seq reads, with -dta parameter78. PCR duplicate reads were removed using Picard tools (https://broadinstitute.github.io/picard/) (v2.18.2). For subsequent analysis on single-copy genes, only uniquely mapped reads were kept. Considering the multi-mapping of reads derived from repeat sequences, we used all mapped reads for further analysis. The expression levels of genes and repeat sequences were independently calculated by StringTie (Version 1.3.4d) with -e -B -G parameters using Release M18 (GRCm38.p6) gene annotations downloaded from GENCODE data portal and annotated repeats (RepeatMasker) downloaded from the UCSC genome browser, respectively. To obtain reliable and cross-comparable expression abundance estimation for each gene and each family of repeat sequence, reads mapped to mm10 were counted as TPM (Transcripts Per Million reads) based on their genome locations. Differential expression analysis of genes in different samples was performed by DESeq2 (v1.32.0) using the reads count matrix produced from a python script “prepair.py” in StringTie website (http://ccb.jhu.edu/software/stringtie/). We selected the genes with stage-specific expression at 0.2 to perform a heatmap clustering analysis using clustermap (v1.0.10) R package. The stage-specific scores of genes expressed during mouse early embryonic development were obtained by entropy-based measure86.

Unsupervised hierarchical clustering was carried out to compare the expression of genes from our study and other reports of 2C-like cell cycle profiles (GSE66582) were included for comparison. TPM were obtained for each sample using the log2 ratio larger than 0 as effective ChIP-seq signals. For meta-analysis of sequencing signals in L1 regions, we only used their subsets proceeded with downstream for each of the 100 reshuffled maps.

Hi-C data analysis. The paired-end reads of fastq files were aligned, processed, and subsequently filtered by Hi-C-Pro software (version 2.11.1)72. Firstly, short reads trimming were cleaned and then independently mapped to the mouse mm10 reference genome (http://hgdownload.soe.ucsc.edu/downloads.html#mouse) using bowtie2 aligner (v2.3.5.1) with end-to-end-algo- rithm and “--very-sensitive” option. To rescue the chimeric fragments spanning the ligation junction, 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. Paired of aligned reads were then assigned to Mbol restriction fragments. Read pairs from the uncut DNA, self-circle ligation and PCR artifacts were filtered out and the valid read pairs involving two different restriction fragments were used to build the contact matrix. To eliminate the possible effects on data analyses of variance in sequencing depths, we randomly sampled equal numbers read pairs from each condition for downstream analyses involving comparison analyses between conditions. Valid read pairs were then binned at a 40, 150, and 300 kb resolutions by dividing the genome into bins of equal size. The binned interaction matrices were then corrected by Knight-Ruiz method using matching method using standard parameter—correctionMethodKR in HiExplorer (https://hiexplorer.readthedocs.io/en/latest/) (v3.3)88. The Observed/Expected (O/E) Hi-C matrix was obtained by HiExplorer hicTransform command with—method obsExp.norm option. Pearson correlation coefficients Hi-C matrix was obtained by HiExplorer hicTrans- form command with—method pearson option.

A and B compartments were identified using the first eigenvector (PCI) from principal component analysis on correlation Hi-C matrix89. We used Homer software (v4.11) with parameters -res 500,000 -window 1,000,000 to obtain the PCI value based on Pearson Correlation Coefficients (PCI) Hi-C matrix. Sometimes the entry signs of PCI were inverted to ensure correct orientations of Hi-C interactions for individual regions. As GC content is well correlated with A and B compartments90, we calculated the GC content of each region and inverted the eigenvector sign if the average GC content of negative-eigenvector entries is higher than that of positive-eigenvector entries. To obtain the heatmap plot of enrichment of A/B interaction, an A/B compartment profile for each chromosome was then separated into five bins (min to max): 0 (0–1), 1 (1–2), 2 (2–3), 3 (3–4), 4 (4–5) at 40 kb resolution. For each domain of length L, a map for the region (((Start – L) to (End + L))) was obtained. This produced a contact that is three times bigger than a given domain. This contact map was then rescaled to a (90 × 90) pixel map using linear interpolation and block-averaging. In the resulting map, the mid-domain contact density was calculated as the sum of the central square of the enrichment map, rows 30–59 and columns 3–5 of compartment enrichment matrix AB. The Observed/Expected (O/E) Hi-C matrix was obtained by HiExplorer hicTransform command with—method obsExp.norm option. Pearson correlation coefficients Hi-C matrix was obtained by HiExplorer hicTrans- form command with—method pearson option.

Hi-C data analysis. The paired-end reads of fastq files were aligned, processed, and subsequently filtered by Hi-C-Pro software (version 2.11.1)72. Firstly, short reads trimming were cleaned and then independently mapped to the mouse mm10 reference genome (http://hgdownload.soe.ucsc.edu/downloads.html#mouse) using bowtie2 aligner (v2.3.5.1) with end-to-end-algo- rithm and “--very-sensitive” option. To rescue the chimeric fragments spanning the ligation junction, 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. Paired of aligned reads were then assigned to Mbol restriction fragments. Read pairs from the uncut DNA, self-circle ligation and PCR artifacts were filtered out and the valid read pairs involving two different restriction fragments were used to build the contact matrix. To eliminate the possible effects on data analyses of variance in sequencing depths, we randomly sampled equal numbers read pairs from each condition for downstream analyses involving comparison analyses between conditions. Valid read pairs were then binned at a 40, 150, and 300 kb resolutions by dividing the genome into bins of equal size. The binned interaction matrices were then corrected by Knight-Ruiz method using matching method using standard parameter—correctionMethodKR in HiExplorer (https://hiexplorer.readthedocs.io/en/latest/) (v3.3)88. The Observed/Expected (O/E) Hi-C matrix was obtained by HiExplorer hicTransform command with—method obsExp.norm option. Pearson correlation coefficients Hi-C matrix was obtained by HiExplorer hicTrans- form command with—method pearson option.

A and B compartments were identified using the first eigenvector (PCI) from principal component analysis on correlation Hi-C matrix89. We used Homer software (v4.11) with parameters -res 500,000 -window 1,000,000 to obtain the PCI value based on Pearson Correlation Coefficients (PCI) Hi-C matrix. Sometimes the entry signs of PCI were inverted to ensure correct orientations of Hi-C interactions for individual regions. As GC content is well correlated with A and B compartments90, we calculated the GC content of each region and inverted the eigenvector sign if the average GC content of negative-eigenvector entries is higher than that of positive-eigenvector entries. To obtain the heatmap plot of enrichment of A/B interaction, an A/B compartment profile for each chromosome was then separated into five bins (min to max): 0 (0–1), 1 (1–2), 2 (2–3), 3 (3–4), 4 (4–5) at 40 kb resolution. For each domain of length L, a map for the region (((Start – L) to (End + L))) was obtained. This produced a contact that is three times bigger than a given domain. This contact map was then rescaled to a (90 × 90) pixel map using linear interpolation and block-averaging. In the resulting map, the mid-domain contact density was calculated as the sum of the central square of the enrichment map, rows 30–59 and columns 3–5 of compartment enrichment matrix AB. The Observed/Expected (O/E) Hi-C matrix was obtained by HiExplorer hicTransform command with—method obsExp.norm option. Pearson correlation coefficients Hi-C matrix was obtained by HiExplorer hicTrans- form command with—method pearson option.

A and B compartments were identified using the first eigenvector (PCI) from principal component analysis on correlation Hi-C matrix89. We used Homer software (v4.11) with parameters -res 500,000 -window 1,000,000 to obtain the PCI value based on Pearson Correlation Coefficients (PCI) Hi-C matrix. Sometimes the entry signs of PCI were inverted to ensure correct orientations of Hi-C interactions for individual regions. As GC content is well correlated with A and B compartments90, we calculated the GC content of each region and inverted the eigenvector sign if the average GC content of negative-eigenvector entries is higher than that of positive-eigenvector entries. To obtain the heatmap plot of enrichment of A/B interaction, an A/B compartment profile for each chromosome was then separated into five bins (min to max): 0 (0–1), 1 (1–2), 2 (2–3), 3 (3–4), 4 (4–5) at 40 kb resolution. For each domain of length L, a map for the region (((Start – L) to (End + L))) was obtained. This produced a contact that is three times bigger than a given domain. This contact map was then rescaled to a (90 × 90) pixel map using linear interpolation and block-averaging. In the resulting map, the mid-domain contact density was calculated as the sum of the central square of the enrichment map, rows 30–59 and columns 3–5 of compartment enrichment matrix AB. The Observed/Expected (O/E) Hi-C matrix was obtained by HiExplorer hicTransform command with—method obsExp.norm option. Pearson correlation coefficients Hi-C matrix was obtained by HiExplorer hicTrans- form command with—method pearson option.

A and B compartments were identified using the first eigenvector (PCI) from principal component analysis on correlation Hi-C matrix89. We used Homer software (v4.11) with parameters -res 500,000 -window 1,000,000 to obtain the PCI value based on Pearson Correlation Coefficients (PCI) Hi-C matrix. Sometimes the entry signs of PCI were inverted to ensure correct orientations of Hi-C interactions for individual regions. As GC content is well correlated with A and B compartments90, we calculated the GC content of each region and inverted the eigenvector sign if the average GC content of negative-eigenvector entries is higher than that of positive-eigenvector entries. To obtain the heatmap plot of enrichment of A/B interaction, an A/B compartment profile for each chromosome was then separated into five bins (min to max): 0 (0–1), 1 (1–2), 2 (2–3), 3 (3–4), 4 (4–5) at 40 kb resolution. For each domain of length L, a map for the region (((Start – L) to (End + L))) was obtained. This produced a contact that is three times bigger than a given domain. This contact map was then rescaled to a (90 × 90) pixel map using linear interpolation and block-averaging. In the resulting map, the mid-domain contact density was calculated as the sum of the central square of the enrichment map, rows 30–59 and columns 3–5 of compartment enrichment matrix AB. The Observed/Expected (O/E) Hi-C matrix was obtained by HiExplorer hicTransform command with—method obsExp.norm option. Pearson correlation coefficients Hi-C matrix was obtained by HiExplorer hicTrans- form command with—method pearson option.
bins. We normalized insolation scores by dividing each region’s score by the average scores of the nearest 50 regions, and log2-transforming the resulting vector, thus accounting for local biases in insolation score. Visualization of Hi-C matrix was carried out by the Juicer tool (https://github.com/aidenlab/juicer) (v1.9.9) and R software (https://www.r-project.org/) (v4.0.2). For heatmap visualization in Juicer, we converted valid read pairs into .hic format files with the juicer tool pre command.

Statistical analysis. All statistical analyses for Next Generation Sequencing (NGS) data were performed with R (v4.0.2)/Biocductor (v3.10) software utilizing custom R scripts. The other statistical analyses were performed with GraphPad Prism 8 software. For p-value produced by GraphPad, only 15 digits after decimal, the upper limit of this software, are shown in the figure legend. Details of individual tests are outlined within each figure legend, including number of replications performed (n) and the reported error as standard error of the mean (SEM). All statistics are *p < 0.05, **p < 0.01, ***p < 0.001, and were calculated by Wilcoxon signed rank test (for paired samples), Mann-Whitney U-test (for independent samples), two-way ANOVA and chi-square test as described in the figure legends.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The data that support this study are available from the corresponding author upon reasonable request. All bulk RNA-seq, single-cell RNA-seq, ChIP-seq, ATAC-seq, and Hi-C data generated in this study have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database under the accession codes GSE166041 and GSE164420. Previously published RNA-Seq data that were re-analyzed here are available under accession codes GSE33923, GSE51682, GSE63525. Supplementary Table 4 provides a summary for all analyzed NGS datasets used in this study. Source data are provided with this paper.

References

1. Macfarlan, T. S. et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 487, 57–63 (2012).
2. Choi, Y. J. & Lin, C. P. De novo formation of inter-chromosomal hubs by the transcription factor KAP1. Cell Rep. 22, 237–240 (2018).
3. Zalzman, M. et al. Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 464, 858–863 (2010).
4. Matsui, T. et al. Provilc silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature 464, 927–931 (2010).
5. Rowe, H. M. et al. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature 463, 237–240 (2010).
6. Fu, X., Wu, X. & Djeidkel, M. N. Myc and Dnm1e impede the pluripotent to totipotent stage transition in embryonic stem cells. Nat. Cell Biol. 21, 835–844 (2019).
7. Dan, J. et al. Zscan4 inhibits maintenance DNA methylation to facilitate telomere elongation in mouse embryonic stem cells. Cell Rep. 20, 1936–1949 (2017).
8. Guiller, D. et al. RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. Nat. Genet. 50, 443–451 (2018).
9. Yang, B. X. et al. Systematic identification of factors for proovir silencing in embryonic stem cells. Cell 163, 230–245 (2015).
10. De Iaco, A. et al. DUX-family transcription factors regulate syzygic genome activation in placental mammals. Nat. Genet. 49, 941–945 (2017).
11. Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat. Genet. 49, 925–934 (2017).
12. Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. Nat. Genet. 49, 935–940 (2017).
Bolger, A. M., Marc, L. & Bjoern, U. Trimmomatic: a 20
NATURE COMMUNICATIONS | (2021) 12:6365 | https://doi.org/10.1038/s41467-021-26576-2 | www.nature.com/naturecommunications

Wang, S. et al. Spatial organization of chromatin domains and compartments in single chromosomes. Science 353, 598–602 (2016).

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging: spatially resolved, highly multiplexed RNA profiling in single cells. Science 348, 983–985 (2015).

Bolger, A. M., Marc, L. & Bjorn, U. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).

Daehwan, K., Ben, L. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360 (2015).

Mihaela, P. et al. StringFISH enabling improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33, 290–295 (2015).

Cabili, M. N. et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genome Dev. 25, 1915 (2011).

Tsai, K. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatica 28, 882–883 (2012).

Zilionis, R. et al. Single-cell barcoding and sequencing using droplet microfluidics. Nat. Protoc. 12, 44–73 (2016).

Saitia, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. 33, 495–502 (2015).

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

Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165 (2016).

Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–590 (2010).

Servant, N. et al. HIC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 16, 259 (2015).

Yildiz, R. et al. High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat. Commun. 9, 189 (2018).

Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293 (2009).

Iakoumakis, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods 9, 999–1003 (2012).

Flyyänen, J. et al. Single-nucleus Hi-C reveals unique chromosome reorganization at oocyte-to-zygote transition. Nature 544, 110–114 (2017).

Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 3, 95–98 (2016).
