rRNA Biogenesis Regulates Mouse 2C-like State by 3D Structure Reorganization of Peri-Nucleolar Heterochromatin

Hua Yu[1, #], Zhen Sun[1, #], Tianyu Tan[1, #], Hongru Pan[1, #], Jing Zhao[1], Ling Zhang[1], Jiayu Chen[3], Anhua Lei[1], Yuqing Zhu[1], Lang Chen[1], Yuyan Xu[1], Ming Chen[3], Shaorong Gao[4], George Q. Daley[5], Jin Zhang[1, 2, 6, *]

[1] Center for Stem Cell and Regenerative Medicine, Department of Basic Medical Sciences, Zhejiang University School of Medicine, Hangzhou, 310058, China. Institute of Hematology, Zhejiang University, Hangzhou, 310058, China.

[2] Zhejiang Laboratory for Systems & Precision Medicine, Zhejiang University Medical Center, 1369 West Wenyi Road, Hangzhou 311121, China.

[3] College of Life Sciences, Zhejiang University, Hangzhou, China

[4] Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, Shanghai Key Laboratory of Signaling and Disease Research, School of Life Sciences and Technology, Tongji University, Shanghai, China

[5] 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.[6] Lead Contact.

[#] These authors contributed equally to this work.

[*] Correspondence: zhgene@zju.edu.cn.

Abstract

Nucleolus is the organelle for ribosome biogenesis and for sensing various types of stress. Its role in regulating stem cell fate is unclear. Here, we present multiple lines of evidence that nucleolar stress induced by interfering rRNA biogenesis can drive 2-cell stage embryo-like (2C-like) transcriptional program and induce an expanded 2C-like cell population in mouse embryonic stem (mES) cells. Mechanistically, the nucleolar integrity mediated by rRNA biogenesis maintains the normal liquid-liquid phase separation (LLPS) of nucleolus and the formation of peri-nucleolar heterochromatin (PNH). Upon rRNA biogenesis defect, the natural LLPS of nucleolus is disrupted,

causing dissociation of NCL/TRIM28 complex from PNH and changes of epigenetic states and reorganization of the 3D structure of PNH, which leads to *Dux*, a 2C program transcription factor gene, to be released from the PNH region and activation of 2C-like program. Correspondingly, embryos with rRNA biogenesis defect are incompatible 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 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.

Main

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 [1-3]. 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 [1,4,5] and DNA methylation [6,7]. In addition, it was also found that RNA hydroxymethylation [8] and protein sumoylation [9] can 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 [10-15], 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 [16,17]. Moreover, 2C-like cells can be induced by

downregulation of chromatin remodeling factor CAF-1 [18]. 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 [19-22]. Two-cell embryo or 2C-like cells show contrasting differences from inner cell mass or ES cells [21,23], 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[14,24,25]. However, it is not completely known how nucleolus integrity influences higher-order chromatin structure and how the chromatin structure determines *Dux* expression.

Here, we found 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, 3D structure of 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 for the first time provided a novel mechanistic perspective of rRNA biogenesis in regulating the homeostasis between 2C-like and mES cells and highlighted 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: 2uM, CX-5461 treatment time: 12h; rotenone treatment dosage: 1uM, rotenone treatment time: 12h; rapamycin treatment dosage: 2uM, rapamycin treatment time: 12h). We found that nucleolar stress induced by CX-5461[26,27] activated the 2-cell marker genes *Zscan4d*, *Dux* and *Gm12794* and repressed pluripotent marker gene *Pou5f1*. However, the other two cellular stresses upon rotenone or rapamycin treatment did not influence the expression of these genes (Fig.1a-1b and Extended Data Fig.S1a-S1b). 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 Extended Data Fig.S1e), particularly GSAT_MM, MERVL-int and MT2_Mm sub-classes, in CX-5461-treated mES cells (Fig.1c and Extended Data Fig.S1c-S1d). However, the other two types of stress, did not activate these repeat elements (Fig.1c and Extended Data Fig.S1c-S1d). Using unsupervised K-means clustering analysis, we identified four gene clusters specifically expressed in different stages during mouse pre-implantation embryo development (Extended Data Fig.S1f)[28]. 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 Extended Data Fig.S1f). Yet, other two cellular stresses did not induce this expression pattern (Fig.1d and Extended Data 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 2CLC signatures from other studies as well as the 2C embryos[1,11,14,16-18,29-31] (Fig.1e). As expected, we

observed the abundance of rRNA is significantly reduced under CX-5461 treatment (Extended Data 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 (Extended Data Fig.S1h-S1j). As expected, we found a marked expansion of the population with MERVL expression in mES cells treated with CX-5461 (Fig.1f). Strikingly, we observed that the expression level of MERVL genes showed significant negative correlation with ribosomal protein genes (Fig.1g-1i and Extended Data Fig.S1k). 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.1j and Extended Data Fig.S1l). Moreover, the percentage of tdTomato positive cells was largely maintained even at 24 hours after CX-5461 withdrawal (Fig.1k). 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[32] (Extended Data Fig.S1m-S1o), 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 [33,34].

Deficiency of rRNA biogenesis disrupted normal nucleolar LLPS and epigenetic state of PNH

As it has been reported that ribosomal RNA plays a critical role in maintaining phase

separation of nucleolus[35-37] and phase-to-phase transition is involved in nucleolar stress[37,38], 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.2a). 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- and NCL-marked “ring” structure (Fig.2b-2c). Immunofluorescence of FBL and RPA194 protein also showed abnormal localization with aggregated pattern in the nucleolus upon treatment (Fig.2d-2e). Consistently, Fluorescence Recovery After Photobleaching (FRAP) analysis revealed that CX-5461 treatment markedly increased the mobility of NCL, NPM1 and FBL (Fig.2f-2h). Together, these data demonstrate that nucleolar stress caused disrupted assembly of the phase-separated nucleolar sub-compartments, which became fused and more dynamic liquid-like droplets[37,38].

As it has been observed that phase separation can regulate the epigenetic state of chromatin[39-41], 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 nucleolus[42]. In addition, the Nucleolar Associated Domains (NAD) or LINE1/L1 repeat sequence regions are also defined as repressive chromosomal segments enriched with peri-nucleolar heterochromatin[43-49]. We thus examined the epigenetic changes on these regions and found decreased H3K9me3 and H3K27me3 levels in CX-5461 treated cells (Fig.2i-2k, Extended Data Fig.S2a and Extended Data Fig.S2b). Moreover, we observed increased H3K4me3 and H3K27ac levels and improved chromatin accessibility at Inactive Hub[42], NAD[43,45] and L1 regions (downloaded from UCSC Table Browser) in CX-5461 treated cells (Fig.2l-2n, Extended Data Fig.S2c-S2g). 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 cells[14]. The disappeared NCL-marked "ring" structure suggested that rRNA 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.2o-2q, Extended Data Fig.S2e and Extended Data Fig.S2f). Altogether, these results demonstrated the rRNA biogenesis defect affected the normal nucleolar phase separation and changed the epigenetic state of the heterochromatic regions at the periphery of nucleolus 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 transcription[10-12]. As *Dux* expression has been reported to be influenced by nucleolar protein NCL[14], 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.3a). In line with this, we found that 1229 CX-5461-induced genes are p53 direct target genes (Fig.3b, hypergeometric test, p-value=0)[50], consistent with the fact that p53 signaling is usually activated under nucleolar stress[51-57]. We further analyzed the overlap between CX-5461 treatment-induced genes and *Dux* over-expression-induced genes using published RNA-seq data [11], and found 621 genes are overlapped (Fig.3b, hypergeometric test, p-value= 2.96e-195). Using p53 and Dux ChIP-seq data in mES cells [11,50], 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.3c). 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.3d), and to ERV genes induced by CX-5461 (Fig.3e-3f). 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 10173 CX-5461-induced ERV genes in *Dux*-overexpressed mES cells [11] (Extended Data Fig.S3a). In addition, we observed the decreased H3K9me3 and H3K27me3 levels and increased H3K4me3 and K3K27ac levels around Dux locus (Extended Data Fig.S3b). 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.3g). 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 are increased, in contrast, Dux binding is decreased for CX-5461 induced 2C genes (Fig 3h-3j). When the Dux silenced mES cells are treated with CX-5461, we observed that H3K9me3 and H3K27me3 levels, and Dux binding on CX-5461 induced 2C genes are reversed (Fig 3h-3j). 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 is decreased in Dux silenced mES cells, and this reduction is reversed by CX-5461 treatment (Fig 3k). These results demonstrated that nucleolar stress induced 2C/ERV gene activation through a 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 embryos [28,29,58] (Extended Data Fig.S3c-S3h). In addition, we also found increased H3K4me3 and H3K27ac levels of 621 commonly induced

genes and 10173 CX-5461 induced ERV genes in CX-5461 treated mES cells (Extended Data Fig.S3i-S3j). 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 more than four hundred million sequenced raw read pairs per sample. We observed obviously decreased higher-order chromatin interactions within PNH region indicated by the Inactive Hub, NAD and L1 regions in the treated cells (Fig.4a-4c, compared with randomly selected genomic regions, Mann-Whitney U test, the replicates of experiment n=10, averaged *p*-values=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.4a-4c and Fig.4g-4i, compared with randomly selected genomic regions, Mann-Whitney U test, the replicates of experiment n=10, averaged *p*-values=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*-values=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, *p*-values=1.86E-6, 5.03E-6, 4.02E-5 for Inactive Hub, NAD and L1, respectively) (Fig.4d-4f and Fig.4j-4l). 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.4m). Interestingly, by performing the same analysis as above on public Hi-C data of mouse pre-implantation embryo development[21], we observed that a similar trend of 3D chromatin structure reorganization of PNH and *Dux* regions when comparing early 2-cell embryos with ICM stage embryos, namely, the 2-cell embryos showed less organized PNH and less contacts between the *Dux* Locus and the PNH (Extended Data Fig.S4a-S4d), 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 2-cell to late 2-cell transition, which coincides with the beginning of shutting down *Dux* gene expression during murine early embryo development (Extended Data Fig.S4a-S4d and Extended Data Fig.S6d).

As it has been reported that 2CLCs display increased three-dimensional structural plasticity relative to mES cells[23], 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 full differentiated cells[59]. 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 (Extended Data Fig.S4e-S4h). However, at topologically associating domain (TAD) or chromatin loop level, we found a mild increase in their strength in CX-5461-treated mES cells (Extended Data Fig.S4i-S4m). To specifically investigate 2C-related genes, we further performed an analysis of local architectural differences around *MERVL* loci. We found that the insulation scores[60] of chromatin around *MERVL* genes activated by CX-5461 treatment is markedly increased both globally (Extended Data Fig.S4n) and at local *MERVL* sites (Fig.4o), similar as observed in 2-cell embryos compared with ICM[23] (Fig.4o), and the topological associated domain (TAD) structure around *MERVL* gene loci is more obvious (Fig.4o and Extended Data Fig.S4o), showing a more similar pattern as that in 2-cell embryos[21]

(Fig.4n). The chromatin structure reorganization around *MERVL*s is accompanied with more open chromatin states around *MERVLs* both globally (Extended Data Fig.S4n) and locally (Fig.4o and Extended Data Fig.S4o) and with their increased expression (Extended Data Fig.S4n). 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 cell lines: 1) a line with degraded Pol I protein (PRA1) by an auxin-inducible degron system[61] (Fig S5a), and 2) a snoRNA knockout line (SNORD113-114 gene cluster, a gift from Pengxu Qian) (Fig S5b), 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.5a-5d). 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.5e-5h and Extended Data Fig.S5c-S5f). 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.5i and Fig.5j). 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.5k, Fig.5l, Extended Data Fig.S5g and Extended Data Fig.S5h). 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.

rRNA biogenesis is critically required at the 2-cell-to-4-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 2-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 2-cell and middle 2-cell stages but increased sharply from the late 2-cell stage to the blastocyst stage (Fig.6a). We then analyzed the expression levels of different subunit genes of RNA polymerase I (Pol I) in pre-implantation embryos. Different from rRNA expression, Pol I gene mRNA levels increased significantly from the late 2-cell stage to 4-cell stage but decreased markedly as embryos progressed through 8-cell, morula stages, and reached to blastocysts which still had a higher level than those stages before the late 2-cell (Fig.6b). Consistent with Pol I genes, we observed the similar pattern of increased expression of ribosome biogenesis genes (Extended Data Fig.S6a). This indicated that while the levels of rRNA was gradually accumulated during pre-implantation embryo development, the rRNA biogenesis rate reached to a peak during the late 2-cell to the 4-cell stage. In contrast, the ERV and 2C marker genes, such as *Dux*, *Zscan4d* and *Gm12794* were significantly decreased during the late 2-cell-to-4-cell stage (Extended Data Fig.S6b-S6d). 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 2-cell to the 4-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 4-cell stage (Extended Data Fig.S6e). We further divided mouse embryos into four groups according to the different stages of CX-5461 treatment, including transitions of zygote-to-2-cell, 2-cell-to-4-cell, and morula-to-blastocyst, respectively (Fig.6c). Compared with the untreated control, we found that blastocyst formation rates of all

three CX-5461-treated groups were decreased, and the 2-cell-to-4-cell-treated group showed the strongest decrease of the blastocyst formation rate at both early and late blastocyst stages (Fig.6d). This is consistent with the pattern of Pol I gene RNA expression and the pattern of PNH reshaping after early 2-cell stage during pre-implantation embryo development (Extended Data Fig.S4a-S4d), indicating that rRNA biogenesis is most critically required during the 2-cell-to-4-cell transition compared with other stages. In well support with this, the knockout of *Polr1a* gene led to mouse embryos arrested at 2-cell [62]. Moreover, upon successful inhibition of rRNA biogenesis at the morula/blastocyst stage (Fig.6e), we found the disappearance of the NPM1- and NCL-marked "ring" structure and abnormal localization and reduced signal density of FBL in the nucleolus (Extended Data Fig.S6f). Importantly, we observed increased expression of 2C genes such as *Zscan4d*, *Gm4340*, *Dux* and *Mervl-pol* (Fig.6f). RNA-seq also demonstrated up-regulated C1 and C2 clusters of 2C genes (Fig 6g) defined before (Fig S1e) as well as ERV genes (Fig 6h-6j) in CX-5461-treated embryos compared with controls, consistent with the results from mES cells described above (Fig.1c and Extended Data Fig.S1d). Altogether, these data demonstrated rRNA biogenesis and nucleolar integrity is a molecular switch for the transition from the 2-cell to the 4-cell embryos.

Discussion

Starting from zygotic genome activation (ZGA) at the 2-cell stage, an embryo undergoes through the four/eight-cell, morula, blastocyst stages, and then prepares itself for implantation. Along with this pre-implantation development process, basic anabolic metabolism and translational processes become more active. Nucleoli, the organelles involved in translation, functionally mature from nucleolar precursor bodies (NPB) during this process[63-65]. Interestingly, a recent research work reported that TRIM28/Nucleolin/LINE1 complex that can mediate both ZGA gene Dux repression and rRNA expression[14], suggesting that shutting down ZGA and initiating nucleoli formation are not independent events but interconnected. However, a complete

picture on how the transition between ZGA and nucleolar formation occurs in the nucleus is not fully understood.

Here, we reported that nucleolar rRNA biogenesis and higher-order 3D chromatin structure remodeling of PNH might coordinate to develop during the 2-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 novel role of rRNA biogenesis in regulating the 2C-like program and the homeostasis between 2C-like cells and mES cells (Fig 7). In the unperturbed mES cells, nucleolar integrity mediated by rRNA biogenesis maintains the normal the liquid-liquid phase separation (LLPS) of nucleolus and the formation of peri-nucleolar heterochromatin (PNH) containing *Dux*, and this normal nucleolar LLPS facilitated NCL/TRIM28 complex occupancy on the *Dux* locus to repress *Dux* expression. In contrast, in the rRNA biogenesis-repressed mES cells, the natural liquid-like phase of nucleolus is disrupted, causing dissociation of the NCL/TRIM28 complex from the PNH and changes of epigenetic state and 3D structure of the PNH, which eventually leads to *Dux* to be released from the 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 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[37,38]. 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[44,57,66]. In contrast, our work for the first time provided a novel 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[39-41,67-70], and shed light on a previously neglected area of nucleolus-associated condensates in chromatin control during early development. It is being worth mentioned that we do not intend to overstate that the fate transition of mES to 2C-like cell triggered by rRNA biogenesis defect are explained by LLPS[71]. What we observed is that the integrity of nucleolus mediated by rRNA biogenesis maintains the normal nucleolar LLPS and 3D structure of PNH. It is possible that 3D structure reshaping of PNH mediated by nucleolar LLPS is a common thread of RNA and protein-mediated Dux silences and 2C repression (e.g., rRNA, snoRNA, LINE1 RNA, NCL, TRIM28 and LIN28). The NCL/TRIM28 complex or other nucleolar proteins of GC layer and heterochromatin proteins, e.g., NPM1 and HP1, is possible the key regulating factors connecting nucleolar LLPS and the establishment and maintenance of PNH. Our study gives a starting clue that the LLPS model is important for the assembly and function of nucleolus[37] with the implication of gene regulation, and the quantitative and mechanistic models merit future investigations.

Methods

Cell culture

E14 mES cells were cultured on 0.1% gelatin-coated plates with MEF feeder cells in N2/B27/LIF/2i medium (1:1 mix of DMEM/F12 (11320-033, Gibco) and Neurobasal medium (21103-049, Gibco) containing 1×N2 and B27 supplements (17502-048/17504-044, Life Technologies), 100 μM non-essential amino acids (GNM71450, GENOM), and 1,000 U/ml LIF (PEPRO TECH), 1 μM PD03259010 and 3 μM CHIR99021 (STEMCELL Technologies) and 100 U/ml penicillin, 100 μg/ml streptomycin (15140-122, Gibco). For primed state media, 20 ng/ml Activin, 10 ng/ml FGF2, and 1% KSR were added to the 1:1 DMEM/F12 and Neurobasal medium containing N2 and B27. To investigate DUX binding, an N-terminal FLAG-DUX protein was expressed in our clonal cell lines. In Control group, mES cells were treated with doxycycline for 12h to induce FLAG-DUX expression and then treatment of negative-control Silencer Select siRNA. In

siDux group, mES cells were treated with doxycycline for 12h and then siDUX for two days. In CX-5461 treatment group, mES cells were treated with doxycycline for 12h and then treatment of CX5461. In siDux+CX-5461 group, mES cells were treated with doxycycline for 12h and then siDUX for two days followed by treatment of CX-5461.

Fluorescence activated cell sorting (FACS) analysis

The E14 2C::*tdTomato* cells were generated as in [1]. E14 mES cells were transfected with 2C::*tdTomato* using Lipofectamine 2000 and selected with 150 μg/ml hygromycin 48 hr after transfection and for 7 days. Mouse E14 wild-type cells were subjected to 0.4μM CX-5461 treatment for 12h or 2μM CX-5461 treatment for 12h. Cells were isolated by FACS to measure the ratio of 2C-like cells. Apoptosis was measured using the Annexin V and DAPI Staining.

Cell line immunofluorescence staining

E14 mES cells were grown on gelatin-coated glass coverslips with MEFs and cultured 12 h before fixed with 4% PFA for 10 min, and then permeabilized with 0.5% Triton X-100 in PBS for 20 min at room temperature (RT). The cell samples were blocked in blocking buffer (3% BSA, 2% donkey serum in PBS) for 10 min at RT and then stained with a primary antibody (1:200, Nucleolin, CST, 145745; 1:400, NPM1, Sigma, B0556; 1:200, Fibrillarin , Abcam, ab4566; RPA194, 1:50, Santa Cruz, sc-48385) for 12 h at 4 °C. After three washes with 0.1% Triton X-100 in PBS, cells were stained with a secondary antibody (1:200, Goat polyclonal Secondary Antibody to Mouse IgG, Abcam, ab150113) for 2-12 h at 4 °C. Followed by washing three times with 0.5% Triton X-100/PBS, DAPI was used for nucleus staining. The samples were then imaged by Zeiss LSM880 fluorescence microscope at a 63×oil objective. For high regulation microscopy imaging, LSM800 with Airyscan module was used.

Fluorescence recovery after photobleaching (FRAP) analysis

E14 wild-type and CX-5461 treatment mES cells cultured on MEF cells were grown in N2B27/LIF/2i conditions and maintained at 37 °C and with 5% CO2 during image

acquisition. Cells were transduced with Lenti-NCL-GFP / Lenti-NPM1-GFP / Lenti-FBL-mCherry lentivirus. FRAP experiments were performed on a ZEISS (Jena, Germany) LSM800 confocal laser scanning microscope equipped with a ZEISS Plan-APO 63x/NA1.46 oil immersion objective. Circular regions of constant size were bleached and monitored overtime for fluorescence recovery. Bleaching was once every 10 seconds for a total of 10 minutes. Fluorescence intensity data was corrected for background fluorescence and normalized to initial intensity before bleaching using GraphPad software. Resulting FRAP curves were fitted with Four parameter logistic (4PL) curve.

siRNA-mediated knockdown in mES cells

siRNA transfections were performed in mES cells with Lipofectamine 2000 (Thermo Fisher Scientific). mES cells were seeded into 12-well plate and cultured in LIF/2i medium for overnight. The next day, 800 µl LIF/2i medium without antibiotics was added into each well. Then, the transfect mixture (40 pmol of 3 independent siRNA targeting each gene/a non-targeting siRNA (negative control, NC) and 2 µl of Lipo 2000 which was diluted in 200 µl Opti-MEM medium (Gibco)) was added into each well and incubated for 6 hr at 37 °C. After incubation, the medium was exchanged for fresh complete LIF/2i medium and cells were harvested for RNA extraction approximately 48 hr later. The sequences of siRNA are listed in Supplementary Table 1.

Cell line RNA extraction and qRT-PCR

Total RNA was isolated from mES cells using miRNeasy kit (217004, QIAGEN) according to the manufacturer's protocol, and 1 µg RNA was reverse transcribed to cDNA with HiScript II Q RT Super Mix (R223-01, Vazyme). Gene expression was analyzed with SYBR-Green qPCR Master mix (Bio-Rad) on Bio-Rad PCR machine (CFX-96 Touch). Each gene was normalized to Actin or Gapdh. All primers used are listed in Supplementary Table 2.

Oligopaint DNA FISH

The Oligopaint DNA FISH probe was designed according to several previous publication. The encoding probes were designed as previous described[72,73] and the targeted genomic region was designed by OligoMiner[74]. The probe pools synthesized by Synbio Technologies were used as templates and the dye-labeled secondary probes were produced by Sunya Biotechnology. All sequences used in this work were listed in Supplementary Table 3. Briefly, the synthesized probe pool was first used as template to amplify via 30 PCR cycles and was subsequently purified by ammonium acetate precipitation. Then, the previous PCR products was used as template to amplify and convert into RNA via an *in vitro* transcription of high yield (New England Biolabs, E2040S); the RNA product above were converted back into single-stranded DNA via reverse transcription. At last, the product was subjected to the alkaline hydrolysis to remove template RNA and was further purified by ammonium acetate precipitation. For the secondary probe, a 30 bp random oligo was designed as previous and attached with Cy3 or Alexa Flour 647 at 5' end.

For DNA FISH, the cell samples were fixed in 4% Paraformaldehyde (Sigma, 158127) for 10 minutes and washed two times with PBS, followed by permeabilizing with 0.5% Triton X-100 (Sigma, T8787) in PBS. Then the samples were incubated in 0.1% w/v sodium borohydride (Sigma, 71320) for 10 minutes and treated with 0.1M HCl for 5 minutes. After that, the cell samples were incubated in 0.1 mg/ml RNase A diluted in PBS for 45 minutes in 37℃. After washing three times in 2x SSCT (2x SSC + 0.1% Tween-20), the samples were immersed in 50% formamide diluted in 2x SSCT for 15 minutes at room temperature then transferred to 85˚C for 10 minutes. The primary probe and secondary probe were freshly mixed into hybridization buffer (2x SSC, 50% formamide, 20% dextran sulfate) at 6 μM and 1 μM final concentration and dropped on samples. The samples were heated at 85 ˚C for 20 minutes 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, 145745) for 12 h at 4℃. After three times of washing with PBS, Donkey

anti-Rabbit secondary antibody (1:200, Abcam, 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, E14 mES cells cultured in one 10 cm dish were collected, cleaned from feeder cells, and supplemented with 2.5% glutaraldehyde. The cell pellet was dispersed into small clusters and fixed at least 6 hr at 4 °C. After that, the cell samples were treated with standard procedures. Then the slices were imaged on FEI Spirit 120 kV LaB6 Routine Cryo-EM Capable Electron Microscope.

Embryo collection and culture

The C57BL/6J mice were housed in the animal facility of Zhejiang University. All experimental procedures were performed in accordance with the Animal Research Committee guidelines of Zhejiang University. To collect pre-implantation embryos, C57BL/6J female mice (4–6 weeks old) were intraperitoneally injected with 7.5 IU each of PMSG (San-Sheng Pharmaceutical) for 48h followed by injection of 7.5 IU of hCG (San-Sheng Pharmaceutical). The superovulated female mice were mated with adult males overnight after hCG administration. Embryos at different stages of pre-implantation development were collected at defined time periods after the administration of hCG: 30 h (early 2-cell), 44–48 h (2-cell), 54–56 h (4-cell), 68–70 h (8-cell), 76-78 h (morula) and 92–94 h (blastocysts) in HEPES-buffered CZB medium. Zygotes were collected from ampullae of oviducts and released with hyaluronidase for removing cumulus cells.

Embryo immunofluorescence 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 minutes in PBS/

0.25% Triton X-100 and blocked-in blocking buffer (PBS/0.2% BSA/0.01% Tween-20) for 1 hr 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, Santa Cruz, sc-48385) at 4 degree or for 1hr at 37 °C. Subsequently, embryos were washed four times for 10 min each and incubated with a secondary antibody (daylight 488-conjugated anti-rabbit, 1:100 or daylight 594-conjugated anti-mouse, 1:200) for 1hr at 37°C and washed three times with PBS. Nuclei were stained with DAPI for 1 min. Embryos were observed under Zeiss LSM880 fluorescence microscope at 63× magnification with an oil immersion objective.

Embryo collection, cDNA synthesis and qRT-PCR

10 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-dT30 (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 quenched on ice. After the reaction tube was centrifuged, 2 µl was used for reverse transcription with Super Script II Reverse Transcriptase 5x first strand buffer, 0.25 µl RNase inhibitor (40 U), 0.06 µl MgCl2 (1 M), 2 µl betaine (5 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 Green Master (Vazyme) in a final volume of 20 µl per reaction as manufacturer's instructions.

Embryo RNA-seq library preparation and sequencing.

Embryos were collected (5 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

582 (TruePrep DNA Library Prep Kit V2 for Illumina, Vazyme). Libraries were sequenced on
583 a HiSeq-PE150, with paired end reads of 150 bp length each.
584
585 ## Bulk RNA-seq library preparation and sequencing
586 A total amount of 2 μg RNA per sample was used as input materials for the RNA sample
587 preparation. mRNA was purified from total RNA using poly-T oligo-attached magnetic
588 beads. Purified mRNA was fragmented at 94 °C for 15 min by using divalent cations
589 under elevated temperature in NEBNext first strand synthesis reaction buffer (5X). First
590 strand cDNA was synthesized using random primer and ProtoScript II reverse
591 transcriptase in a preheated thermal cycler as follows: 10 min at 25 °C; 15 min at 42 °C;
592 15 min at 70 °C. Immediately finished, second strand synthesis reaction was performed
593 by using second strand synthesis reaction buffer (10X) and enzyme mix at 16 °C for 1
594 hr. The library fragments were purified with QiaQuick PCR kits and elution with EB
595 buffer, then terminal repair, A-tailing and adapter added were implemented. The
596 products were retrieved, and PCR was performed for library enrichment. The libraries
597 were sequenced on an Illumina platform.
598
599 ## 10X single-cell mRNA library preparation and sequencing
600 Single-cell suspensions of control and CX-5461 treated mES cells were resuspended in
601 DPBS-0.04% BSA at 1x106 cells/mL. Then scRNA-seq libraries were generated from the
602 10X Single Cell 3' Solution Reagents V2 according to the manufacturer's protocol (10x
603 Genomics). After the GEM-RT incubation, barcoded-cDNA was purified with
604 DynaBeads cleanup mix, followed by 10-cycles of PCR amplification (98°C for 3 min;
605 [98°C for 15 s, 67°C for 20 s, 72°C for 1 min] x 10; 72°C for 1 min). The total cDNA of
606 single-cell transcriptomes was fragmented, double-size selected with SPRI beads
607 (Beckman), followed by 12 cycles sample index PCR amplification (98°C for 45 s; [98°C
608 for 20 s, 54°C for 30 s, 72°C for 1 min] x 10; 72°C for 1 min), then another double-size
609 selection with SPRI beads was performed before sequencing. Libraries were
610 sequenced on the Illumina Hiseq X10 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, 1100 rpm, and wash twice in ice-cord PBS. Cell pellet was resuspended with lysis buffer containing 1× Protease Inhibitor Cocktail and incubated on ice for 10 min, then vortexed vigorously for 10 seconds and centrifuged at 3000 rpm for 5 minutes. The pellet was re-suspended in ChIP lysis buffer and incubated on ice for 10 minutes 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 seconds on, 30 seconds 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 (abcam, ab8898), H3K27me3 Rabbit mAb (CST, 9733), H3K4me3 Rabbit mAb (CST, 9751), H3K27ac Rabbit mAb (CST, 8173), Nucleolin (D4C7O) Rabbit (CST, 14574), TRIM28 Mouse monoclonal (20C1) (abcam, 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 hr 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.

In situ Hi-C library preparation and sequencing

10^6 cells were cross-linked for 10 min with 1% final concentration fresh formaldehyde

and quenched with 0.2M final concentration glycine for 5 min. The cross-linked cells were subsequently lysed in lysis buffer (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 for 15 min. The DNA in the nuclei was digested by adding 30 µl 10x NEB buffer 2.1 (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 100 µg/ml BSA, pH 7.9) and 150U of MboI, 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, 1µl of 10 mM dATP, 1 µl of 10 mM dGTP, 2 µl of 5mM 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 10x blunt-end ligation buffer (300 mM Tris-HCl, 100 mM MgCl2, 100 mM DTT, 1 mM ATP, pH 7.8), 100µl 10% Triton X-100 and 20 U T4 DNA ligase were added to start proximity ligation. The ligation reaction was placed at 16 °C for 4 h. After ligation, the cross-linking was reversed by 200 µg/mL proteinase K (Thermo) at 65°C overnight. DNA purification was achieved through QIAamp DNA Mini Kit (Qiagen) according to manufacturer's instructions. Purified DNA was sheared to a length of ~400 bp. Point ligation junctions were pulled down by Dynabeads® MyOne™ Streptavidin C1 (Thermofisher) according to manufacturer's instructions. The Hi-C library for Illumina sequencing was prepped by NEBNext® Ultra™ II DNA library Prep Kit for Illumina (NEB) according to manufacturer's instructions. The final library was sequenced on the Illumina HiSeq X Ten platform (San Diego, CA, United States) with 150PEmode. Two replicates were generated for one group material.

Bulk RNA-seq data analysis

All bulk RNA-seq reads were trimmed using Trimmomatic software (Version 0.36) with the following settings "ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36"[75] and were further quality-filtered using FASTX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) fastq_quality_trimmer command

with the minimum quality score 20 and minimum percent of 80% bases that has a quality score larger than this cutoff value. The high-quality reads were mapped to the mm10 genome by HISAT2, a fast and sensitive spliced alignment program for mapping RNA-seq reads, with -dta parameter[76]. PCR duplicate reads were removed using Picard tools (https://broadinstitute.github.io/picard/). 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[77] (Version v1.3.4d) 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-sample 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 using the reads count matrix produced from a python script “prepDE.py” provided in StringTie website (http://ccb.jhu.edu/software/stringtie/). We selected the genes with stage-specific scores larger than 0.2 to perform K-mean clustering analysis using pheatmap R package. The stage-specific scores of genes expressed during mouse early embryo development were obtained by entropy-based measure [78].

Unsupervised hierarchical clustering was carried out to compare the transcriptomes of mES cells from our study and other reports of 2C-like cell (2C::*tdTomato*+ and 2C::*tdTomato*-(GEO accession GSE33923); Zscan4_Em+ and Zscan4_Em- (GEO accession GSE51682); Kap1_KO and Kap1_WT (GEO accession GSE74278); CAF1_KO and CAF1_WT (GEO accession GSE85632), Dux_GFPpos and Dux_GFPneg (GEO accession GSE85632); LINE1 ASO and RC ASO (GEO accession GSE100939); Dppa4_GFPpos and Dppa4_GFPneg (GEO accession GSE120953), NELFA_GFPpos and NELFA_GFPneg (GEO accession GSE113671); Lin28a_KO and Lin28a_WT (GEO

accession GSE164420). Additionally, pre-implantation mouse embryos of different developmental stages (GSE66582) were included for comparison. TPM were obtained for each sample using the StringTie described above. Only genes that were expressed TPM ≥ 1 were included for analysis. A log2 transformation was applied after adding one pseudo-count (that is, log2[TPM+1]). The ComBat function from the sva package (https://bioconductor. org/packages/release/bioc/html/sva.html) was applied on log2 expression values to correct for batch effects caused by different experiments and sequencing platforms.

Single-cell RNA-seq data analysis

For single cell RNA-seq, we used 10x Genomics system following the manufacturer's protocol. We followed the previously published pipeline[79] to produce digital gene expression matrices of the droplet microfluidics-based single-cell RNA-seq sequencing data derived from control and CX-5461 treated mouse ES cells. Single-cell gene expression matrix was further analyzed with Seurat (https://satijalab.org/seurat/) (v2.3.4). We excluded the genes with expressed cell number smaller than 3 and the cells with nUMIs smaller than 500 or the expression percentages of mitochondrial genes larger than 0.2 and used 16 Principle Components (PCs) for tSNE analysis. Especially, we modified the published pipeline by substituting the aligner of Bowtie with HISAT2 for calculating the expression of repeat sequences. Considering the multi-mapping of reads derived from repeat sequences, we used all mapped reads for further analysis.

ChIP-seq and ATAC-seq data analysis

To tailor and filter ATAC-seq and ChIP-seq reads, we used the same procedure as RNA-seq reads processing. To avoid the potential effects of inconsistent sequencing depths on subsequent data analyses, we randomly sampled equal numbers read pairs from each experimental sample. For each sample, the ATAC-seq and ChIP-seq reads were first aligned to mm10 genomes using Bowtie2 (version 2.3.4.1)[80]. The ATAC-seq reads

were aligned with the parameters: -t -q -N 1 -L 25 -X 2000 no-mixed no-discordant. The ChIP-seq reads were aligned to mm10 with the options: -t -q -N 1 -L 25. The ATAC-seq reads were aligned with the parameters: -t -q -N 1 -L 25 -X 2000 no-mixed no-discordant. The ChIP-seq reads were aligned to mm10 with the options: -t -q -N 1 -L 25. For meta-analysis of genome regions, all unmapped reads, multiple mapped reads, and PCR duplicates were removed. For demonstrating the sequencing signal around Dux locus in UCSC genome browser, we maintained all multiple mapped reads for visualization. The bamCoverage and bamCompare commands contained in deepTools [81] (version 2.5.3) were adopted for downstream analysis. Using BamCoverage command with the parameters: -normalizeUsing BPM -of bigwig -binSize 100, we normalized the raw reads signal to Bins Per Million mapped reads (BPM) signal and converted the alignment bam files to bigwig signal files. The bigwig files were imported into UCSC genome browser for visualization. To minimize the effect of chromatin structure and sequencing bias in our ChIP-seq data, we corrected ChIP-seq signal using $\log_2$ ratio transformation between H3K9me3 signal and input signal by BamCompare command. We only considered the $\log_2$ ratio larger than 0 as effective ChIP-seq signals. The “computeMatrix” and “plotProfile” commands of deepTools were used to produce the reads density distribution plot of ATAC-seq and ChIP-seq signal in the given genomic region. For meta-analysis of sequencing signals in L1 regions, we only used their subsets which have overlaps with Inactive Hub regions. Homer (v4.11)[82] was used for motif discovery and enrichment analysis. For motifs across gene promoters, the search space was defined as a 4 kilobase (kb) window centered at the transcription start site (findMotifs.pl geneInput.txt mouse out/ -start -2000 -end 2000 -len 8,12 -p 4).

Hi-C data analysis

The paired-end reads of fastq files were aligned, processed, and iteratively corrected using HiC-Pro (version 2.11.1) as previously described[83]. Briefly, short sequencing reads were first independently mapped to mouse mm10 reference genome using the

bowtie2 aligner with end-to-end algorithm 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. Pairs of aligned reads were then assigned to MboI 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 variable 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 40kb, 150kb and 500 kb resolutions by dividing the genome into bins of equal size. The binned interaction matrices were then corrected by Knight–Ruiz matrix balancing method using hicCorrectMatrix command with the parameter --correctionMethod KR in HiCExplorer (https://hicexplorer.readthedocs.io/en/latest/) (v3.3)[85]. The Observed/Expected (O/E) Hi-C contact matrix was produced by HiCExplorer hicTransform command with --method obs_exp_norm option. Pearson Correlation Coefficients Hi-C matrix was obtained by HiCExplorer hicTransform command with --method pearson option.

A and B compartments were identified using the first eigenvector (PC1) from principal component analysis on correlation Hi-C matrix as described previously[85] with minor modifications. We used HOMER software with parameters '-res 500000 -window 1000000' to obtain the PC1 value based on Pearson Correlation Coefficients (PCC) Hi-C matrix. Sometimes the entry signs of PC1 need to be inverted to ensure that we are assigning the correct signs to individual regions. As GC content is well correlated with A and B compartments[86], 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 5 bins: (min to 20th percentile), (20th percentile to 40th percentile),

etc. For each pair of bins (25 pairs total), the averaged "O/E" values were then calculated for loci belonging to each pair of bins. The compartment strength was then calculated as $(AA+BB)/AB^2$ as done previously[87]. For the error bar in evaluating the compartment strength, we obtained 100 5×5 compartment enrichment matrices by bootstrapping. For each pixel of the 5x5 compartment enrichment map, we took all the O/E values that contributed to this pixel and took a random sample with replacement of the same size that the contributing values. We then proceeded with downstream for each of the 100 reshuffled maps.

TADs and loops annotated in CH12.LX were obtained from[59] and lifted over to the mm10 genome version using the UCSC genome browser liftOver tool. Aggregated plots of TAD enrichment map were obtained by averaging O/E values over annotated TAD positions at 40kb resolution as previously described[87]. 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-region pixels 30 to 60 correspond to the TAD body. TAD strength for boxplots was quantified as the ratio of two numbers. The first number is the within-TAD intensity: the sum of the central square of the enrichment map, rows 30 to 59 and columns 30 to 59. The second number is the between-TAD intensity, ½ of the sums of the regions [0:30, 30:60] and [30:60, 60:90]. Aggregated plots of loop enrichment map were obtained by averaging O/E values around loop anchors of 310kb window at 10kb resolution. Loop strength was calculated as previous described[87].

Hi-C "de novo boundary" aggregate plots at MERVLs are centered on 5' to 3'-oriented MERVL and show a window of 2mb around the MERVL element at 40kb resolution. For illustrating the change of insulation around MERVL genes, the $\log_2$ transformed Hi-C O/E matrix was further scaled by *z*-score normalization across each row. We calculated the insulation score as originally defined[60] with minor modifications. Briefly, for each region i in the genome, we calculated the average number of O/E interactions in 40kb Hi-C matrix in a quadratic window with the lower left corner at (i-1, i+1), and the top

right corner at (i-5, i+5), where 5 is the window size in bins. We normalized insulation scores by dividing each region's score by the average scores of the nearest 50 regions, and log_2-transforming the resulting vector, thus accounting for local biases in insulation score. Visualization of Hi-C matrix was carried out by Juicer tool (https://github.com/aidenlab/juicer)[88] and R software (https://www.r-project.org/). 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/Bioconductor software utilizing custom R scripts**.** The other statistical analyses were performed with GraphPad Prism software. 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 (s.e.m). All statistics are * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and were calculated by Wilcox signed rank test (for paired samples), Mann-Whitney U test (for independent samples), two-way ANOVA as described in the figure legends.

Code availability

All the analysis in this study was made based on custom perl and R codes and can be available upon reasonable request.

Data availability

All the 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 Gene Expression Omnibus (GEO) database under the accession number of GSE166041. Previously published RNA-Seq data that were re-analyzed here are available under accession codes GSE33923, GSE51682, GSE74278, GSE85632, GSE100939, GSE120953, GSE113671, GSE97778 and GSE66582. Published ChIP-Seq data for DUX are available

under accession code GSE85632. Published ChIP-Seq data for p53 are available under accession code GSE26360. Published ATAC-seq data are available under accession code (GSE66390 and GSE85632). Published Hi-C data of mouse pre-implantation embryos are available under accession code GSE82185. Published Hi-C data of lymphoblastoid cells are available under accession code GSE63525. Supplementary Table 5 provided a summary for all analyzed NGS datasets used in this study. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Acknowledgements

We thank Dr. Xiong Ji for providing the Pol I degraded mES cell lines. We thank Dr. Pengxu Qian for providing the snoRNA KO cell lines. We thank Dr. Todd Macfarlan for kind discussion. J.Z. is supported by the National Key Research and Development Program of China (No. 2018YFA0107100, No. 2018YFA0107103, No. 2018YFC1005002), the National Natural Science Foundation projects of China (No. 31871453, No. 91857116) and the Zhejiang Natural Science Foundation projects of China (No. LR19C120001). H.Y. is supported by the Zhejiang Natural Science Foundation projects of China (No. LQ21C120002).

Author contributions

J.Z. and H.Y. conceived and designed the study and experiments. H.Y. and J.Z. wrote the manuscript with contributions from all authors. H.Y. designed and performed all computational analysis. Z.S., T.T., H.P., A.L., Y.Z., L.C. and L.Z. performed the molecular experiments in mES cells. J.Z., L.Z. and J.C. performed the experiments in mouse embryos. H.Y. and Y.X. assisted with the experiment sample preparation. M.C., S.G. and G.D provided computational support and gave critical suggestions about the study design and paper writing. All authors analyzed the results and approved the final version of the manuscript.

Additional information

Competing interests: The authors declare no competing interests.

Figure Legends

Fig.1: Inhibition of rRNA biogenesis activated 2C-like transcriptional program and induced an expanded 2C-like cell population in mES cells. a) Volcano plots of RNA-sequencing data comparing gene expression of control and cellular stress inducer-treated mES cells (GEO accession GSE166041). **b)** UCSC Genome Browser viewing of RNA-sequencing results at the *Dux* locus (GEO accession GSE166041). **c)** MA plots of RNA-sequencing data comparing repeat sequence expression of control and cellular stress inducer-treated mES cells (GEO accession GSE166041). **d)** Violin plots demonstrating the expressional changes of stage-specific gene clusters of mouse pre-implantation embryos under different types of cellular stress treatment; **: $p<0.01$, ***: $p<0.001$, Mann-Whitney U test (GEO accession GSE166041). **e)** Hierarchical clustering of transcriptomes from our study, published 2C-like cell model studies and pre-implantation mouse embryos; Control, CX-5461, Rotenone and Rapamycin (GEO accession GSE166041), 2C::tdTomato+ and 2C::tdTomato- (GEO accession GSE33923); Zscan4_Em+ and Zscan4_Em- (GEO accession GSE51682); Kap1_KO and Kap1_WT (GEO accession GSE74278); CAF1_WT and CAF1_KO (GEO accession GSE85632), Dux_GFPpos and Dux_GFPneg (GEO accession GSE85632); LINE1 ASO and RC ASO (GEO accession GSE100939); Dppa4_GFPpos and Dppa4_GFPneg (GEO accession GSE120953), NELFA_GFPpos and NELFA_GFPneg (GEO accession GSE113671); Lin28a_KO and Lin28a_WT (GEO accession GSE164420); MII-Oocyte, Zygote, Early 2-cell, 2-cell, 4-cell, 8-cell, ICM and mES cells (GEO accession GSE66390). **f)** tSNE feature plots demonstrating the averaged expression levels of *MERVL* genes in 2981 control mES cells and 3219 CX-5461-treated mES cells. **g)** tSNE feature plots demonstrating the averaged expression levels of ribosome genes in 2981 control mES cells and 3219 CX-5461-treated mES cells. **h)** tSNE feature plots demonstrating the averaged expression levels of ribosome biogenesis genes in 2981 control mES cells and 3219 CX-5461-

901 treated mES cells (GEO accession GSE166041). **i)** Scatter plots demonstrating negative
902 correlation of expression level between *MERVL/MERVL-int/MT2_Mm* and ribosome
903 genes; Each dot represents a single cell with detectable ERV expression; *r* denotes
904 correlation coefficient; *p*-value was obtained by cor.test function in R software (GEO
905 accession GSE166041). **j)** FACS analysis on 2C::tdTomato+ mES cells upon different
906 treatment doses of CX-5461, showing the change of percentage of 2C-like cells. **k)** FACS
907 analysis on 2C::tdTomato+ mES cells after 12 hour treatment and 24 hour withdrawal
908 of CX-5461, showing the change of percentage of 2C-like cells.
909 **Fig.2: Deficiency of rRNA biogenesis disrupted normal nucleolar LLPS and epigenetic**
910 **state of PNH region. a)** CX-5461 treatment causes abnormal nucleolar structure with
911 electron microscopy. **b)** Immunofluorescence staining of NCL in control mES cells and
912 CX-5461-treated mES cells. **c)** Immunofluorescence staining of NPM1 in control mES
913 cells and CX-5461-treated mES cells. **d)** Immunofluorescence staining of FBL in control
914 mES cells and CX-5461 treated mES cells. **e)** Immunofluorescence staining of RPA194
915 in control and CX-5461 treated mES cells. **f)** FRAP analysis showing CX-5461 treatment
916 causes accelerated recovery after photobleaching of NCL. Shown images are
917 representative of 4 times of experiments. **g)** FRAP analysis showing CX-5461 treatment
918 causes accelerated recovery after photobleaching of NPM1. Shown images are
919 representative of 4 times of experiments. **h)** FRAP analysis showing CX-5461 treatment
920 causes accelerated recovery after photobleaching of FBL. Shown images are
921 representative of 4 times of experiments. **i)** Heatmap plots demonstrate the levels of
922 H3K9me3 and H3K27me3 on within 1mb region around start and end sites of Inactive
923 Hub. The regions of different lengths of Inactive Hub fragments were fitted to 1mb
924 (GEO accession GSE166041). **j)** Heatmap plots demonstrate the levels of H3K9me3 and
925 H3K27me3 within 1mb region around start and end sites of NAD. The regions of
926 different lengths of NAD fragments were fitted to 1mb (GEO accession GSE166041). **k)**
927 Heatmap plots demonstrate the levels of H3K9me3 and H3K27me3 within 1kb region
928 around start and end sites of L1 (GEO accession GSE166041). The regions of different
929 lengths of L1 sequences were fitted to 1kb. **l)** Heatmap plots demonstrate the level of

H3K4me3 and H3K27ac within 1mb region around start and end sites of Inactive Hub (GEO accession GSE166041). The regions of different lengths of Inactive Hub fragments were fitted to 1mb. **m)** Heatmap plots demonstrate the levels of H3K4me3 and H3K27ac within 1mb region around start and end sites of NAD. The regions of different lengths of NAD fragments were fitted to 1mb (GEO accession GSE166041). **n)** Heatmap plots demonstrate the level of H3K4me3 and H3K27ac on within 1kb region around start and end sites of L1 (GEO accession GSE166041). The regions of different lengths of L1 sequences were fitted to 1kb. **o)** Heatmap plots demonstrate the level of NCL and TRIM28 within 1mb region around start and end sites of Inactive Hub (GEO accession GSE166041). The regions of different lengths of Inactive Hub fragments were fitted to 1mb. **p)** Heatmap plots demonstrate the levels of NCL and TRIM28 within 1mb region around start and end sites of NAD. The regions of different lengths of NAD fragments were fitted to 1mb (GEO accession GSE166041). **q)** Heatmap plots demonstrate the level of NCL and TRIM28 within 1kb region around start and end sites of L1 (GEO accession GSE166041). The regions of different lengths of L1 sequences were fitted to 1kb.

Fig.3: 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 5kb region around transcription start sites of 5258 CX-5461 induced genes using published p53 ChIP-seq data and Dux ChIP-seq data (GEO accession GSE26360 for p53 and GEO accession GSE85632 for Dux). **d)** Line plots demonstrate the meta-analysis results of p53 and Dux protein on within 5kb 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 GSE26360 for p53 and GEO accession 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 10173 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 5kb (GEO accession GSE26360 for p53 and GEO accession GSE85632 for Dux). **f)** Heatmap plots demonstrate the binding of p53 and Dux proteins on within 5kb 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 5kb (GEO accession GSE26360 for p53 and GEO accession GSE85632 for Dux). **g)** qRT-PCR showing the expression of *Dux* or 2C-related genes in *Dux* silenced mES cells; **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3; error bar: standard error of the mean. **h)** ChIP-PCR showing H3K9me3 levels of Dux or 2C-related genes in Dux silenced mES cells; *: p<0.05, **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3, error bar: standard error of mean. **i)** ChIP-PCR showing H3K27me3 levels of Dux or 2C-related genes in Dux silenced mES cells; *: p<0.05, **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3, error bar: standard error of mean. **j)** ChIP-PCR showing DUX protein binding levels on 2C-related genes in Dux silenced mES cells; *: p<0.05, **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3, error bar: standard error of mean. **k)** Line plots demonstrate the meta-analysis results of chromatin accessibility in Dux silenced mES cells within 5kb region around transcription start sites or transcription start and end sites of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression and 10173 CX-5461 induced ERV genes using published ATAC-seq data. The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE166041)**.** p53-ctrl: untreated mES cells, p53-adr: mES cells treated with adriamycin, a DNA damage agent widely used to activate p53, Dux (12h): mES induced with doxycycline for 12 hours, Dux (18h): mES induced with doxycycline for 18 hours.

Fig.4: rRNA biogenesis defect drove 3D chromatin structure reorganization of PNH and *MERVL* regions towards the 2C-like state. a) Hi-C contact maps of Inactive Hub and 1.5 Mb genomic regions around *Dux* at 150kb resolution (GEO accession GSE166041). **b)** Hi-C contact maps of NAD and 1.5 Mb genomic regions around *Dux* at

150kb resolution (GEO accession GSE166041). **c)** Hi-C contact maps of L1 and 1.5 Mb genomic regions around *Dux* at 150kb 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 GSE166041). **d)** Hi-C pearson correlation coefficient (PCC) heat maps of Inactive Hub and 1.5Mb genomic regions around *Dux* at 150kb resolution (GEO accession GSE166041). **e)** Hi-C pearson correlation heat maps of NAD and 1.5Mb genomic regions around *Dux* at 150kb resolution (GEO accession GSE166041). **f)** Hi-C pearson correlation heat maps of L1 and 1.5Mb genomic regions around *Dux* at 150kb 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 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 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 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 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 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 GSE166041). **l)** Scatter plot demonstrates the PCC difference between L1 and different types of genes in control and CX-5461 treated mES cells. The difference of PCC is defined as the average (PCC) of Inactive Hub regions and different types of genes in CX-5461 treated mES cells minus the average PCC of Inactive Hub regions and different types of genes in wild type mES cells. PCC: Pearson Correlation Coefficient, *MERVL-int*: Up-regulated *MERVL-int* genes, *MT2_Mm*: Up-regulated *MT2_Mm* genes, UG: Up-regulated genes, DG: Down-regulated genes (GEO accession 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. ***: p<0.001, two-way ANOVA, error bar: standard error of the mean, n denotes the number of observed mES cells (GEO accession GSE166041). **n)** Aggregate Observed(O)/Expected(E) Hi-C matrices centered on CX-5461 induced *MERVL* genes in control ,CX-5461 treated mES cells (GEO accession GSE166041) and lymphoblastoid cells (GEO accession GSE63525) and mouse embryos throughout mouse pre-implantation embryonic development (GEO accession GSE82185). **o)** Representative 40kb 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 GSE166041 and GSE82185).

Fig.5: Genetic interferences of rRNA biogenesis recapitulate CX-5461-induced 2C-like molecular phenotypes. a) DNA FISH analysis with a *Dux* locus probe and Inactive Hub locus probe, and co-immunostained with NCL protein in control and Pol I degradation mES cell lines. **b)** The percentage of Nucleolus-localized (overlapped with NCL) and Nucleoplasm-localized (nonoverlapped with NCL) of FISH signals in control and Pol I degradation mES cell lines; **: p<0.01, ***: p<0.001, two-way ANOVA, error bar: standard error of the mean, n denotes the number of observed mES cells. **c)** 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. **d)** The percentage of Nucleolus-localized (overlapped with NCL) and Nucleoplasm-localized (nonoverlapped with NCL) of FISH signals in control and snoRNA knockout mES cell lines; *: p<0.05, ***: p<0.001, two-way ANOVA, error bar: standard error of the mean, n denotes the number of observed mES cells. **e)** FRAP analysis showing Pol I degradation causes accelerated recovery after photobleaching of NCL, the replicates of experiment n = 4. **f)** FRAP analysis showing Pol I degradation causes accelerated recovery after

photobleaching of NPM1, the replicates of experiment n = 4. **g)** FRAP analysis showing snoRNA knockout causes accelerated recovery after photobleaching of NCL, the replicates of experiment n = 4. **h)** FRAP analysis showing snoRNA knockout causes accelerated recovery after photobleaching of NPM1, the replicates of experiment n = 4. **i)** qRT-PCR quantification of 2C marker gene expression in control mES cells and Pol I degraded mES cell lines; **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3; error bar: standard error of the mean. **j)** qRT-PCR quantification of 2C marker gene expression in control mES cells and snoRNA knockout mES cell lines; **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3; error bar: standard error of the mean. **k)** The percentage of 2C::tdTomato positive cells was quantified using FACS analysis in control mES cells and Pol I degraded mES cells; Data are means ± SD, SD: Standard Deviation, ***: p<0.001, two-way ANOVA, the replicates of experiment n=5. **l)** The percentage of 2C::tdTomato positive cells was quantified using FACS analysis in control mES cells and snoRNA knockout mES cells; Data are means ± SD, SD: Standard Deviation, **: p<0.01, two-way ANOVA, the replicates of experiment n=5.

Fig.6: rRNA biogenesis is critically required at the 2-cell-to-4-cell stage transition during pre-implantation embryo development. a) Expression of pre-rRNA and 28S rRNA across different embryo developmental stages. **b)** Expression of different subunit genes of RNA polymerase I across different embryo developmental stages. **c)** Different schemes of treatment with CX-5461. The 24hrs 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. *: p<0.05, **: p<0.01, two-way ANOVA, the replicates of experiment n=3. **f)** qRT-PCR showing 2C marker gene expression level in blastocysts, after CX-5461 treatment of morula embryos followed by *in vitro* culture of the treated embryos. ***: p<0.001, two-way ANOVA, the replicates of experiment n=3. **g)** Violin plots demonstrating the expression

level changes of stage-specific gene clusters of mouse pre-implantation embryos (as defined in Extended Data Fig S1) in CX-5461-treated and control blastocyst embryos, ***: p<0.001, Mann-Whitney U test (GEO accession GSE166041). **h)** Violin plots show expression levels of major ERV gene classes in control blastocyst embryos and CX-5461 treated blastocyst embryos; n denotes the number of sub-classes of ERV genes; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **i)** Violin plots show expression levels of ERV gene sub-classes of *MERVL-int* and *MT2_Mm* in control blastocyst embryos and CX-5461 treated blastocyst embryos; seq_n denotes the number of annotated *MERVL-int* and *MT2_Mm* sequences in the mouse mm10 reference genome; *: ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **j)** 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 *GSAT_MM* sequences in the mouse mm10 reference genome; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041).

Fig.7: A mechanistic model for the role of rRNA biogenesis in regulating the 2C-like program and the homeostasis between 2C-like cells and mES cells. In the unperturbed mES cells, nucleolar integrity mediated by rRNA biogenesis maintains the normal the liquid-liquid phase separation (LLPS) of nucleolus and the formation of peri-nucleolar heterochromatin (PNH) containing *Dux*, and this normal nucleolar LLPS facilitated NCL/TRIM28 complex occupancy on the *Dux* locus to repress *Dux* expression. In contrast, in the rRNA biogenesis-inhibited mES cells, the natural liquid-like phase of nucleolus is disrupted, causing dissociation of the NCL/TRIM28 complex from the PNH and changes of epigenetic state and 3D structure of the PNH, which eventually leads to *Dux* to be released from the PNH, activation of 2C-like program and transition of mES cells to 2C-like cells.

Extended Data Fig.S1: Inhibition of rRNA biogenesis activated 2C-like transcriptional program and induced an expanded 2C-like cell population in mES cells, related to Fig.1. a) UCSC Genome Browser viewing of RNA-sequencing results in *Zscan4d* gene locus (GEO accession GSE166041). **b)** UCSC Genome Browser viewing of RNA-sequencing results in *Gm12794* gene locus (GEO accession GSE166041). **c)** UCSC

Genome Browser viewing of RNA-sequencing results in a representative ERV (*MERVL-int* and *MT2_Mm*) gene locus (GEO accession GSE166041). **d)** UCSC Genome Browser viewing of RNA-sequencing results in a representative *GSAT_MM* gene loci (GEO accession GSE166041). **e)** Boxplots show the expression levels of major ERV gene classes under different types of cellular stress treatment. n denotes the number of sub-classes of ERV genes, ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **f)** A heatmap plot demonstrating four developmental stage-specific gene clusters derived from RNA-seq data of pre-implantation mouse embryos (GEO accession GSE97778). **g)** qRT-PCR quantification of rRNA expression in control mES cells and CX-5461 treated mES cells, **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3. **h)** Boxplots demonstrating the expression levels of 2C marker genes of Dux, Zscan4d and Gm12794 in control mES cells and CX-5461 treated mES cells; Each point denoted a single cell; The number of parentheses denotes the percentage of cells expressing these genes; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **i)** Boxplots demonstrating the averaged expression levels of *MERVL* genes in control mES cells and CX-5461 treated mES cells with single cell RNA-sequencing analysis; Each point denoted a single cell; The number of parentheses denotes the percentage of cells expressing *MERVL* genes; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **j)** The expression levels of pluripotency genes in control mES cells and CX-5461 treated mES cells; Each point denoted a single cell; The number of parentheses denotes the percentage of cells expressing these genes; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **k)** Scatter plots demonstrating negative correlation of expression level between *MERVL/MERVL-int/MT2_Mm* and ribosome biogenesis genes; Each dot represents a single cell with detectable ERV expression; *r* denotes correlation coefficient; *p*-value was obtained by cor.test function in R software (GEO accession GSE166041). **l)** The percentage of 2C::tdTomato positive cells was quantified using FACS analysis in control mES cells and CX-5461-treated mES cells; Data are means ± SD; SD: Standard Deviation; **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=5. **m)** FACS analysis on

Annexin-V FITC, marker for early cell apoptosis, and DAPI, marker for late cell apoptosis upon different treatment doses of CX-5461; **n)** FACS analysis on Annexin-V FITC and 2C::tdTomato positive mES cells upon different treatment doses of CX-5461. **o)** FACS analysis on DAPI and 2C::tdTomato positive mES cells upon different treatment doses of CX-5461. In **e)**, **h)**, **i)** and **j),** the centre of the box plots represents the median value and the lower and upper lines represent the 5% and 95% quantile, respectively.

Extended Data Fig.S2: Deficiency of rRNA biogenesis disrupted normal nucleolar LLPS and epigenetic state of PNH region, related to Fig.2. a) Boxplots demonstrate the averaged H3K9me3 and H3K27me3 levels of 101 Inactive Hub fragments, 578 NAD fragments and 34888 L1 sequences; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **b)** UCSC Genome Browser viewing of H3K9me3 and H3K27me3 ChIP-seq signals in control and CX-5461 treated mES cells around a representative PNH fragment at chr15:4,699,343-5,778,242 (GEO accession GSE166041). **c)** Boxplots demonstrate the averaged H3K4me3 and H3K27ac levels of 101 Inactive Hub fragments, 578 NAD fragments and 34888 L1 sequences; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **d)** UCSC Genome Browser viewing of H3K4me3 and H3K27ac ChIP-seq signals in control mES cells and CX-5461 treated mES cells around a representative PNH fragment at chr10:3,632,504-3,926,058 (GEO accession GSE166041). **e)** Heatmap plots demonstrate the levels of chromatin accessibility on within 1mb region around start and end sites of Inactive Hub and NAD, and within 1kb region around start and end sites of L1 (GEO accession GSE166041). The regions of different lengths of Inactive Hub and NAD fragments were fitted to 1mb. The regions of different lengths of L1 sequences were fitted to 1kb. SS: start site of a chromatin fragment of PNH; ES: end site of a chromatin fragment of PNH. The PNH fragment was defined as the L1 contained regions overlapped with Inactive Hub and NAD. **f)** Boxplots demonstrate the averaged chromatin accessibility of 101 Inactive Hub fragments, 578 NAD fragments and 34888 L1 sequences; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **g)** UCSC Genome Browser viewing of ATAC-seq signals in control and CX-5461 treated mES cells around a representative PNH fragment at

chr19:18,666,449-18,747,509 (GEO accession GSE166041). **h)** Boxplots demonstrate the averaged binding signals of NCL and TRIM28 on 101 Inactive Hub fragments, 578 NAD fragments and 34888 L1 sequences; **: p<0.01, ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **i)** UCSC Genome Browser viewing of NCL and TRIM28 ChIP-seq signals in control and CX-5461 treated mES cells around a representative PNH fragment at chr18:7,718,063-8,074,255 (GEO accession GSE166041). In **a)**, **c)**, **f)** and **h),** the centre of the box plots represents the median value and the lower and upper lines represent the 5% and 95% quantile, respectively.

Extended Data Fig.S3: 2C/ERV genes were activated through Dux, related to Fig.3. **a)** Line plots demonstrate the meta-analysis results of chromatin accessibility in Dux-GFP positive mES cells and Dux-GFP negative mES cells within 5kb region around transcription start sites or transcription start and end sites of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression, 4637 specifically induced genes by CX-5461 and 10173 CX-5461 induced ERV genes using published ATAC-seq data. The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE85632). **b)** UCSC Genome Browser viewing of ChIP-seq signals around *Dux* locus (GEO accession GSE166041). **c)** Line plots demonstrate the meta-analysis results of chromatin accessibility of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression and 10173 CX-5461 induced ERV genes in control mES cells and CX-5461 treated mES cells. The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE166041). **d)** Line plots demonstrate the meta-analysis results of chromatin accessibility of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression and 10173 CX-5461 induced ERV genes in 2-cell embryo and ICM embryo. The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE66390). **e)** Line plots demonstrates the meta-analysis results of H3K9me3 and H3K27me3 levels of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression in control mES cells and CX-5461 treated mES cells (GEO accession GSE166041). **f)** Line plots demonstrate the meta-analysis results of

H3K9me3 and H3K27me3 levels of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression in 2-cell embryo and ICM embryo (GEO accession GSE66390). **g)** Line plots demonstrate the meta-analysis results of H3K9me3 and H3K27me3 levels of 10173 CX-5461 induced ERV genes in control mES cells and CX-5461 treated mES cells; The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE166041). **h)** Line plots demonstrate the meta-analysis results of H3K9me3 and H3K27me3 levels of 10173 CX-5461 induced ERV genes in 2-cell embryo and ICM embryo; The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE66390). **i)** Line plots demonstrate the meta-analysis results of H3K4me3 and H3K27ac levels of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression in control mES cells and CX-5461 treated mES cells (GEO accession GSE166041). **j)** Line plots demonstrate the meta-analysis results of H3K4me3 and H3K27ac levels of 10173 CX-5461 induced ERV genes in control mES cells and CX-5461 treated mES cells; The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE66390).

Extended Data Fig.S4: rRNA biogenesis defect drove 3D chromatin structure reorganization of PNH and MERVL regions towards the 2C-like state, related to Fig.4. **a)** Hi-C contact maps of Inactive Hub/NAD/L1 and 1.5 Mb genomic regions around *Dux* at 150kb resolution during mouse pre-implantation embryos development. The zoomed-in regions aim to demonstrate the change of Hi-C contacts between *Dux* and chromosome 10 during mouse pre-implantation embryos development (GEO accession GSE82185). **b)** Scatter plot demonstrates the $\log_2$(fold change) of Hi-C contacts between Inactive Hub/NAD/L1 and different types of genes in early 2-cell and ICM stage embryos (GEO accession GSE82185). **c)** Hi-C pearson correlation heat maps of Inactive Hub/NAD/L1 and 1.5 Mb genomic regions around *Dux* at 150kb resolution during mouse pre-implantation embryos development. The zoomed-in regions aim to demonstrate the change of Hi-C PCC between *Dux* and chromosome 10 during mouse pre-implantation embryos development (GEO accession GSE82185). **d)** Scatter plot demonstrates the PCC difference between Inactive Hub/NAD/L1 and different types of

genes in early 2-cell and ICM stage embryos (GEO accession GSE82185). **e)** A/B interaction profile showing contact enrichment between active and inactive compartments (GEO accession GSE166041 and GSE63525). **f)** Quantification of compartment strength; *: p<0.05, ***: p<0.001, Wilcox signed rank test. **g)** Pearson correlation heat maps for chromosome 16 at 500kb resolution to demonstrate A/B compartment (GEO accession GSE166041 and GSE63525). **h)** Hi-C contact maps for chromosome 16 at 500kb resolution for A/B compartment profile. **i)** Observed/Expected (O/E) aggregate plot of TADs (GEO accession GSE166041 and GSE63525). **j)** Quantification of TAD strength; *: p<0.05, ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041 and GSE63525). **k)** O/E aggregate plots of chromatin loops (GEO accession GSE166041 and GSE63525). **l)** Quantification of loop strength; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041 and GSE63525). **m)** Hi-C contact maps for chromosome 16:27.3-36.2mb region at 40kb resolution to demonstrate TAD and chromatin loop (GEO accession GSE166041 and GSE63525). **n)** Insulation score, expression (TPM), ATAC-seq (BPM), H3K9me3 (BPM), H3K27me3 (BPM), H3K4me3 (BPM) and H3K27ac (BPM) signals of control and CX-5461-treated mES cells centered on CX-5461-induced *MERVL* genes (GEO accession GSE166041). **o)** Representative 40kb Hi-C O/E interaction matrices of a *MERVL* loci (chr15:76,145,960-86,151,258) located at TAD boundaries are shown as heatmaps, along with 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 GSE166041).

Extended Data Fig.S5: Genetic interferences of rRNA biogenesis recapitulate CX-5461-induced 2C-like molecular phenotypes, related to Fig.5. a) Western Blotting experiment showing the Pol I protein degradation after 24h of Auxin treatment. **b)** PCR experiment showing that a 400bp band was observed in the snoRNA KO mES cells, but not in the wild-type (WT) mES cells. As a band of 400bp was designed especially in the snoRNA KO mES cells, this result indicates that the homologs of human SNORD113-114 gene cluster was successfully knocked-out.

c) Shown images are representative of 4 times of NCL FRAP experiments in control mES cells and Pol I degraded mES cells. **d)** Shown images are representative of 4 times of of NPM1 FRAP experiments in control mES cells and Pol I degraded mES cells. **e)** Shown images are representative of 4 times of NCL FRAP experiments in control mES cells and snoRNA knockout mES cells. **f)** Shown images are representative of 4 times of NPM1 FRAP experiments in control mES cells and snoRNA knockout mES cells. **g)** FACS analysis on 2C::tdTomato+ mES cells in Pol I degraded mES cell lines, showing the change of percentage of 2C-like cells**. h)** FACS analysis on 2C::tdTomato+ mES cells in snoRNA knockout mES cell lines, showing the change of percentage of 2C-like cells**.**

Extended Data Fig.S6: rRNA biogenesis is critically required at the 2-cell-to-4-cell stage transition during pre-implantation embryo development, related to Fig.6. a) Expression pattern of Ribosome biogenesis gene set across different early embryo developmental stages. n denotes the number of sub-classes of *MERVL* genes. **b)** Expression pattern of *MERVL-int* genes across different embryo developmental stages. seq_n denotes the number of annotated *MERVL-int* sequences in the mouse mm10 reference genome. **c)** Expression pattern of *MT2_Mm* genes across different embryo developmental stages. seq_n denotes the number of annotated *MT2_Mm* sequences in the mouse mm10 reference genome. **d)** Expression pattern of 2C marker genes, *Dux*, *Zscan4d* and *Gm12794*, across different embryo developmental stages. **e)** Representative images of mouse embryos produced from control and CX-5461 treatment during two different developmental stages; Dpc: Days post-coitum. This experiment was repeated three times independently with similar results. **f)** Immunofluorescence staining of NCL, NPM1 and FBL in control blastocyst embryos and CX-5461-treated blastocyst embryos.

References

1 Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57-63, doi:10.1038/nature11244 (2012).

2 Choi, Y. J. & Lin, C. P. Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. *Science* **355**, doi:10.1126/science.aag1927 (2017).

3 Zalzman, M. *et al.* Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* **464**, 858-863, doi:10.1038/nature08882 (2010).

4 Matsui, T. *et al.* Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927-931, doi:10.1038/nature08858 (2010).

5 Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237-240, doi:10.1038/nature08674 (2010).

6 Fu, X., Wu, X. & Djekidel, M. N. Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells. *Nat Cell Biol* **21**, 835-844, doi:10.1038/s41556-019-0343-0 (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, doi:10.1016/j.celrep.2017.07.070 (2017).

8 Guallar, D. *et al.* RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nat Genet* **50**, 443-451, doi:10.1038/s41588-018-0060-9 (2018).

9 Yang, B. X. *et al.* Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* **163**, 230-245, doi:10.1016/j.cell.2015.08.037 (2015).

10 De Iaco, A. *et al.* DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* **49**, 941-945, doi:10.1038/ng.3858 (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, doi:10.1038/ng.3844 (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, doi:10.1038/ng.3846 (2017).

13 Yang, F. *et al.* DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces a Totipotent 2C-like State. *Cell Stem Cell* **26**, 234-250.e237, doi:10.1016/j.stem.2020.01.004 (2020).

14 Percharde, M. *et al.* A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **174**, 391-405 e319, doi:10.1016/j.cell.2018.05.043 (2018).

15 Wu, K. *et al.* SETDB1-Mediated Cell Fate Transition between 2C-Like and Pluripotent States. *Cell Rep* **30**, 25-36.e26, doi:10.1016/j.celrep.2019.12.010 (2020).

16 Eckersley-Maslin, M. & Alda-Catalinas, C. Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Gene Dev* **33**, 194-208, doi:10.1101/gad.321174.118 (2019).

17 Hu, Z. & Tan, D. E. K. Maternal factor NELFA drives a 2C-like state in mouse

1318 embryonic stem cells. *Nat Cell Biol* **22**, 175-186, doi:10.1038/s41556-019-0453-8
1319 (2020).

1320 18 Ishiuchi, T. *et al.* Early embryonic-like cells are induced by downregulating
1321 replication-dependent chromatin assembly. *Nat Struct Mol Biol* **22**, 662-671,
1322 doi:10.1038/nsmb.3066 (2015).

1323 19 Zhang, Y., Li, T. & Preissl, S. Transcriptionally active HERV-H retrotransposons
1324 demarcate topologically associating domains in human pluripotent stem cells. *Nat*
1325 *Genet* **51**, 1380-1388, doi:10.1038/s41588-019-0479-7 (2019).

1326 20 Chen, X. *et al.* Key role for CTCF in establishing chromatin structure in human
1327 embryos. *Nature* **576**, 306-310, doi:10.1038/s41586-019-1812-0 (2019).

1328 21 Du, Z. *et al.* Allelic reprogramming of 3D chromatin architecture during early
1329 mammalian development. *Nature* **547**, 232-235, doi:10.1038/nature23263 (2017).

1330 22 Hug, C. B. & Vaquerizas, J. M. The Birth of the 3D Genome during Early
1331 Embryonic Development. *Trends Genet* **34**, 903-914, doi:10.1016/j.tig.2018.09.002
1332 (2018).

1333 23 Kruse, K. *et al.* Transposable elements drive reorganisation of 3D chromatin
1334 during early embryogenesis. *bioRxiv*, 523712, doi:10.1101/523712 (2019).

1335 24 Guetg, C. & Santoro, R. Formation of nuclear heterochromatin: the nucleolar
1336 point of view. *Epigenetics* **7**, 811-814, doi:10.4161/epi.21072 (2012).

1337 25 Németh, A. & Längst, G. Genome organization in and around the nucleolus.
1338 *Trends Genet* **27**, 149-156, doi:10.1016/j.tig.2011.01.002 (2011).

1339 26 Bywater, M. J. et al. Inhibition of RNA polymerase I as a therapeutic strategy to
1340 promote cancer-specific activation of p53. *Cancer Cell* **22**, 51-65,
1341 doi:10.1016/j.ccr.2012.05.019 (2012).

1342 27 Fujiwara, Y. et al. A Nucleolar Stress-Specific p53-miR-101 Molecular Circuit
1343 Functions as an Intrinsic Tumor-Suppressor Network. *EBioMedicine* **33**, 33-48,
1344 doi:10.1016/j.ebiom.2018.06.031 (2018).

1345 28 Wang, C., Liu, X. & Gao, Y. Reprogramming of H3K9me3-dependent
1346 heterochromatin during mammalian embryo development. *Nat Cell Biol* **20**, 620-631,
1347 doi:10.1038/s41556-018-0093-4 (2018).

1348 29 Wu, J. et al. The landscape of accessible chromatin in mammalian
1349 preimplantation embryos. *Nature* **534**, 652-657, doi:10.1038/nature18606 (2016).

1350 30 Akiyama, T. et al. Transient bursts of Zscan4 expression are accompanied by the
1351 rapid derepression of heterochromatin in mouse embryonic stem cells. DNA
1352 research : an international journal for rapid publication of reports on genes and
1353 genomes 22, 307-318, doi:10.1093/dnares/dsv013 (2015).

1354 31 Ecco, G. *et al.* Transposable Elements and Their KRAB-ZFP Controllers Regulate

Gene Expression in Adult Tissues. *Dev Cell* **36**, 611-623, doi:10.1016/j.devcel.2016.02.024 (2016).

32 Bastola, P., Neums, L., Schoenen, F. J. & Chien, J. VCP inhibitors induce endoplasmic reticulum stress, cause cell cycle arrest, trigger caspase-mediated cell death and synergistically kill ovarian cancer cells in combination with Salubrinal. *Mol Oncol* **10**, 1559-1574, doi:10.1016/j.molonc.2016.09.005 (2016).

33 Hung, S. S. et al. Repression of global protein synthesis by Eif1a-like genes that are expressed specifically in the two-cell embryos and the transient Zscan4-positive state of embryonic stem cells. *DNA Res* **20**, 391-402, doi:10.1093/dnares/dst018 (2013).

34 Eckersley-Maslin, M. A. et al. MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. *Cell Rep* **17**, 179-192, doi:10.1016/j.celrep.2016.08.087 (2016).

35 Yao, R. W., Xu, G., Wang, Y., Shan, L. & Chen, L. L. Nascent Pre-rRNA Sorting via Phase Separation Drives the Assembly of Dense Fibrillar Components in the Human Nucleolus. *Mol Cell* **76**, 767-783.e711 (2019).

36 Feric, M. et al. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686-1697, doi:10.1016/j.cell.2016.04.047 (2016).

37 Lafontaine, D. L. J., Riback, J. A., Bascetin, R. & Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat Rev Mol Cell Biol* **22**, 165-182, doi:10.1038/s41580-020-0272-6 (2020).

38 Frottin, F. & Schueder, F. The nucleolus functions as a phase-separated protein quality control compartment. *Science* **365**, 342-347, doi:10.1126/science.aaw9157 (2019).

39 Strom, A. R. et al. Phase separation drives heterochromatin domain formation. *Nature* **547**, 241-245, doi:10.1038/nature22989 (2017).

40 Sanulli, S. et al. HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature* **575**, 390-394, doi:10.1038/s41586-019-1669-2 (2019).

41 Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23, doi:10.1016/j.cell.2017.02.007 (2017).

42 Quinodoz, S. A. et al. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757.e724, doi:10.1016/j.cell.2018.05.024 (2018).

43 Lu, J. Y. et al. Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Rep* **30**, 3296-3311.e3295, doi:10.1016/j.celrep.2020.02.048 (2020).

44 Gupta, S. & Santoro, R. Regulation and Roles of the Nucleolus in Embryonic Stem

Cells: From Ribosome Biogenesis to Genome Organization. *Stem Cell Rep* **15**, 1206-1219, doi:10.1016/j.stemcr.2020.08.012 (2020).

45 Bizhanova, A., Yan, A., Yu, J., Zhu, L. J. & Kaufman, P. D. Distinct features of nucleolus-associated domains in mouse embryonic stem cells. *Chromosoma* **129**, 121-139, doi:10.1007/s00412-020-00734-9 (2020).

46 Németh, A. et al. Initial genomics of the human nucleolus. *PLoS genetics* **6**, e1000889, doi:10.1371/journal.pgen.1000889 (2010).

47 Guetg, C. et al. The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats. *The EMBO journal* **29**, 2135-2146, doi:10.1038/emboj.2010.17 (2010).

48 Lu, J. Y. et al. Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res* **31**, 613-630, doi:10.1038/s41422-020-00466-6 (2021).

49 Bersaglieri, C. et al. Genome-wide maps of nucleolus interactions reveal distinct layers of repressive chromatin domains. bioRxiv (2020).

50 Li, M. et al. Distinct regulatory mechanisms and functions for p53-activated and p53-repressed DNA damage response genes in embryonic stem cells. *Mol Cell* **46**, 30-42, doi:10.1016/j.molcel.2012.01.020 (2012).

51 Latonen, L. Phase-to-Phase With Nucleoli-Stress Responses, Protein Aggregation and Novel Roles of RNA. *Front Cell Neurosci* **13**, 151, doi:10.3389/fncel.2019.00151 (2019).

52 Boulon, S., Westman, B. J., Hutten, S., Boisvert, F. M. & Lamond, A. I. The nucleolus under stress. *Mol Cell* **40**, 216-227, doi:10.1016/j.molcel.2010.09.024 (2010).

53 Nicolas, E. et al. Involvement of human ribosomal proteins in nucleolar structure and p53-dependent nucleolar stress. *Nat Commun* **7**, 11390, doi:10.1038/ncomms11390 (2016).

54 Yang, K., Yang, J. & Yi, J. Nucleolar Stress: hallmarks, sensing mechanism and diseases. *Cell Stress* **2**, 125-140, doi:10.15698/cst2018.06.139 (2018).

55 Golomb, L., Volarevic, S. & Oren, M. p53 and ribosome biogenesis stress: the essentials. *FEBS Lett* **588**, 2571-2579, doi:10.1016/j.febslet.2014.04.014 (2014).

56 Zhang, Y. & Lu, H. Signaling to p53: ribosomal proteins find their way. *Cancer Cell* **16**, 369-377, doi:10.1016/j.ccr.2009.09.024 (2009).

57 Yang, A. et al. Nucleolin maintains embryonic stem cell self-renewal by suppression of p53 protein-dependent pathway. *J Biol Chem* **286**, 43370-43382, doi:10.1074/jbc.M111.225185 (2011).

58 Liu, X. et al. Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* **537**, 558-562, doi:10.1038/nature19362 (2016).

59 Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).

60 Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244, doi:10.1038/nature14450 (2015).

61 Jiang, Y. et al. Genome-wide analyses of chromatin interactions after the loss of Pol I, Pol II, and Pol III. *Genome Biol* **21**, 158, doi:10.1186/s13059-020-02067-3 (2020).

62 Bello, S. M., Smith, C. L. & Eppig, J. T. Allele, phenotype and disease data at Mouse Genome Informatics: improving access and analysis. *Mamm Genome* **26**, 285-294, doi:10.1007/s00335-015-9582-y (2015).

63 Borsos, M. & Torres-Padilla, M. E. Building up the nucleus: nuclear organization in the establishment of totipotency and pluripotency during mammalian development. *Gene Dev* **30**, 611-621, doi:10.1101/gad.273805.115 (2016).

64 Fulka, H. & Aoki, F. Nucleolus Precursor Bodies and Ribosome Biogenesis in Early Mammalian Embryos: Old Theories and New Discoveries. *Biol Reprod* **94**, 143, doi:10.1095/biolreprod.115.136093 (2016).

65 Kresoja-Rakic, J. & Santoro, R. Nucleolus and rRNA Gene Chromatin in Early Embryo Development. *Trends Genet* **35**, 868-879, doi:10.1016/j.tig.2019.06.005 (2019).

66 Savić, N. et al. lncRNA maturation to initiate heterochromatin formation in the nucleolus is required for exit from pluripotency in ESCs. *Cell Stem Cell* **15**, 720-734, doi:10.1016/j.stem.2014.10.005 (2014).

67 Boija, A. et al. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855.e1816, doi:10.1016/j.cell.2018.10.042 (2018).

68 Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958, doi:10.1126/science.aar3958 (2018).

69 Gibson, B. A. et al. Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, 470-484.e421, doi:10.1016/j.cell.2019.08.037 (2019).

70 Wu M. et al.. lncRNA SLERT controls phase separation of FC/DFCs to facilitate Pol I transcription. *Science* **373**, 547-555. doi: 10.1126/science.abf6582 (2021).

71 Leslie, M. Separation anxiety. *Science* **371**, 336-338, doi:10.1126/science.371.6527.336 (2021).

72 Wang, S. et al. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598-602, doi:10.1126/science.aaf8084 (2016).

1467 73 Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging.
1468 Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**,
1469 aaa6090, doi:10.1126/science.aaa6090 (2015).

1470 74 Beliveau, B. J. et al. Single-molecule super-resolution imaging of chromosomes
1471 and in situ haplotype visualization using Oligopaint FISH probes. *Nat Comm* **6**, 7147,
1472 doi:10.1038/ncomms8147 (2015).

1473 75 Bolger, A. M., Marc, L. & Bjoern, U. Trimmomatic: a flexible trimmer for Illumina
1474 sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

1475 76 Daehwan, K., Ben, L. & Salzberg, S. L. HISAT: a fast spliced aligner with low
1476 memory requirements. *Nat Methods* **12**, 357-360 (2015).

1477 77 Mihaela, P. et al. StringTie enables improved reconstruction of a transcriptome
1478 from RNA-seq reads. *Nat Biotechnol* **33**, 290-295 (2015).

1479 78 Cabili, M. N. et al. Integrative annotation of human large intergenic noncoding
1480 RNAs reveals global properties and specific subclasses. *Gene Dev* **25**, 1915 (2011).

1481 79 Zilionis, R. et al. Single-cell barcoding and sequencing using droplet microfluidics.
1482 *Nat Protoc* **12**, 44-73 (2016).

1483 80 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat*
1484 *Methods* **9**, 357-359. doi: 10.1038/nmeth.1923 (2012).

1485 81 Ramírez F. et al. deepTools2: a next generation web server for deep-sequencing
1486 data analysis. *Nucleic Acids Res* **44**, W160-5. doi: 10.1093/nar/gkw257 (2016).

1487 82 Heinz S. et al. Simple combinations of lineage-determining transcription factors
1488 prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*
1489 **38**, 576-589. doi: 10.1016/j.molcel.2010.05.004 (2010).

1490 83 Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data
1491 processing. *Genome Biol* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).

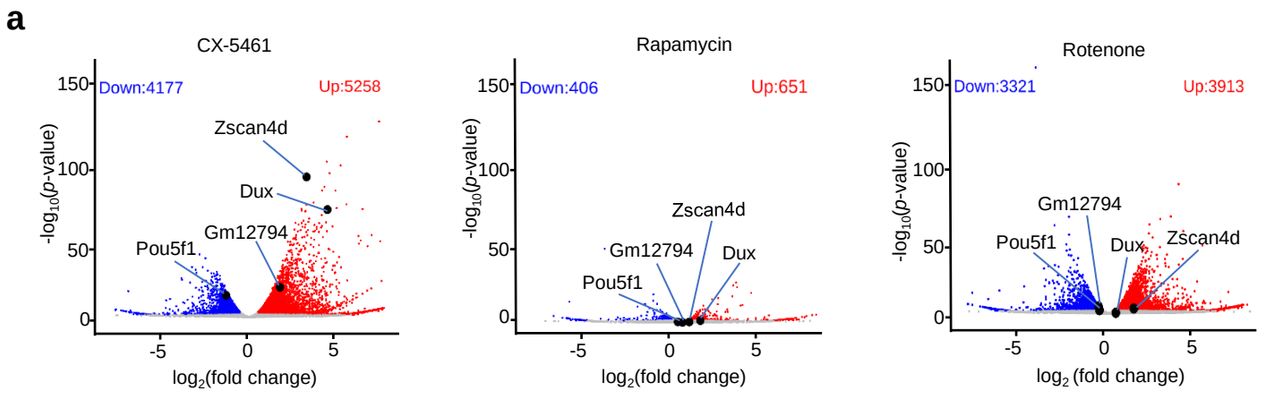
1492 84 Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions
1493 reveals folding principles of the human genome. *Science* **326**, 289-293,
1494 doi:10.1126/science.1181369 (2009).

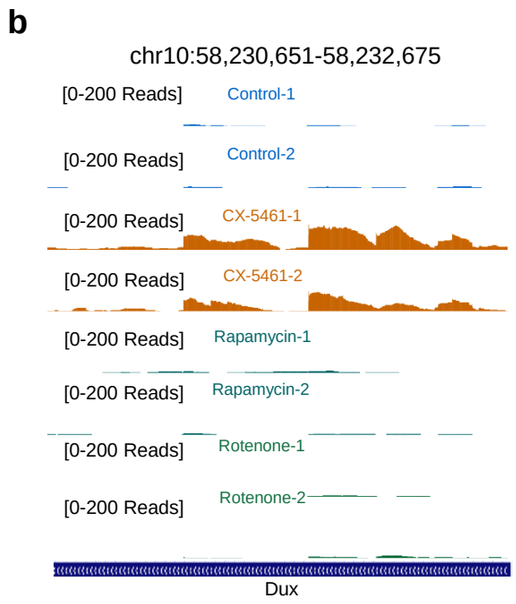
1495 85 Fidel, R. et al. High-resolution TADs reveal DNA sequences underlying genome
1496 organization in flies. *Nat Commun*, **9**, 189, doi.org/10.1038/s41467-017-02525-w
1497 (2018).

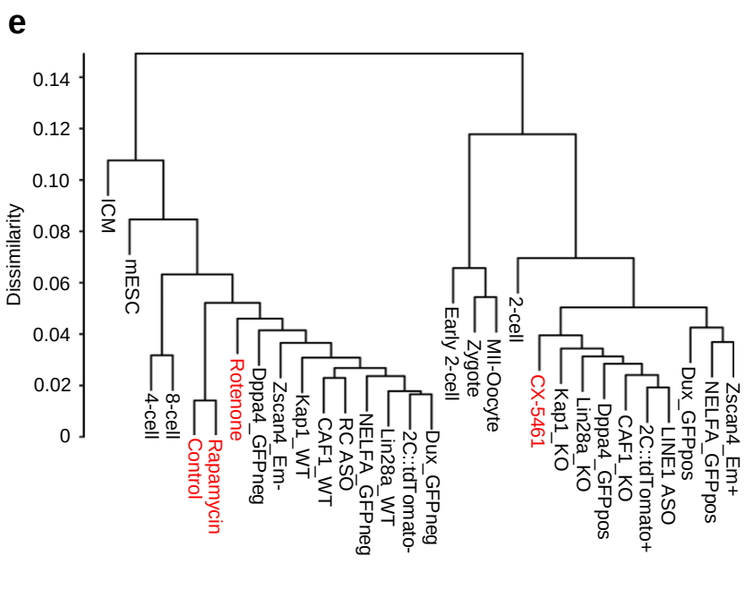
1498 86 Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of
1499 chromosome organization. *Nat Methods* **9**, 999-1003, doi:10.1038/nmeth.2148
1500 (2012).

1501 87 Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization
1502 at oocyte-to-zygote transition. *Nature* **544**, 110-114, doi:10.1038/nature21711
1503 (2017).

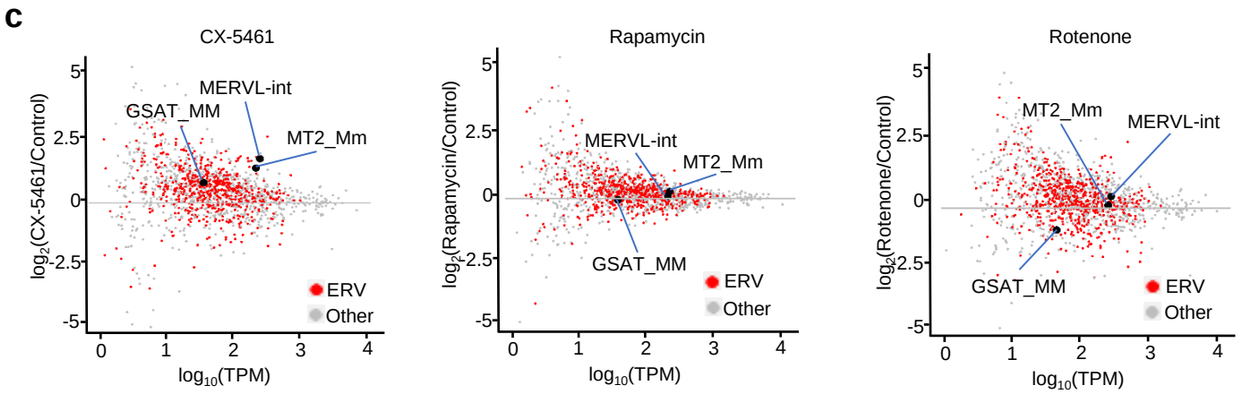
88 Durand N.C. et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98. doi: 10.1016/j.cels.2016.07.002 (2016).

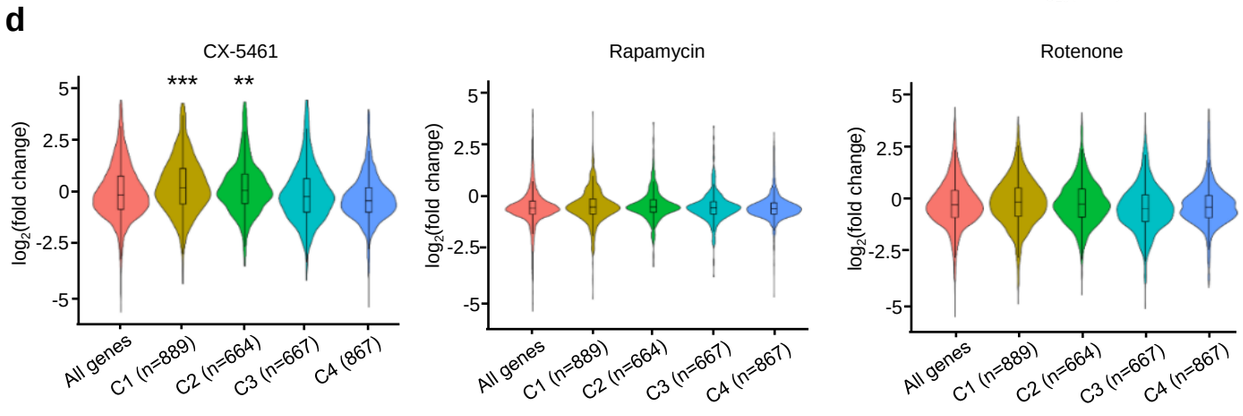

a
CX-5461
Down:4177
Up:5258
Zscan4d
Dux
Gm12794
Pou5f1
-log10(p-value)
log2(fold change)
Rapamycin
Down:406
Up:651
Zscan4d
Gm12794
Dux
Pou5f1
Rotenone
Down:3321
Up:3913
Gm12794
Pou5f1
Dux
Zscan4d

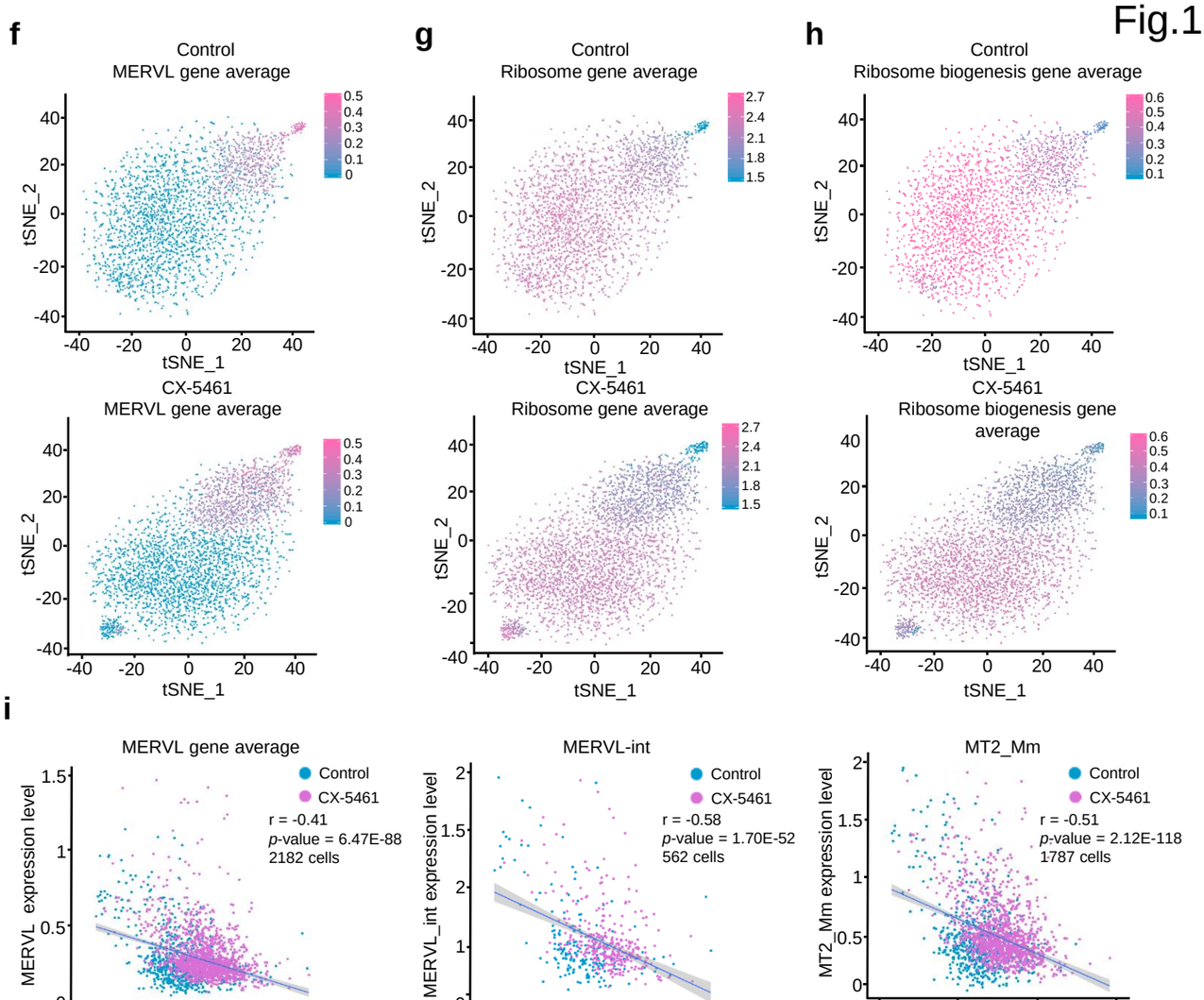

b
chr10:58,230,651-58,232,675
[0-200 Reads] Control-1
[0-200 Reads] Control-2
[0-200 Reads] CX-5461-1
[0-200 Reads] CX-5461-2
[0-200 Reads] Rapamycin-1
[0-200 Reads] Rapamycin-2
[0-200 Reads] Rotenone-1
[0-200 Reads] Rotenone-2
Dux

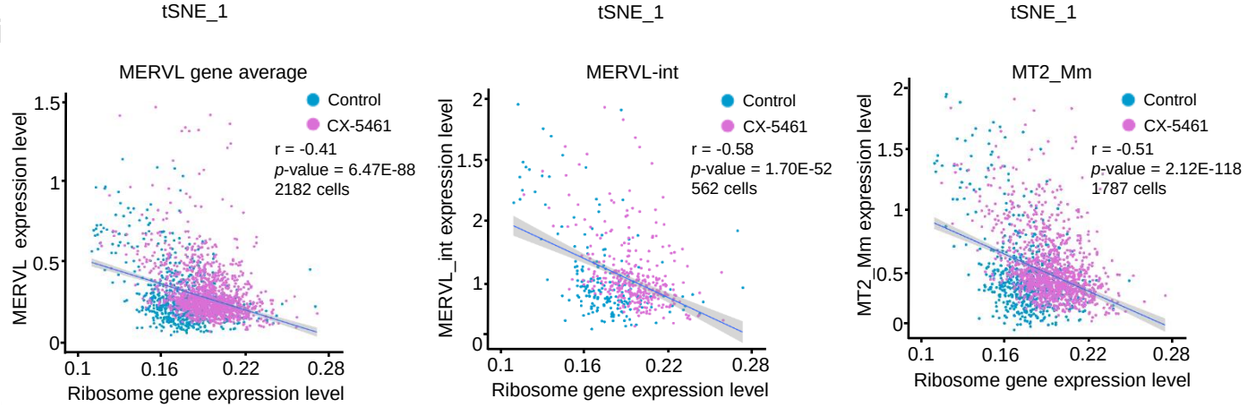

e
Dissimilarity
ICM
mESC
8-cell
4-cell
Rotenone
Rapamycin
Control
Dppa4_GFPneg
Zscan4_Em-
Kap1_WT
CAF1_WT
RC ASO
NELFA_GFPneg
Lin28a_WT
2C::tdTomato-
Dux_GFPneg
2-cell
Early 2-cell
Zygote
MII-Oocyte
CX-5461
Kap1_KO
Lin28a_KO
Dppa4_GFPpos
CAF1_KO
2C::tdTomato+
LINE1 ASO
Dux_GFPpos
NELFA_GFPpos
Zscan4_Em+

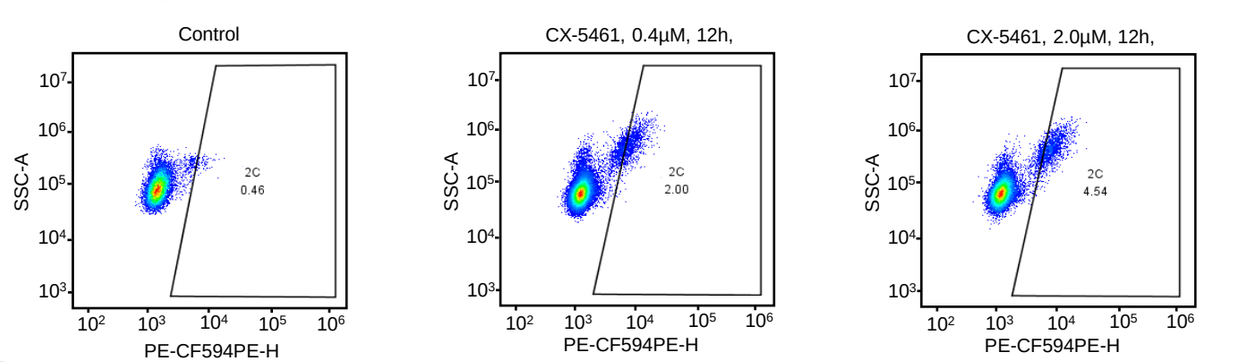
Dux


c
CX-5461
Rapamycin
Rotenone
MERVL-int
GSAT_MM
MT2_Mm
ERV
Other
log2(CX-5461/Control)
log2(Rapamycin/Control)
log2(Rotenone/Control)
log10(TPM)


d
CX-5461
Rapamycin
Rotenone
***
**
log2(fold change)
All genes
C1 (n=889)
C2 (n=664)
C3 (n=667)
C4 (867)


f
Control
MERVL gene average
CX-5461
MERVL gene average
g
Control
Ribosome gene average
CX-5461
Ribosome gene average
h
Control
Ribosome biogenesis gene average
CX-5461
Ribosome biogenesis gene average
tSNE_1
tSNE_2


i
MERVL gene average
MERVL-int
MT2_Mm
Control
CX-5461
r = -0.41
p-value = 6.47E-88
2182 cells
r = -0.58
p-value = 1.70E-52
562 cells
r = -0.51
p-value = 2.12E-118
1787 cells
MERVL expression level
MERVL_int expression level
MT2_Mm expression level
Ribosome gene expression level


j
Control
CX-5461, 0.4μM, 12h,
CX-5461, 2.0μM, 12h,
2C
0.46
2.00
4.54
SSC-A
PE-CF594PE-H

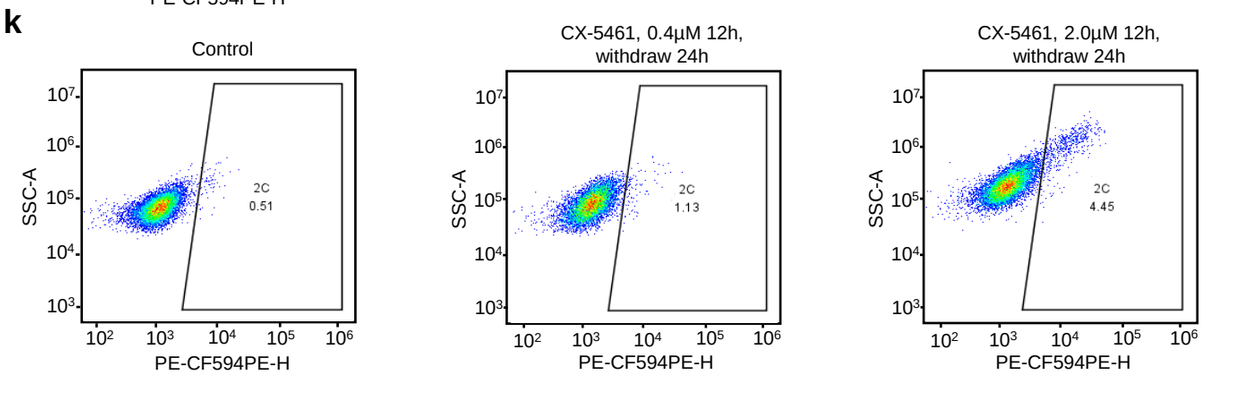
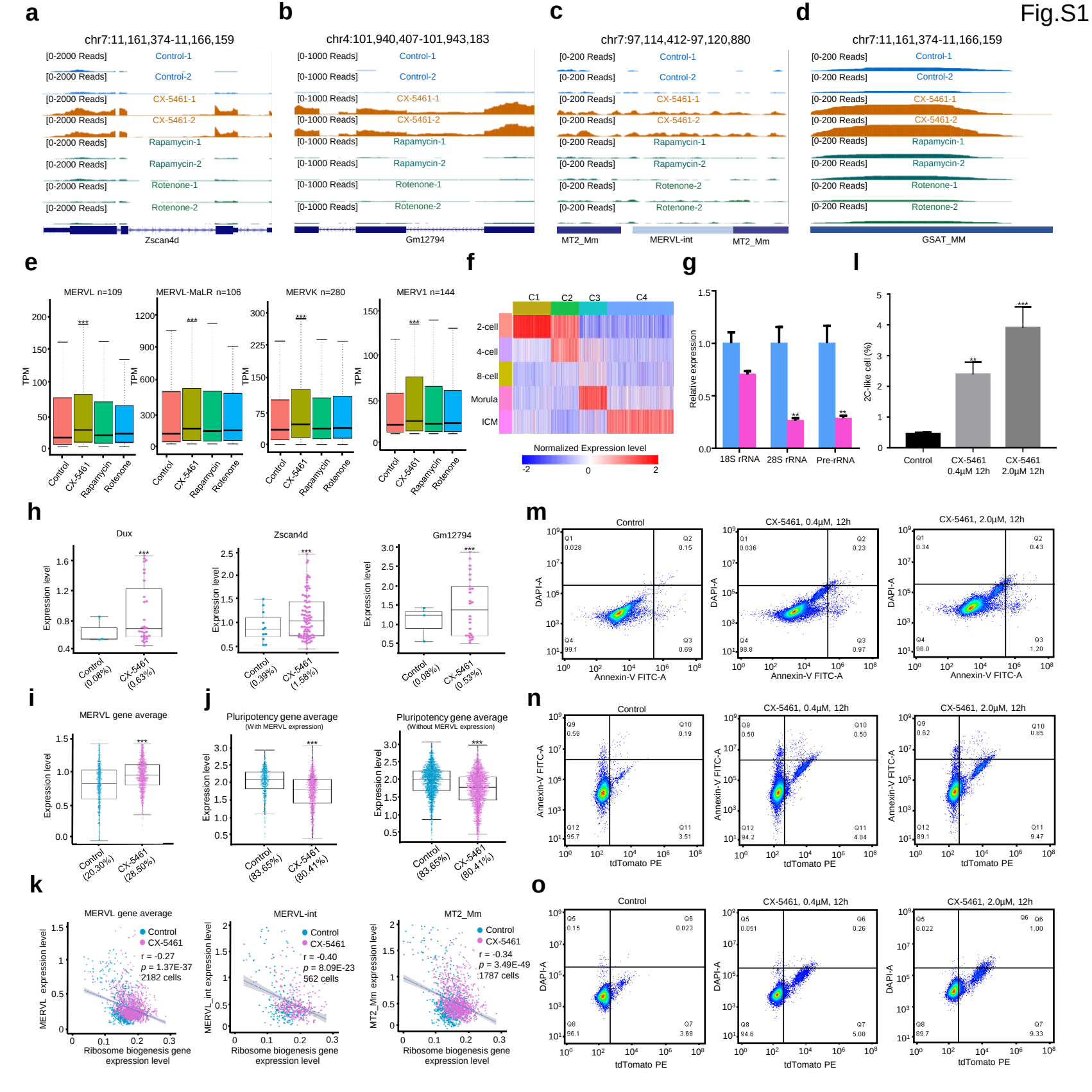

k
Control
CX-5461, 0.4μM 12h, withdraw 24h
CX-5461, 2.0μM 12h, withdraw 24h
2C
0.51
1.13
4.45
SSC-A
PE-CF594PE-H

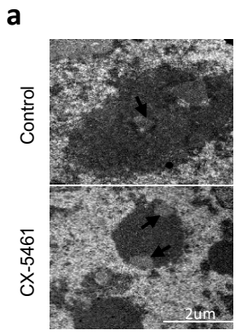
Fig.1: Inhibition of rRNA biogenesis activated 2C-like transcriptional program and induced an expanded 2C-like cell population in mES cells. a) Volcano plots of RNA-sequencing data comparing gene expression of control and cellular stress inducer-treated mES cells (GEO accession GSE166041). **b)** UCSC Genome Browser viewing of RNA-sequencing results at the *Dux* locus (GEO accession GSE166041). **c)** MA plots of RNA-sequencing data comparing repeat sequence expression of control and cellular stress inducer-treated mES cells (GEO accession GSE166041). **d)** Violin plots demonstrating the expressional changes of stage-specific gene clusters of mouse pre-implantation embryos under different types of cellular stress treatment; **: p<0.01, ***: p<0.001, Mann-Whitney U test (GEO accession GSE166041). **e)** Hierarchical clustering of transcriptomes from our study, published 2C-like cell model studies and pre-implantation mouse embryos; Control, CX-5461, Rotenone and Rapamycin (GEO accession GSE166041), 2C::tdTomato+ and 2C::tdTomato- (GEO accession GSE33923); Zscan4_Em+ and Zscan4_Em- (GEO accession GSE51682); Kap1_KO and Kap1_WT (GEO accession GSE74278); CAF1_WT and CAF1_KO (GEO accession GSE85632), Dux_GFPpos and Dux_GFPneg (GEO accession GSE85632); LINE1 ASO and RC ASO (GEO accession GSE100939); Dppa4_GFPpos and Dppa4_GFPneg (GEO accession GSE120953), NELFA_GFPpos and NELFA_GFPneg (GEO accession GSE113671); Lin28a_KO and Lin28a_WT (GEO accession GSE164420); MII-Oocyte, Zygote, Early 2-cell, 2-cell, 4-cell, 8-cell, ICM and mES cells (GEO accession GSE66390). **f)** tSNE feature plots demonstrating the averaged expression levels of *MERVL* genes in 2981 control mES cells and 3219 CX-5461-treated mES cells. **g)** tSNE feature plots demonstrating the averaged expression levels of ribosome genes in 2981 control mES cells and 3219 CX-5461-treated mES cells. **h)** tSNE feature plots demonstrating the averaged expression levels of ribosome biogenesis genes in 2981 control mES cells and 3219 CX-5461-treated mES cells (GEO accession GSE166041). **i)** Scatter plots demonstrating negative correlation of expression level between *MERVL/MERVL-int/MT2_Mm* and ribosome genes; Each dot represents a single cell with detectable ERV expression; *r* denotes correlation coefficient; *p*-value was obtained by cor.test function in R software (GEO accession GSE166041). **j)** FACS analysis on 2C::tdTomato+ mES cells upon different treatment doses of CX-5461, showing the change of percentage of 2C-like cells. **k)** FACS analysis on 2C::tdTomato+ mES cells after 12 hour treatment and 24 hour withdrawal of CX-5461, showing the change of percentage of 2C-like cells.

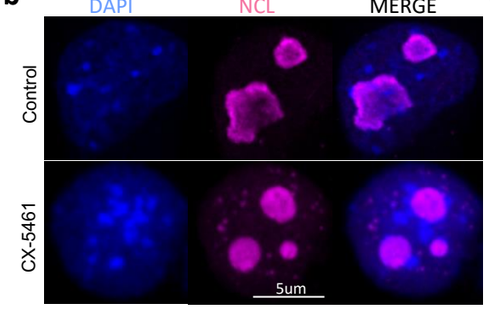
Fig.S1

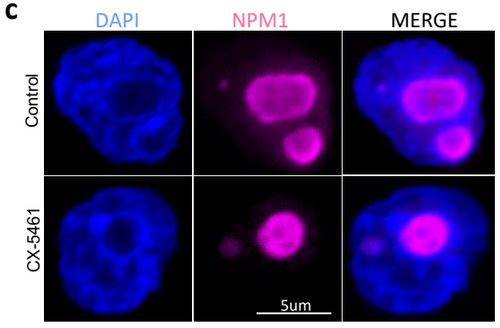

a
chr7:11,161,374-11,166,159
Control-1
Control-2
CX-5461-1
CX-5461-2
Rapamycin-1
Rapamycin-2
Rotenone-1
Rotenone-2
Zscan4d
b
chr4:101,940,407-101,943,183
Gm12794
c
chr7:97,114,412-97,120,880
MT2_Mm
MERVL-int
d
chr7:11,161,374-11,166,159
GSAT_MM
e
MERVL n=109
MERVL-MaLR n=106
MERVK n=280
MERV1 n=144
TPM
Control
CX-5461
Rapamycin
Rotenone
f
C1
C2
C3
C4
2-cell
4-cell
8-cell
Morula
ICM
Normalized Expression level
g
Relative expression
18S rRNA
28S rRNA
Pre-rRNA
l
2C-like cell (%)
Control
CX-5461 0.4μM 12h
CX-5461 2.0μM 12h
h
Dux
Zscan4d
Gm12794
Expression level
Control (0.08%)
CX-5461 (0.63%)
Control (0.39%)
CX-5461 (1.58%)
Control (0.08%)
CX-5461 (0.53%)
m
Control
CX-5461, 0.4μM, 12h
CX-5461, 2.0μM, 12h
DAPI-A
Annexin-V FITC-A
Q1 0.028
Q2 0.15
Q4 99.1
Q3 0.69
Q1 0.036
Q2 0.23
Q4 98.8
Q3 0.97
Q1 0.34
Q2 0.43
Q4 98.0
Q3 1.20
i
MERVL gene average
Control (20.30%)
CX-5461 (28.50%)
j
Pluripotency gene average
(With MERVL expression)
Control (83.65%)
CX-5461 (80.41%)
Pluripotency gene average
(Without MERVL expression)
Control (83.65%)
CX-5461 (80.41%)
n
Control
CX-5461, 0.4μM, 12h
CX-5461, 2.0μM, 12h
Annexin-V FITC-A
tdTomato PE
Q9 0.59
Q10 0.19
Q12 95.7
Q11 3.51
Q9 0.50
Q10 0.50
Q12 94.2
Q11 4.84
Q9 0.62
Q10 0.85
Q12 89.1
Q11 9.47
k
MERVL gene average
MERVL-int
MT2_Mm
Control
CX-5461
r = -0.27
p = 1.37E-37
2182 cells
r = -0.40
p = 8.09E-23
562 cells
r = -0.34
p = 3.49E-49
1787 cells
MERVL expression level
MERVL_int expression level
MT2_Mm expression level
Ribosome biogenesis gene expression level
o
Control
CX-5461, 0.4μM, 12h
CX-5461, 2.0μM, 12h
DAPI-A
tdTomato PE
Q5 0.15
Q6 0.023
Q8 96.1
Q7 3.68
Q5 0.051
Q6 0.26
Q8 94.6
Q7 5.08
Q5 0.022
Q6 1.00
Q8 89.7
Q7 9.33

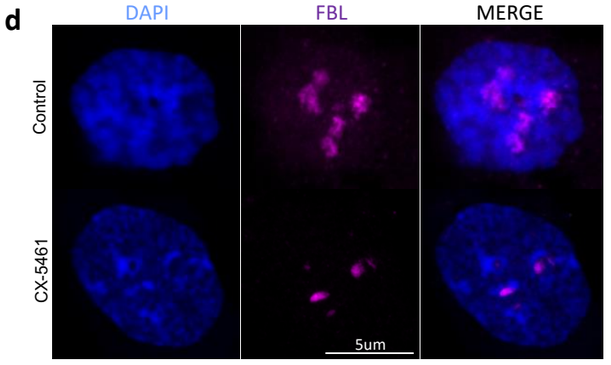
Extended Data Fig.S1: Inhibition of rRNA biogenesis activated 2C-like transcriptional program and induced an expanded 2C-like cell population in mES cells, related to Fig.1. a) UCSC Genome Browser viewing of RNA-sequencing results in *Zscan4d* gene locus (GEO accession GSE166041). **b)** UCSC Genome Browser viewing of RNA-sequencing results in *Gm12794* gene locus (GEO accession GSE166041). **c)** UCSC Genome Browser viewing of RNA-sequencing results in a representative ERV (*MERVL-int* and *MT2_Mm*) gene locus (GEO accession GSE166041). **d)** UCSC Genome Browser viewing of RNA-sequencing results in a representative *GSAT_MM* gene loci (GEO accession GSE166041). **e)** Boxplots show the expression levels of major ERV gene classes under different types of cellular stress treatment. n denotes the number of sub-classes of ERV genes, ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **f)** A heatmap plot demonstrating four developmental stage-specific gene clusters derived from RNA-seq data of pre-implantation mouse embryos (GEO accession GSE97778). **g)** qRT-PCR quantification of rRNA expression in control mES cells and CX-5461 treated mES cells, **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3. **h)** Boxplots demonstrating the expression levels of 2C marker genes of Dux, Zscan4d and Gm12794 in control mES cells and CX-5461 treated mES cells; Each point denoted a single cell; The number of parentheses denotes the percentage of cells expressing these genes; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **i)** Boxplots demonstrating the averaged expression levels of *MERVL* genes in control mES cells and CX-5461 treated mES cells with single cell RNA-sequencing analysis; Each point denoted a single cell; The number of parentheses denotes the percentage of cells expressing *MERVL* genes; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **j)** The expression levels of pluripotency genes in control mES cells and CX-5461 treated mES cells; Each point denoted a single cell; The number of parentheses denotes the percentage of cells expressing these genes; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **k)** Scatter plots demonstrating negative correlation of expression level between *MERVL/MERVL-int/MT2_Mm* and ribosome biogenesis genes; Each dot represents a single cell with detectable ERV expression; *r* denotes correlation coefficient; *p*-value was obtained by cor.test function in R software (GEO accession GSE166041). **l)** The percentage of 2C::tdTomato positive cells was quantified using FACS analysis in control mES cells and CX-5461-treated mES cells; Data are means ± SD; SD: Standard Deviation; **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=5. **m)** FACS analysis on Annexin-V FITC, marker for early cell apoptosis,

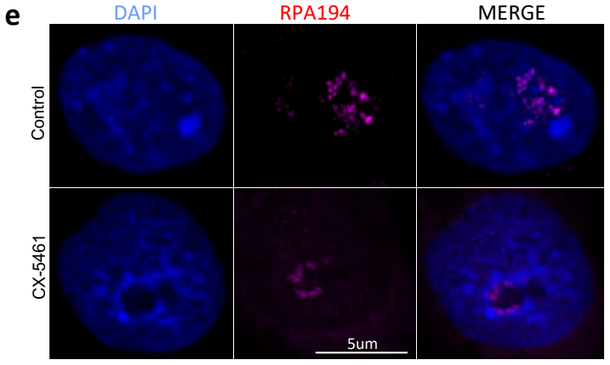
and DAPI, marker for late cell apoptosis upon different treatment doses of CX-5461; **n)** FACS analysis on Annexin-V FITC and 2C::tdTomato positive mES cells upon different treatment doses of CX-5461. **o)** FACS analysis on DAPI and 2C::tdTomato positive mES cells upon different treatment doses of CX-5461.

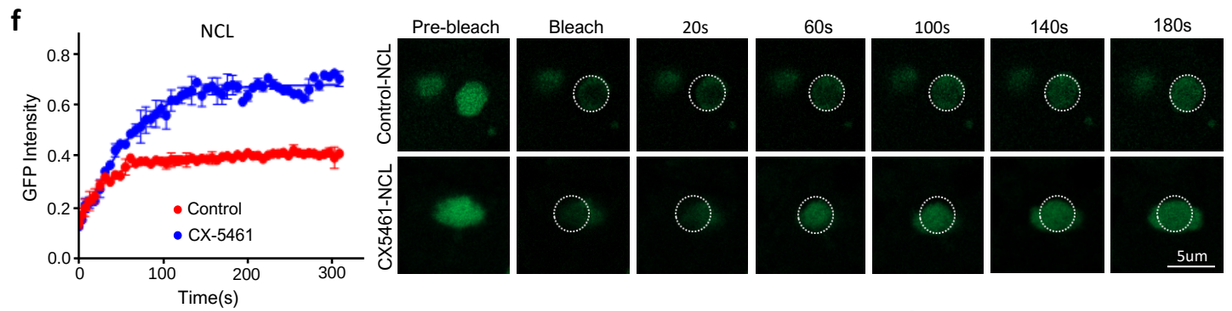

a
Control
CX-5461
2um

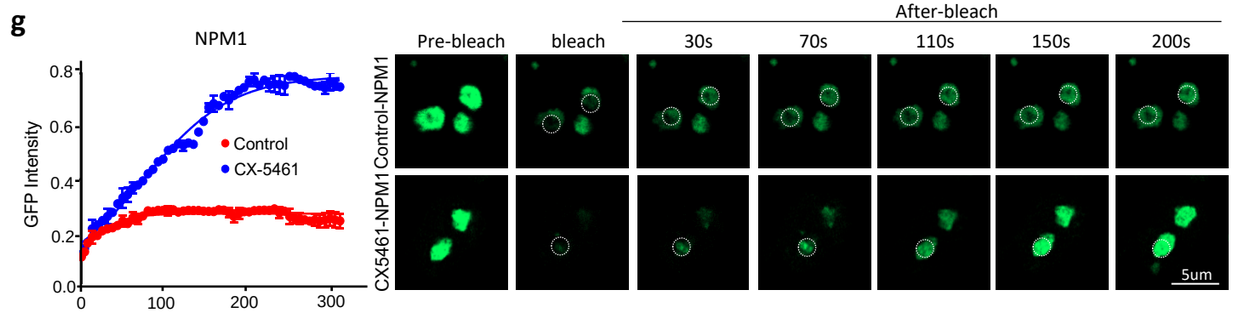

b
DAPI
NCL
MERGE
Control
CX-5461
5um

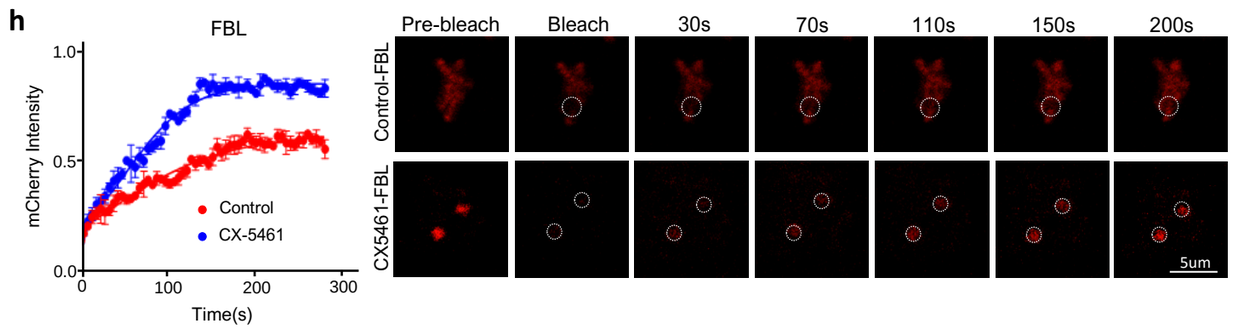

c
DAPI
NPM1
MERGE
Control
CX-5461
5um

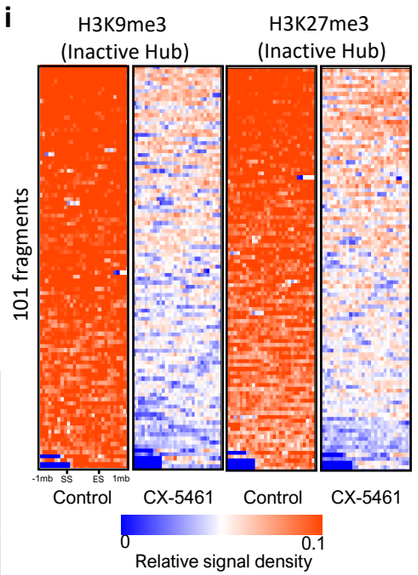

d
DAPI
FBL
MERGE
Control
CX-5461
5um

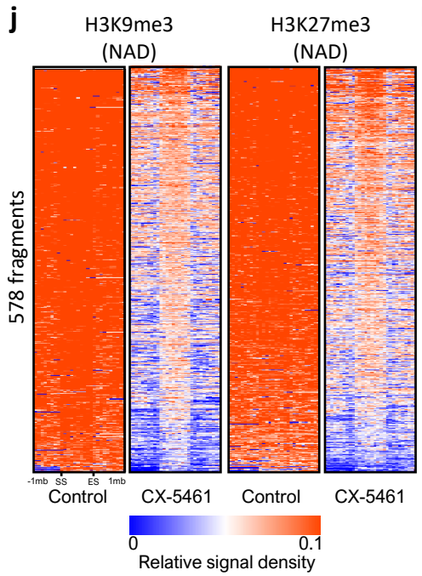

e
DAPI
RPA194
MERGE
Control
CX-5461
5um

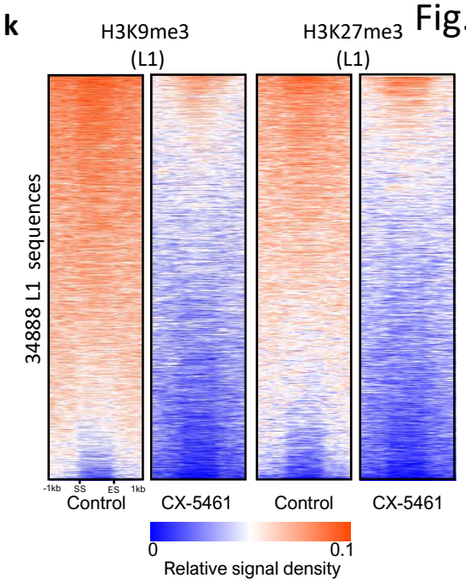

f
NCL
GFP Intensity
Time(s)
Control
CX-5461
Pre-bleach
Bleach
20s
60s
100s
140s
180s
Control-NCL
CX5461-NCL
5um

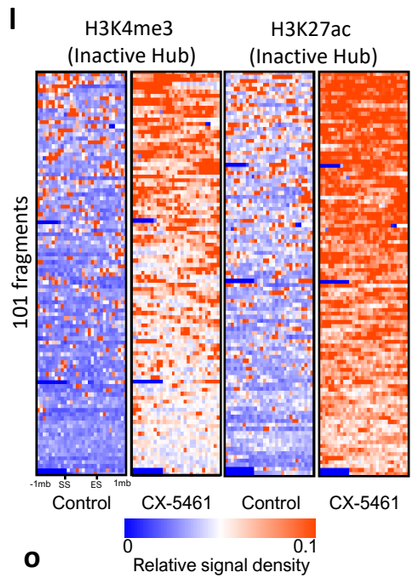

g
NPM1
GFP Intensity
Control
CX-5461
After-bleach
Pre-bleach
bleach
30s
70s
110s
150s
200s
Control-NPM1
CX5461-NPM1
5um

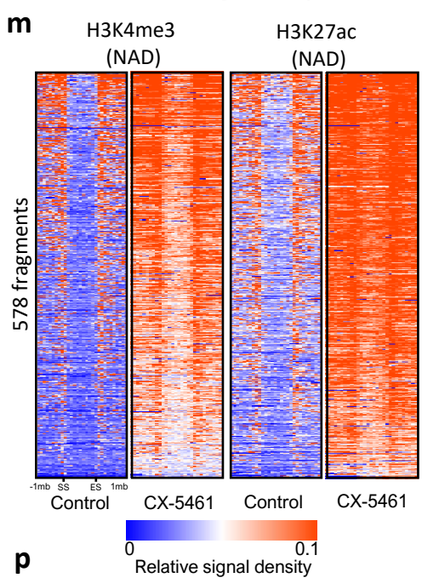

h
FBL
mCherry Intensity
Time(s)
Control
CX-5461
Pre-bleach
Bleach
30s
70s
110s
150s
200s
Control-FBL
CX5461-FBL
5um

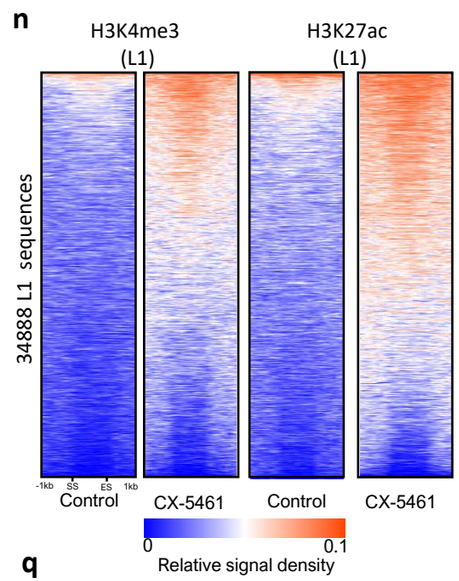

i
H3K9me3
(Inactive Hub)
H3K27me3
(Inactive Hub)
101 fragments
-1mb SS ES 1mb
Control
CX-5461
Control
CX-5461
0
0.1
Relative signal density

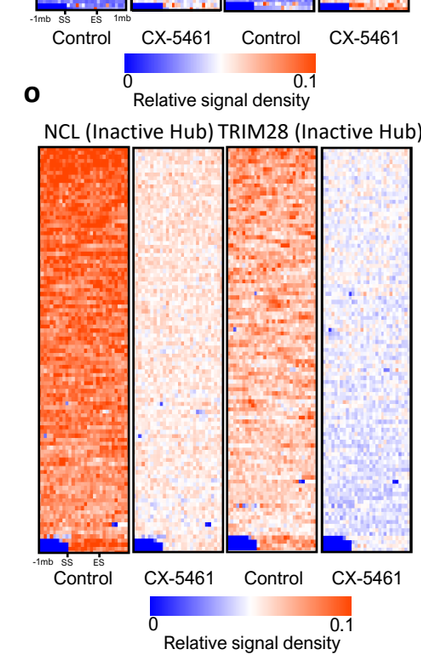

j
H3K9me3
(NAD)
H3K27me3
(NAD)
578 fragments
-1mb SS ES 1mb
Control
CX-5461
Control
CX-5461
0
0.1
Relative signal density

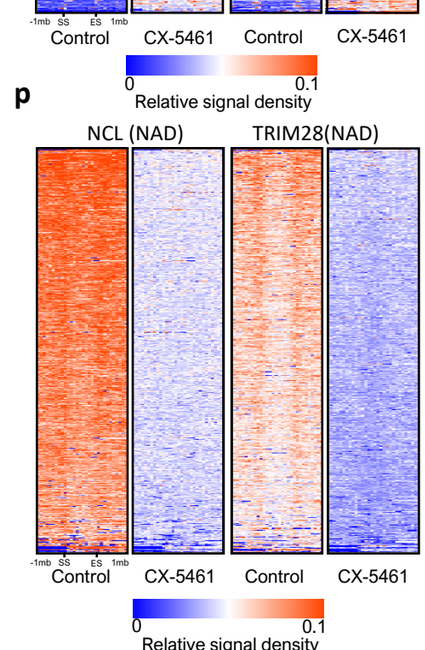

k
H3K9me3
(L1)
H3K27me3
(L1)
34888 L1 sequences
-1kb SS ES 1kb
Control
CX-5461
Control
CX-5461
0
0.1
Relative signal density


l
H3K4me3
(Inactive Hub)
H3K27ac
(Inactive Hub)
101 fragments
-1mb SS ES 1mb
Control
CX-5461
Control
CX-5461
0
0.1
Relative signal density


m
H3K4me3
(NAD)
H3K27ac
(NAD)
578 fragments
-1mb SS ES 1mb
Control
CX-5461
Control
CX-5461
0
0.1
Relative signal density


n
H3K4me3
(L1)
H3K27ac
(L1)
34888 L1 sequences
-1kb SS ES 1kb
Control
CX-5461
Control
CX-5461
0
0.1
Relative signal density


o
NCL (Inactive Hub)
TRIM28 (Inactive Hub)
-1mb SS ES
Control
CX-5461
Control
CX-5461
0
0.1
Relative signal density


p
NCL (NAD)
TRIM28(NAD)
-1mb SS ES 1mb
Control
CX-5461
Control
CX-5461
0
0.1
Relative signal density

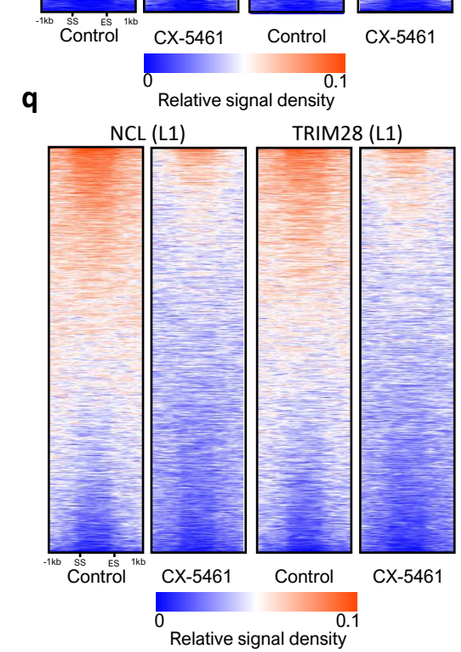
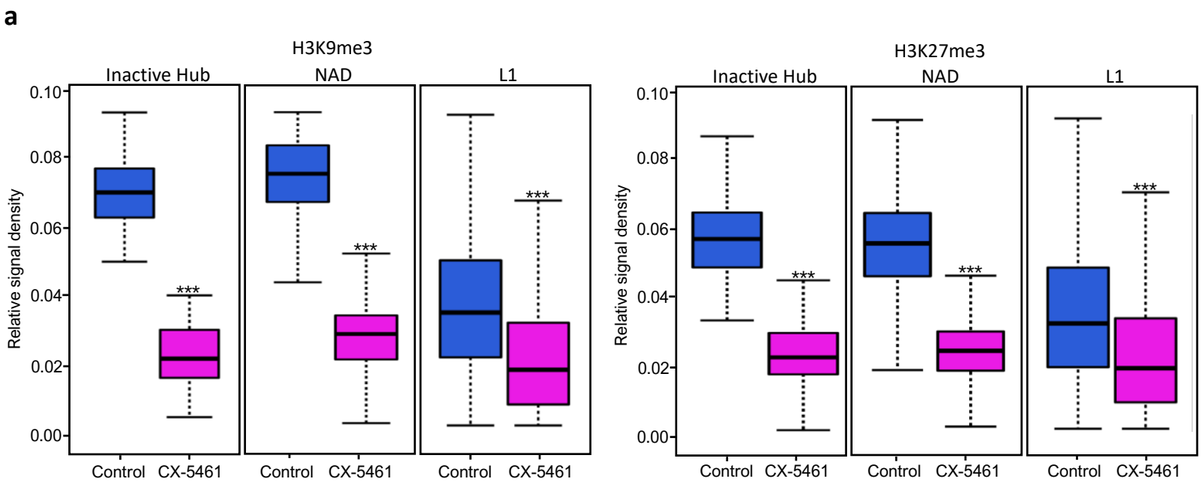
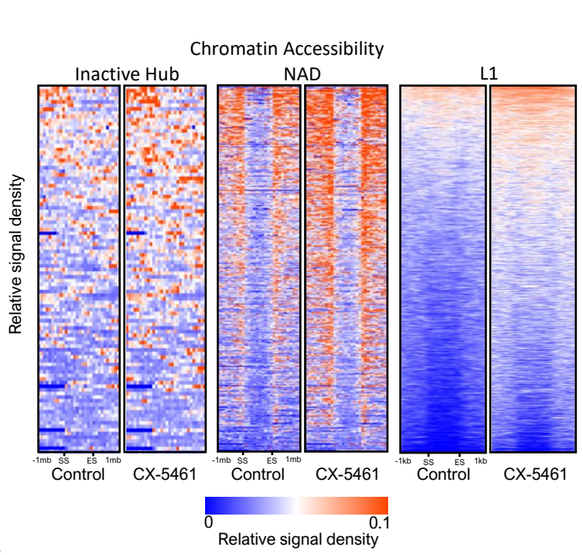
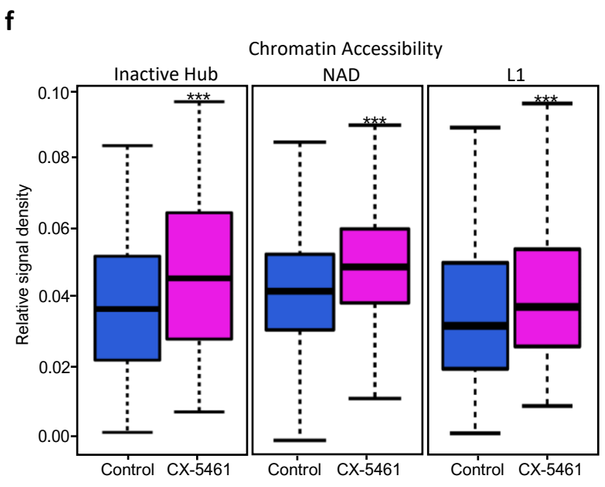

q
NCL (L1)
TRIM28 (L1)
-1kb SS ES 1kb
Control
CX-5461
Control
CX-5461
0
0.1
Relative signal density

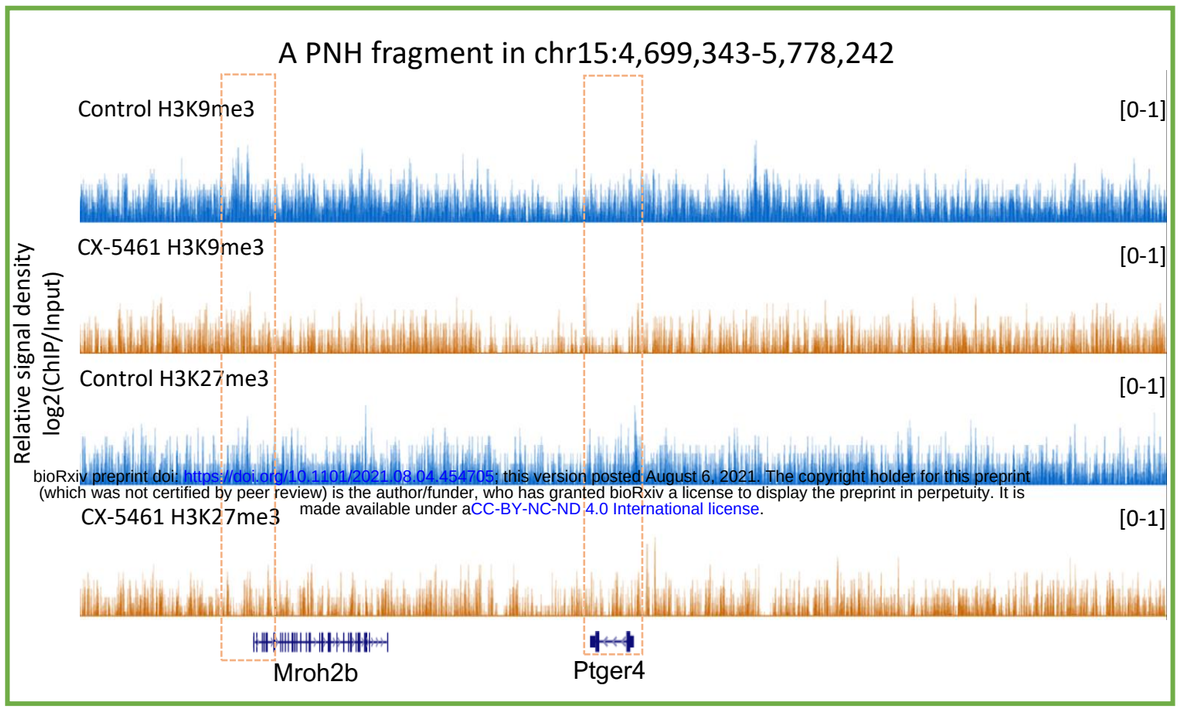
Fig.2: Deficiency of rRNA biogenesis disrupted normal nucleolar LLPS and epigenetic state of PNH region. a) CX-5461 treatment causes abnormal nucleolar structure with electron microscopy. **b)** Immunofluorescence staining of NCL in control mES cells and CX-5461-treated mES cells. **c)** Immunofluorescence staining of NPM1 in control mES cells and CX-5461-treated mES cells. **d)** Immunofluorescence staining of FBL in control mES cells and CX-5461 treated mES cells. **e)** Immunofluorescence staining of RPA194 in control and CX-5461 treated mES cells. **f)** FRAP analysis showing CX-5461 treatment causes accelerated recovery after photobleaching of NCL. Shown images are representative of 4 times of experiments. **g)** FRAP analysis showing CX-5461 treatment causes accelerated recovery after photobleaching of NPM1. Shown images are representative of 4 times of experiments. **h)** FRAP analysis showing CX-5461 treatment causes accelerated recovery after photobleaching of FBL. Shown images are representative of 4 times of experiments. **i)** Heatmap plots demonstrate the levels of H3K9me3 and H3K27me3 on within 1mb region around start and end sites of Inactive Hub. The regions of different lengths of Inactive Hub fragments were fitted to 1mb (GEO accession GSE166041). **j)** Heatmap plots demonstrate the levels of H3K9me3 and H3K27me3 within 1mb region around start and end sites of NAD. The regions of different lengths of NAD fragments were fitted to 1mb (GEO accession GSE166041). **k)** Heatmap plots demonstrate the levels of H3K9me3 and H3K27me3 within 1kb region around start and end sites of L1 (GEO accession GSE166041). The regions of different lengths of L1 sequences were fitted to 1kb. **l)** Heatmap plots demonstrate the level of H3K4me3 and H3K27ac within 1mb region around start and end sites of Inactive Hub (GEO accession GSE166041). The regions of different lengths of Inactive Hub fragments were fitted to 1mb. **m)** Heatmap plots demonstrate the levels of H3K4me3 and H3K27ac within 1mb region around start and end sites of NAD. The regions of different lengths of NAD fragments were fitted to 1mb (GEO accession GSE166041). **n)** Heatmap plots demonstrate the level of H3K4me3 and H3K27ac on within 1kb region around start and end sites of L1 (GEO accession GSE166041). The regions of different lengths of L1 sequences were fitted to 1kb. **o)** Heatmap plots demonstrate the level of NCL and TRIM28 within 1mb region around start and end sites of Inactive Hub (GEO accession GSE166041). The regions of different lengths of Inactive Hub fragments were fitted to 1mb. **p)** Heatmap plots demonstrate the levels of NCL and TRIM28 within 1mb region around start and end sites of NAD. The regions of different lengths of NAD fragments were fitted to 1mb (GEO accession GSE166041). **q)** Heatmap plots demonstrate the level of NCL

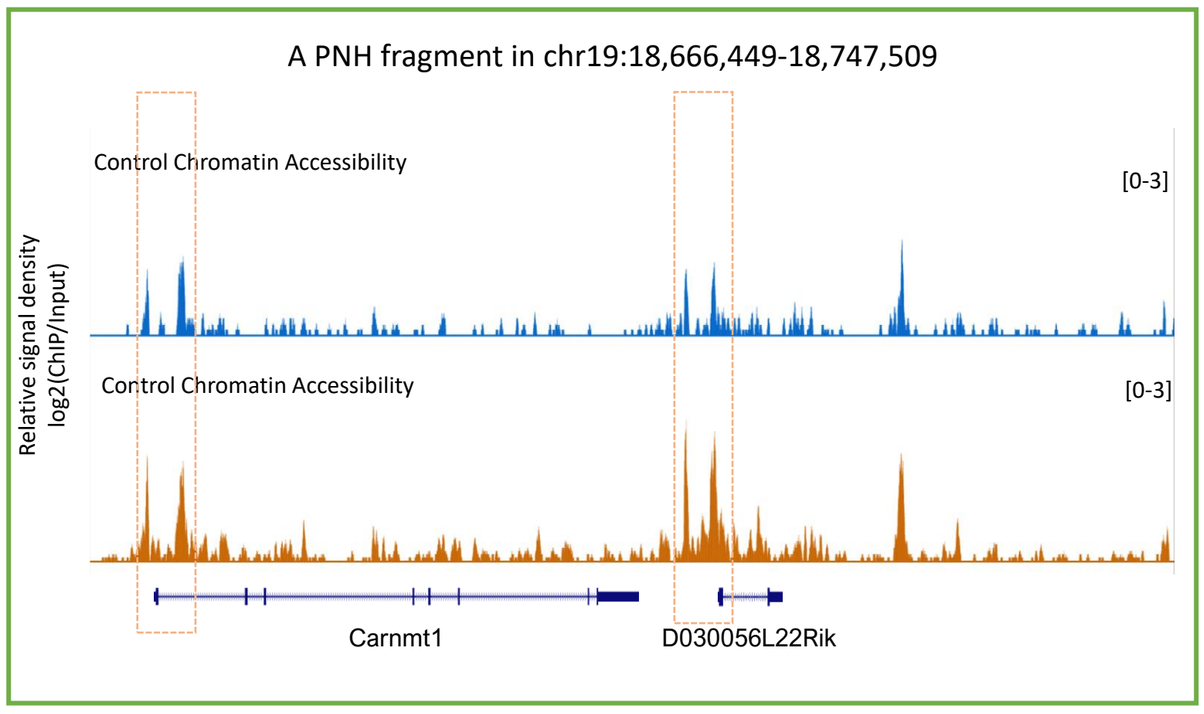
and TRIM28 within 1kb region around start and end sites of L1 (GEO accession GSE166041). The regions of different lengths of L1 sequences were fitted to 1kb.

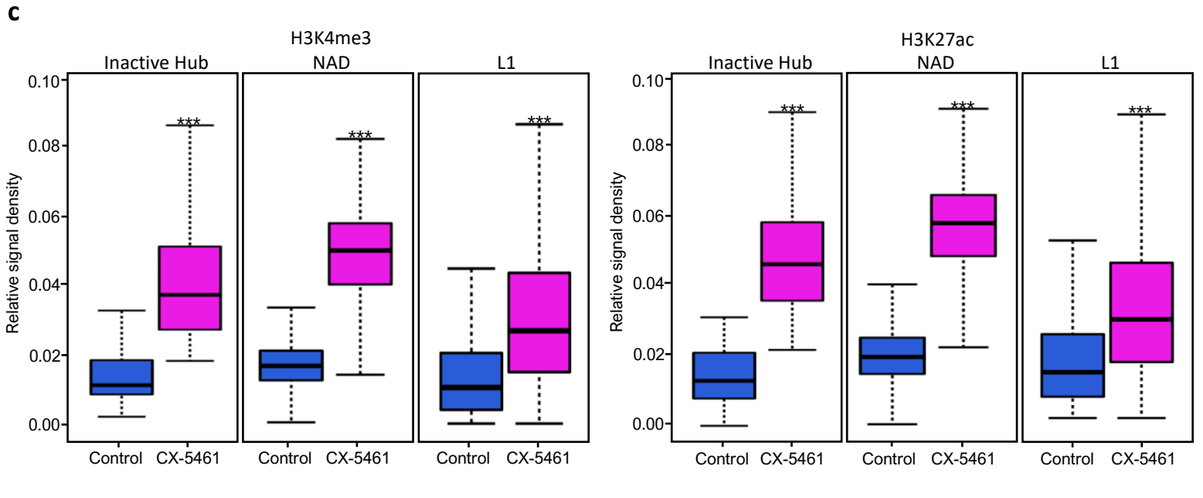


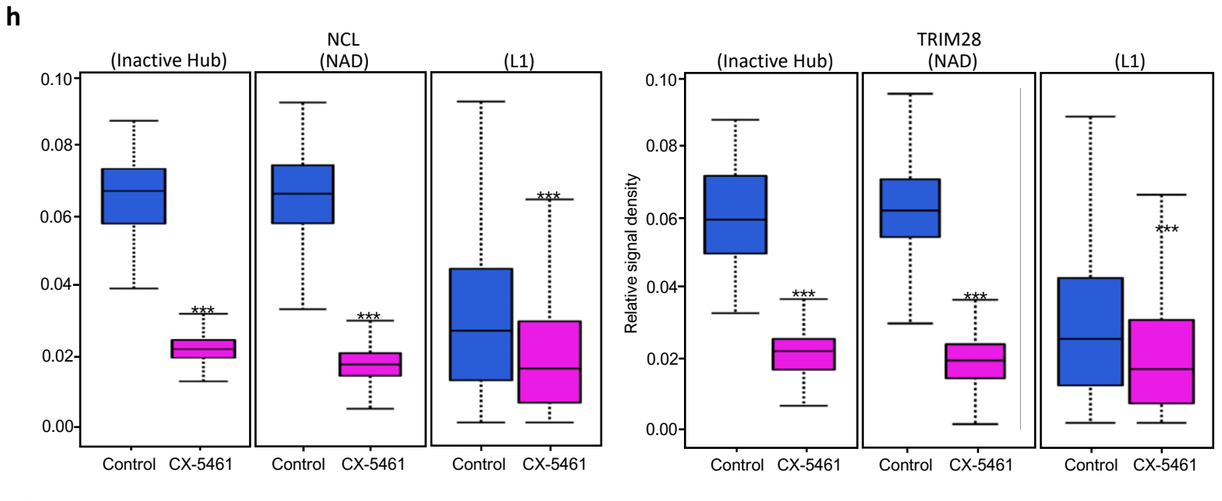


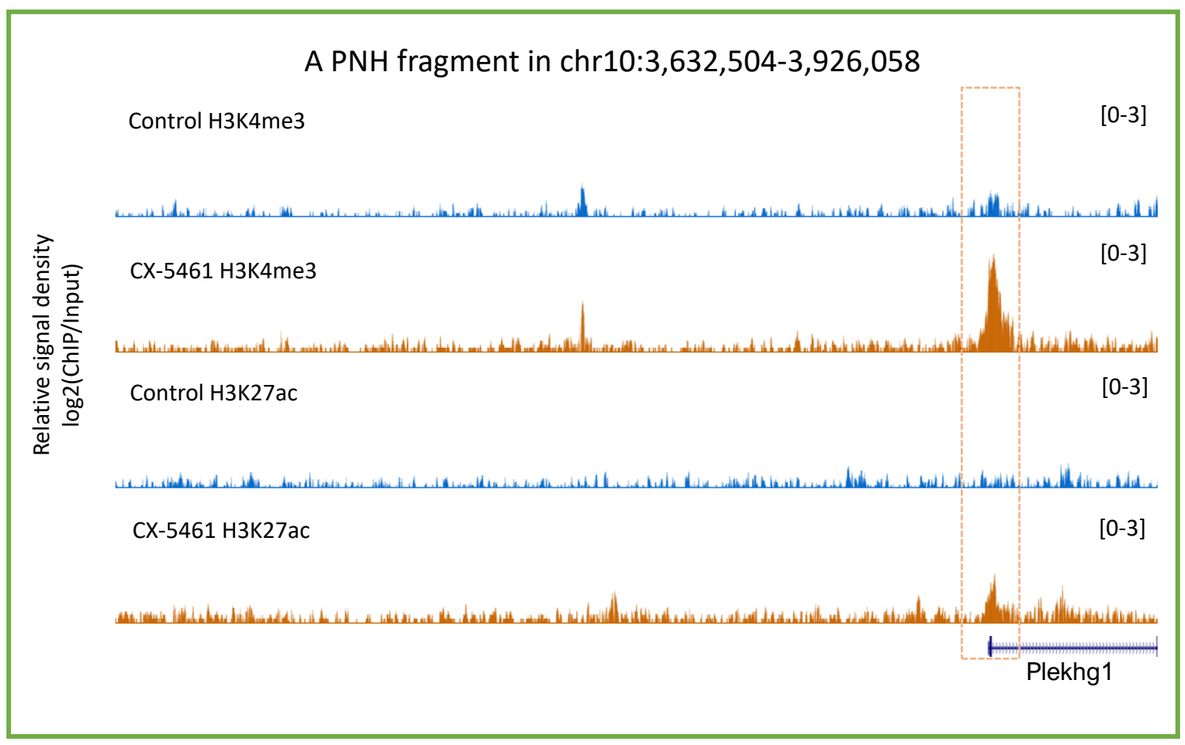


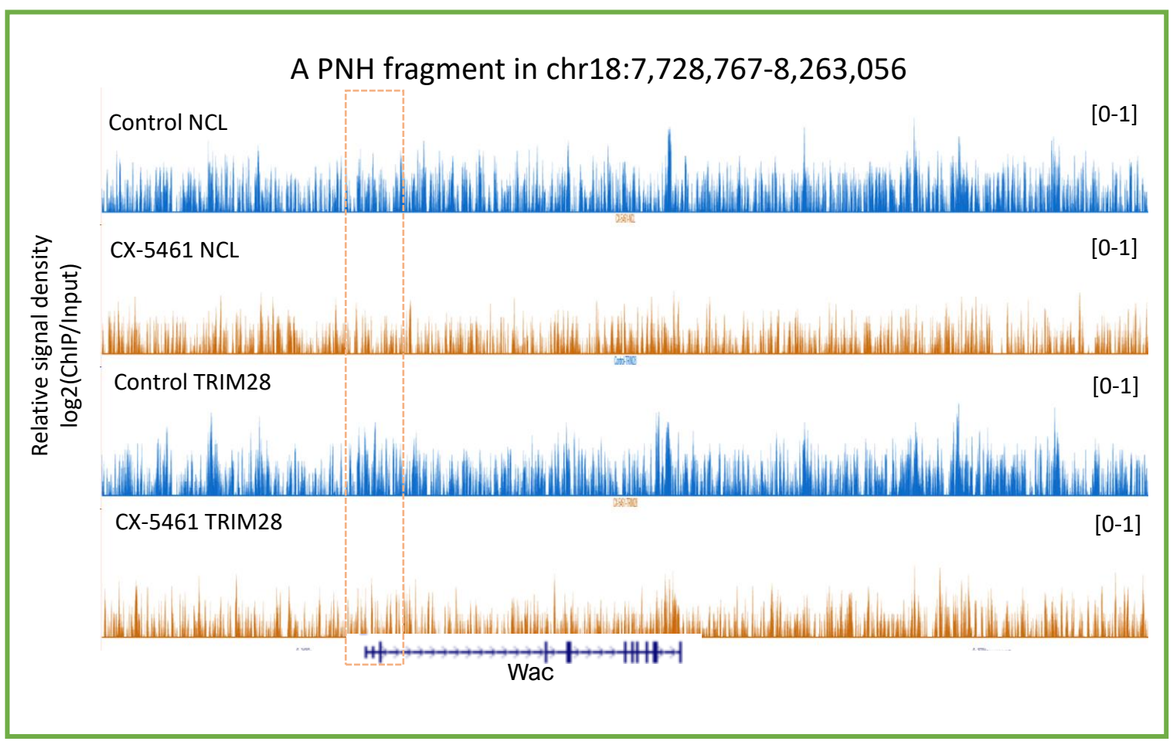












Extended Data Fig.S2: Deficiency of rRNA biogenesis disrupted normal nucleolar LLPS and epigenetic state of PNH region, related to Fig.2. a) Boxplots demonstrate the averaged H3K9me3 and H3K27me3 levels of 101 Inactive Hub fragments, 578 NAD fragments and 34888 L1 sequences; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **b)** UCSC Genome Browser viewing of H3K9me3 and H3K27me3 ChIP-seq signals in control and CX-5461 treated mES cells around a representative PNH fragment at chr15:4,699,343-5,778,242 (GEO accession GSE166041). **c)** Boxplots demonstrate the averaged H3K4me3 and H3K27ac levels of 101 Inactive Hub fragments, 578 NAD fragments and 34888 L1 sequences; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **d)** UCSC Genome Browser viewing of H3K4me3 and H3K27ac ChIP-seq signals in control mES cells and CX-5461 treated mES cells around a representative PNH fragment at chr10:3,632,504-3,926,058 (GEO accession GSE166041). **e)** Heatmap plots demonstrate the levels of chromatin accessibility on within 1mb region around start and end sites of Inactive Hub and NAD, and within 1kb region around start and end sites of L1 (GEO accession GSE166041). The regions of different lengths of Inactive Hub and NAD fragments were fitted to 1mb. The regions of different lengths of L1 sequences were fitted to 1kb. SS: start site of a chromatin fragment of PNH; ES: end site of a chromatin fragment of PNH. The PNH fragment was defined as the L1 contained regions overlapped with Inactive Hub and NAD. **f)** Boxplots demonstrate the averaged chromatin accessibility of 101 Inactive Hub fragments, 578 NAD fragments and 34888 L1 sequences; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **g)** UCSC Genome Browser viewing of ATAC-seq signals in control and CX-5461 treated mES cells around a representative PNH fragment at chr19:18,666,449-18,747,509 (GEO accession GSE166041). **h)** Boxplots demonstrate the averaged binding signals of NCL and TRIM28 on 101 Inactive Hub fragments, 578 NAD fragments and 34888 L1 sequences; **: p<0.01, ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041). **i)** UCSC Genome Browser viewing of NCL and TRIM28 ChIP-seq signals in control and CX-5461 treated mES cells around a representative PNH fragment at chr18:7,718,063-8,074,255 (GEO accession GSE166041).

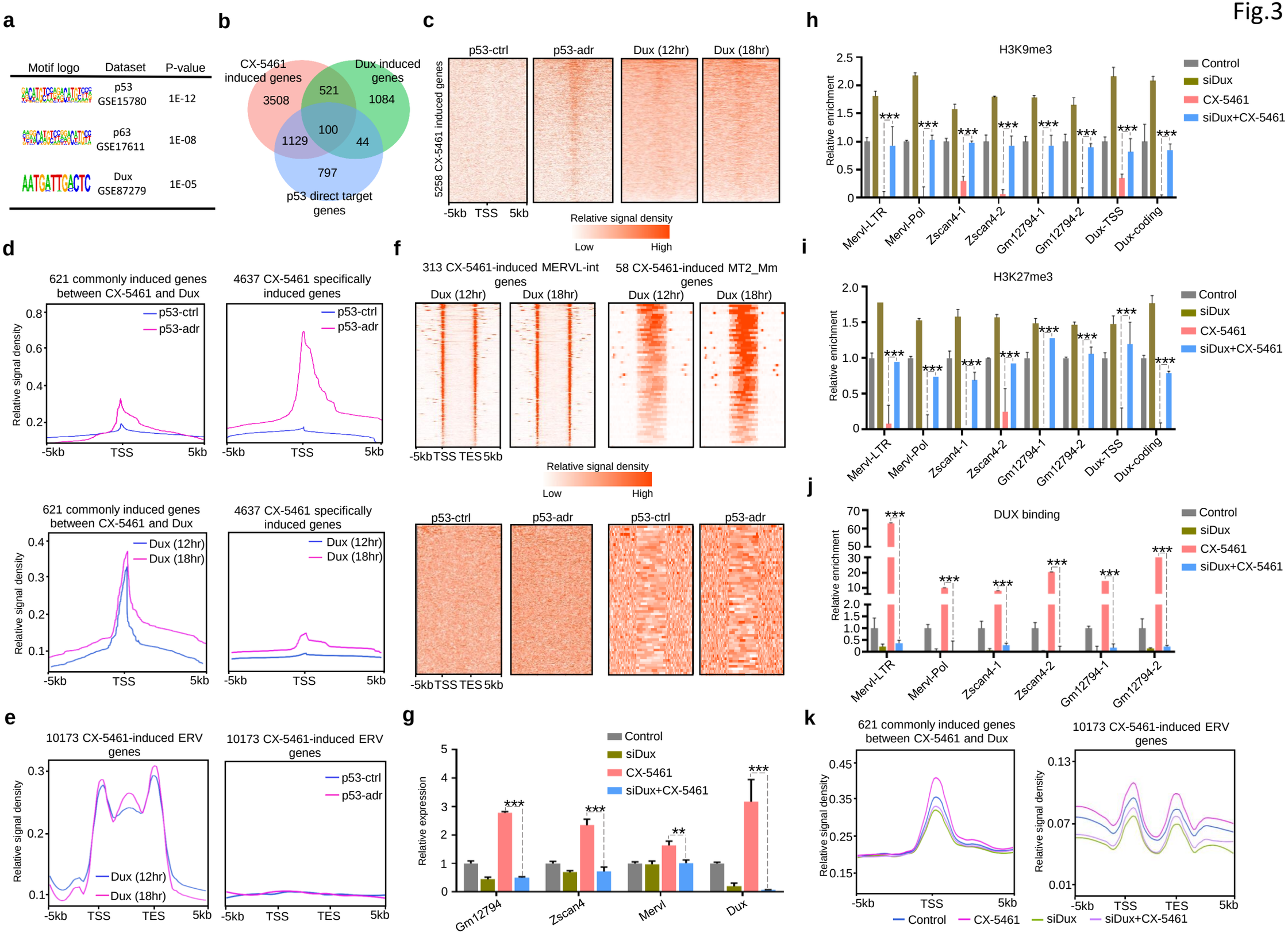
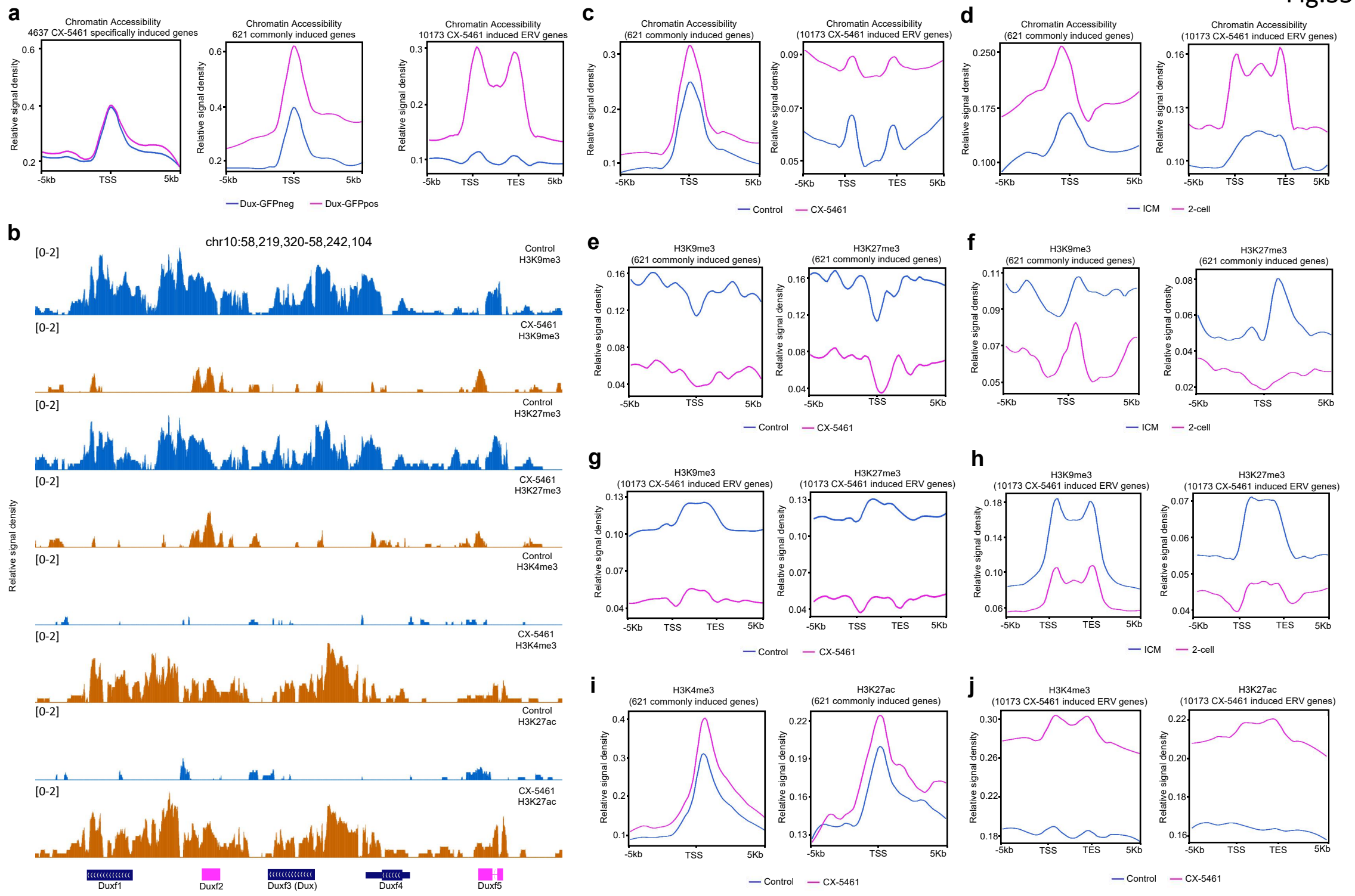

a
Motif logo
Dataset
P-value
p53 GSE15780
1E-12
p63 GSE17611
1E-08
Dux GSE87279
1E-05
b
CX-5461 induced genes
Dux induced genes
p53 direct target genes
3508
521
1084
100
1129
44
797
c
p53-ctrl
p53-adr
Dux (12hr)
Dux (18hr)
5258 CX-5461 induced genes
-5kb TSS 5kb
Relative signal density
Low
High
h
H3K9me3
Relative enrichment
Control
siDux
CX-5461
siDux+CX-5461
MervI-LTR
MervI-Pol
Zscan4-1
Zscan4-2
Gm12794-1
Gm12794-2
Dux-TSS
Dux-coding
d
621 commonly induced genes between CX-5461 and Dux
4637 CX-5461 specifically induced genes
Relative signal density
p53-ctrl
p53-adr
Dux (12hr)
Dux (18hr)
-5kb TSS 5kb
f
313 CX-5461-induced MERVL-int genes
58 CX-5461-induced MT2_Mm genes
-5kb TSS TES 5kb
i
H3K27me3
j
DUX binding
e
10173 CX-5461-induced ERV genes
Dux (12hr)
Dux (18hr)
-5kb TSS TES 5kb
g
Relative expression
Gm12794
Zscan4
MervI
Dux
k
621 commonly induced genes between CX-5461 and Dux
10173 CX-5461-induced ERV genes
Control
CX-5461
siDux
siDux+CX-5461

Fig.3: 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 5kb region around transcription start sites of 5258 CX-5461 induced genes using published p53 ChIP-seq data and Dux ChIP-seq data (GEO accession GSE26360 for p53 and GEO accession GSE85632 for Dux). **d)** Line plots demonstrate the meta-analysis results of p53 and Dux protein on within 5kb 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 GSE26360 for p53 and GEO accession 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 10173 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 5kb (GEO accession GSE26360 for p53 and GEO accession GSE85632 for Dux). **f)** Heatmap plots demonstrate the binding of p53 and Dux proteins on within 5kb 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 5kb (GEO accession GSE26360 for p53 and GEO accession GSE85632 for Dux). **g)** qRT-PCR showing the expression of *Dux* or 2C-related genes in *Dux* silenced mES cells; **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3; error bar: standard error of the mean. **h)** ChIP-PCR showing H3K9me3 levels of Dux or 2C-related genes in Dux silenced mES cells; *: p<0.05, **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3, error bar: standard error of mean. **i)** ChIP-PCR showing H3K27me3 levels of Dux or 2C-related genes in Dux silenced mES cells; *: p<0.05, **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3, error bar: standard error of mean. **j)** ChIP-PCR showing DUX protein binding levels on 2C-related genes in Dux silenced mES cells; *: p<0.05, **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3, error bar: standard error of mean. **k)**

Line plots demonstrate the meta-analysis results of chromatin accessibility in Dux silenced mES cells within 5kb region around transcription start sites or transcription start and end sites of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression and 10173 CX-5461 induced ERV genes using published ATAC-seq data. The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE166041)**.** p53-ctrl: untreated mES cells, p53-adr: mES cells treated with adriamycin, a DNA damage agent widely used to activate p53, Dux (12h): mES induced with doxycycline for 12 hours, Dux (18h): mES induced with doxycycline for 18 hours.

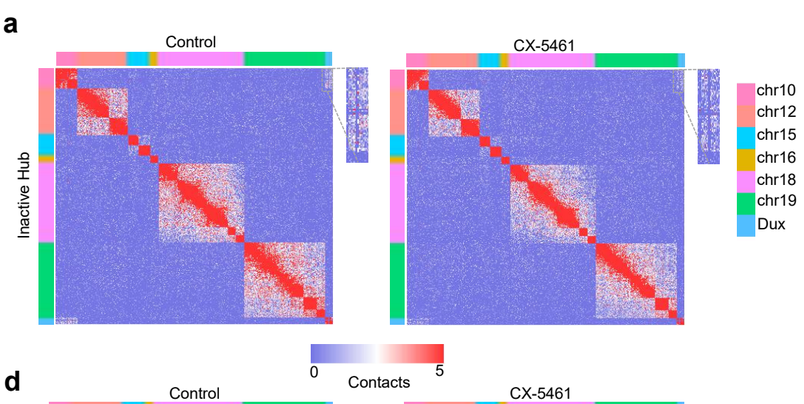
Fig.S3

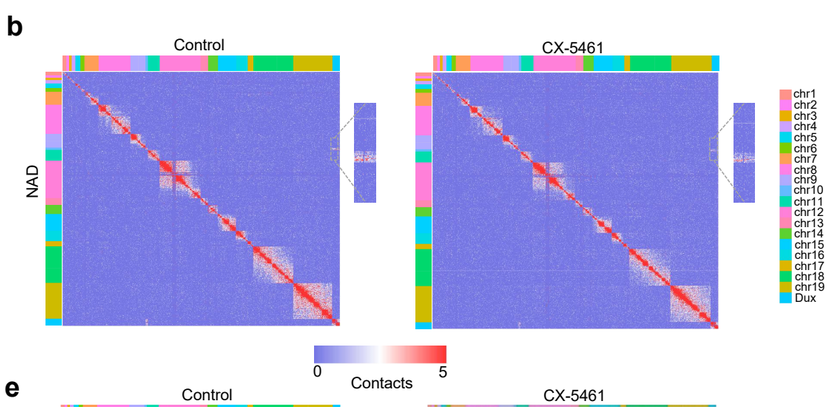

a
Chromatin Accessibility
4637 CX-5461 specifically induced genes
Chromatin Accessibility
621 commonly induced genes
Chromatin Accessibility
10173 CX-5461 induced ERV genes
Relative signal density
-5kb
TSS
TES
5kb
Dux-GFPneg
Dux-GFPpos
b
chr10:58,219,320-58,242,104
[0-2]
Control H3K9me3
CX-5461 H3K9me3
Control H3K27me3
CX-5461 H3K27me3
Control H3K4me3
CX-5461 H3K4me3
Control H3K27ac
CX-5461 H3K27ac
Relative signal density
Duxf1
Duxf2
Duxf3 (Dux)
Duxf4
Duxf5
c
Chromatin Accessibility
(621 commonly induced genes)
Chromatin Accessibility
(10173 CX-5461 induced ERV genes)
Control
CX-5461
d
Chromatin Accessibility
(621 commonly induced genes)
Chromatin Accessibility
(10173 CX-5461 induced ERV genes)
ICM
2-cell
e
H3K9me3
(621 commonly induced genes)
H3K27me3
(621 commonly induced genes)
Control
CX-5461
f
H3K9me3
(621 commonly induced genes)
H3K27me3
(621 commonly induced genes)
ICM
2-cell
g
H3K9me3
(10173 CX-5461 induced ERV genes)
H3K27me3
(10173 CX-5461 induced ERV genes)
Control
CX-5461
h
H3K9me3
(10173 CX-5461 induced ERV genes)
H3K27me3
(10173 CX-5461 induced ERV genes)
ICM
2-cell
i
H3K4me3
(621 commonly induced genes)
H3K27ac
(621 commonly induced genes)
Control
CX-5461
j
H3K4me3
(10173 CX-5461 induced ERV genes)
H3K27ac
(10173 CX-5461 induced ERV genes)
Control
CX-5461

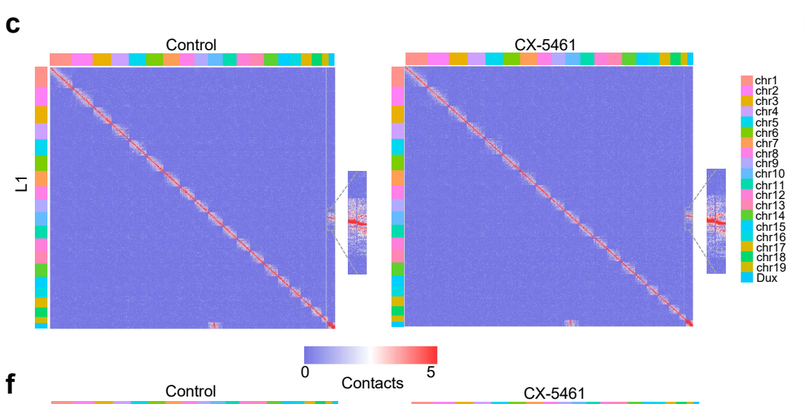
Extended Data Fig.S3: 2C/ERV genes were activated through Dux, related to Fig.3. **a)** Line plots demonstrate the meta-analysis results of chromatin accessibility in Dux-GFP positive mES cells and Dux-GFP negative mES cells within 5kb region around transcription start sites or transcription start and end sites of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression, 4637 specifically induced genes by CX-5461 and 10173 CX-5461 induced ERV genes using published ATAC-seq data. The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE85632). **b)** UCSC Genome Browser viewing of ChIP-seq signals around *Dux* locus (GEO accession GSE166041). **c)** Line plots demonstrate the meta-analysis results of chromatin accessibility of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression and 10173 CX-5461 induced ERV genes in control mES cells and CX-5461 treated mES cells. The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE166041). **d)** Line plots demonstrate the meta-analysis results of chromatin accessibility of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression and 10173 CX-5461 induced ERV genes in 2-cell embryo and ICM embryo. The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE66390). **e)** Line plots demonstrates the meta-analysis results of H3K9me3 and H3K27me3 levels of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression in control mES cells and CX-5461 treated mES cells (GEO accession GSE166041). **f)** Line plots demonstrate the meta-analysis results of H3K9me3 and H3K27me3 levels of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression in 2-cell embryo and ICM embryo (GEO accession GSE66390). **g)** Line plots demonstrate the meta-analysis results of H3K9me3 and H3K27me3 levels of 10173 CX-5461 induced ERV genes in control mES cells and CX-5461 treated mES cells; The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE166041). **h)** Line plots demonstrate the meta-analysis results of H3K9me3 and H3K27me3 levels of 10173 CX-5461 induced ERV genes in 2-cell embryo and ICM embryo; The regions of different lengths of ERV genes were fitted to 5kb (GEO accession GSE66390). **i)** Line plots demonstrate the meta-analysis results of H3K4me3 and H3K27ac levels of 621 commonly induced genes between CX-5461 treatment and *Dux* overexpression in control mES cells and CX-5461 treated mES cells (GEO accession GSE166041). **j)** Line plots demonstrate the meta-analysis results of H3K4me3 and H3K27ac levels of 10173 CX-5461 induced ERV genes in control mES cells and CX-5461 treated mES cells; The regions of different lengths of ERV genes were fitted to 5kb (GEO

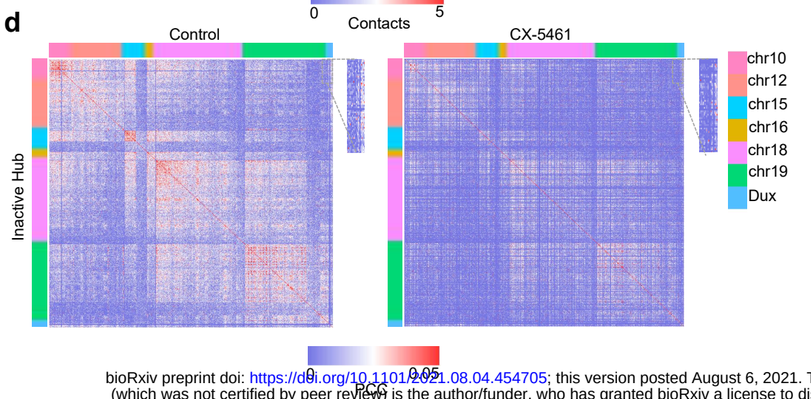
accession GSE66390).

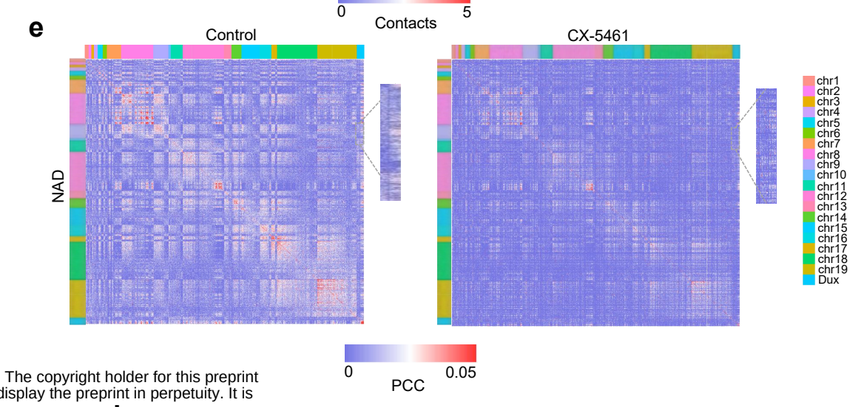
Fig.4

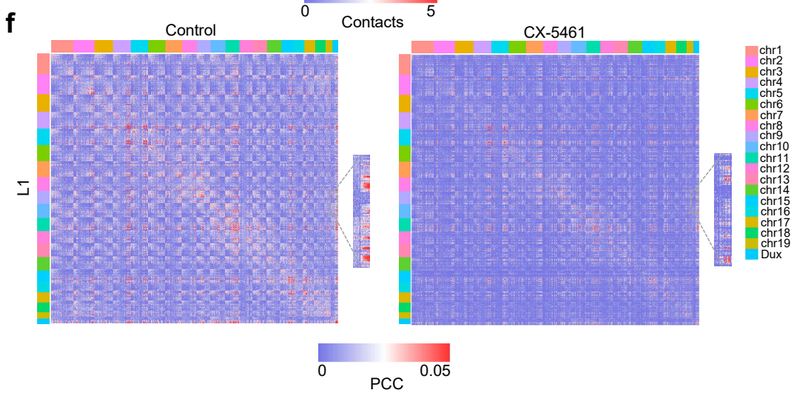


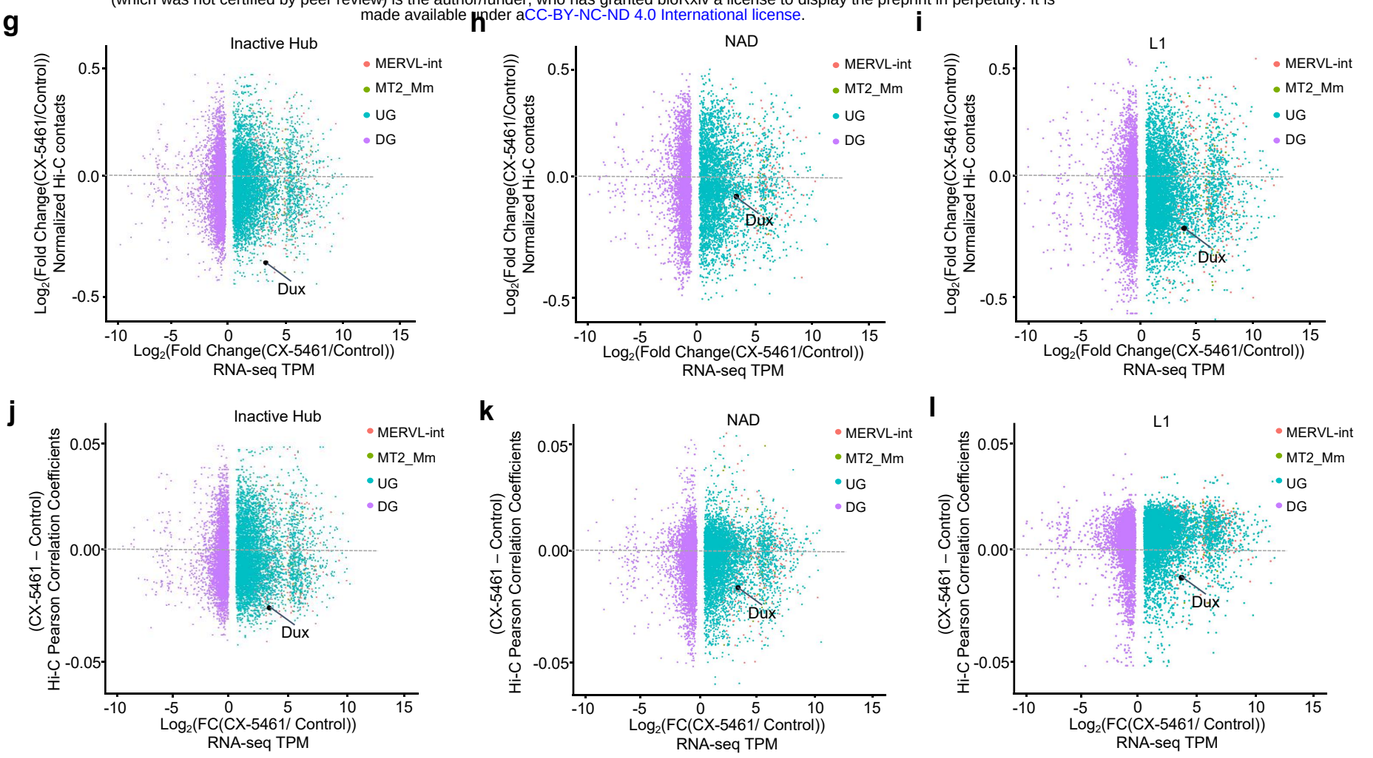


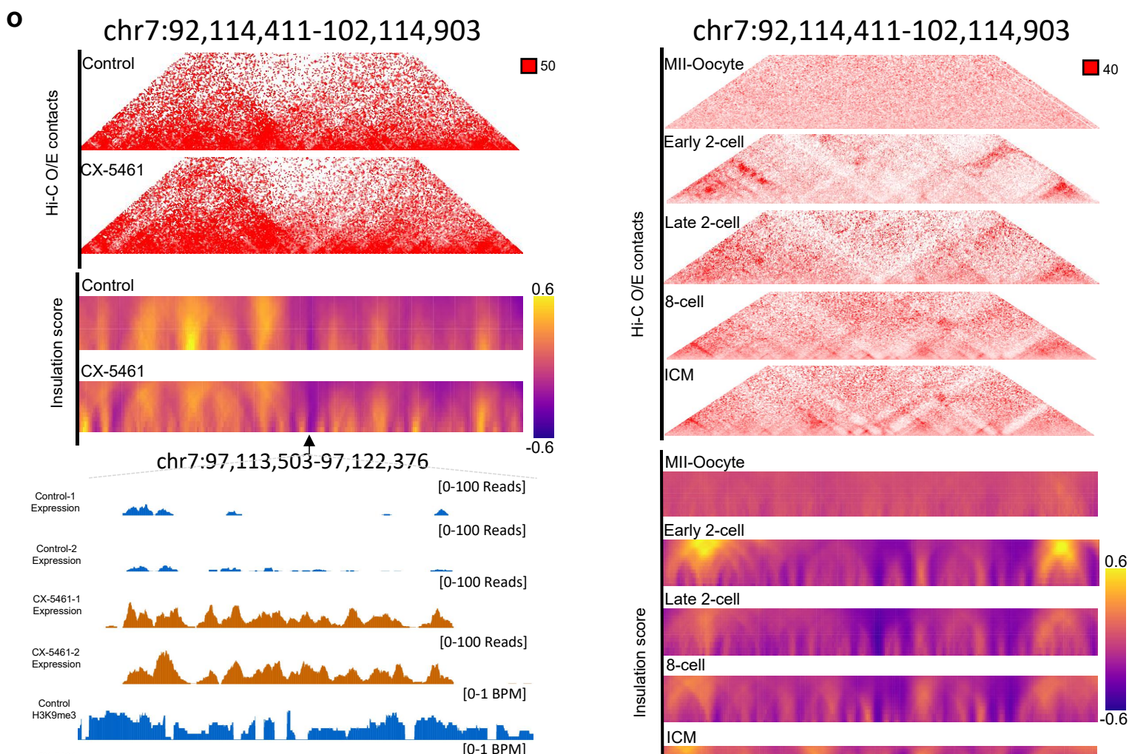


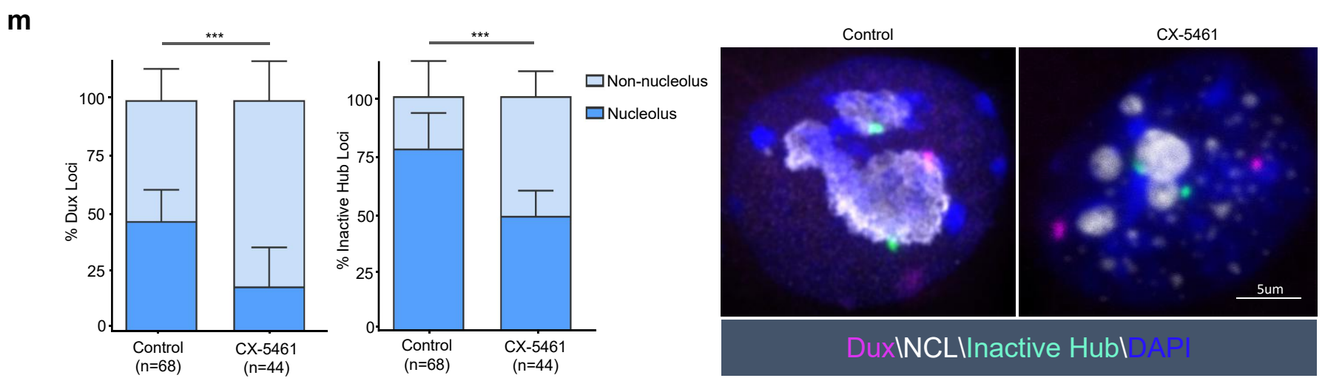












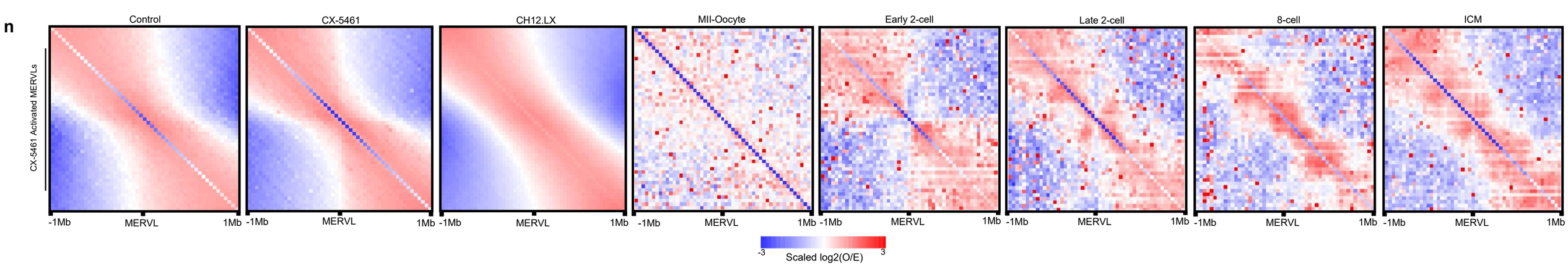


Fig.4: rRNA biogenesis defect drove 3D chromatin structure reorganization of PNH and *MERVL* regions towards the 2C-like state. a) Hi-C contact maps of Inactive Hub and 1.5 Mb genomic regions around *Dux* at 150kb resolution (GEO accession GSE166041). **b)** Hi-C contact maps of NAD and 1.5 Mb genomic regions around *Dux* at 150kb resolution (GEO accession GSE166041). **c)** Hi-C contact maps of L1 and 1.5 Mb genomic regions around *Dux* at 150kb 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 GSE166041). **d)** Hi-C pearson correlation coefficient (PCC) heat maps of Inactive Hub and 1.5Mb genomic regions around *Dux* at 150kb resolution (GEO accession GSE166041). **e)** Hi-C pearson correlation heat maps of NAD and 1.5Mb genomic regions around *Dux* at 150kb resolution (GEO accession GSE166041). **f)** Hi-C pearson correlation heat maps of L1 and 1.5Mb genomic regions around *Dux* at 150kb 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 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 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 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 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 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 GSE166041). **l)** Scatter plot demonstrates the PCC difference between L1 and different types of genes in control and CX-5461 treated mES cells. The difference of PCC is defined as the average (PCC) of Inactive Hub regions and different types of genes in CX-5461 treated mES cells minus the average PCC of Inactive Hub regions and different types of genes in wild type mES cells. PCC: Pearson Correlation Coefficient, *MERVL-int*: Up-regulated *MERVL-int* genes, *MT2_Mm*: Up-regulated *MT2_Mm* genes, UG: Up-regulated genes, DG: Down-regulated genes (GEO accession 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. ***: p<0.001, two-way ANOVA, error bar: standard error of the mean, n denotes the number of observed mES cells (GEO accession GSE166041). **n)** Aggregate Observed(O)/Expected(E) Hi-C matrices centered on CX-5461 induced *MERVL* genes in control ,CX-5461 treated mES cells (GEO accession GSE166041) and lymphoblastoid cells (GEO accession GSE63525) and mouse embryos throughout mouse pre-implantation embryonic development (GEO accession GSE82185). **o)** Representative 40kb 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 GSE166041 and GSE82185).

Extended Data Fig.S4: rRNA biogenesis defect drove 3D chromatin structure reorganization of PNH and MERVL regions towards the 2C-like state, related to Fig.4. a) Hi-C contact maps of Inactive Hub/NAD/L1 and 1.5 Mb genomic regions around *Dux* at 150kb resolution during mouse pre-implantation embryos development. The zoomed-in regions aim to demonstrate the change of Hi-C contacts between *Dux* and chromosome 10 during mouse pre-implantation embryos development (GEO accession GSE82185). **b)** Scatter plot demonstrates the $\log_2$(fold change) of Hi-C contacts between Inactive Hub/NAD/L1 and different types of genes in early 2-cell and ICM stage embryos (GEO accession GSE82185). **c)** Hi-C pearson correlation heat maps of Inactive Hub/NAD/L1 and 1.5 Mb genomic regions around *Dux* at 150kb resolution during mouse pre-implantation embryos development. The zoomed-in regions aim to demonstrate the change of Hi-C PCC between *Dux* and chromosome 10 during mouse pre-implantation embryos development (GEO accession GSE82185). **d)** Scatter plot demonstrates the PCC difference between Inactive Hub/NAD/L1 and different types of genes in early 2-cell and ICM stage embryos (GEO accession GSE82185). **e)** A/B interaction profile showing contact enrichment between active and inactive compartments (GEO accession GSE166041 and GSE63525). **f)** Quantification of compartment strength; *: p<0.05, ***: p<0.001, Wilcox signed rank test. **g)** Pearson correlation heat maps for chromosome 16 at 500kb resolution to demonstrate A/B compartment (GEO accession GSE166041 and GSE63525). **h)** Hi-C contact maps for chromosome 16 at 500kb resolution for A/B compartment profile. **i)** Observed/Expected (O/E) aggregate plot of TADs (GEO accession GSE166041 and GSE63525). **j)** Quantification of TAD strength; *: p<0.05, ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041 and GSE63525). **k)** O/E aggregate plots of chromatin loops (GEO accession GSE166041 and GSE63525). **l)** Quantification of loop strength; ***: p<0.001, Wilcox signed rank test (GEO accession GSE166041 and GSE63525). **m)** Hi-C contact maps for chromosome 16:27.3-36.2mb region at 40kb resolution to demonstrate TAD and chromatin loop (GEO accession GSE166041 and GSE63525). **n)** Insulation score, expression (TPM), ATAC-seq (BPM), H3K9me3 (BPM), H3K27me3 (BPM), H3K4me3 (BPM) and H3K27ac (BPM) signals of control and CX-5461-treated mES cells centered on CX-5461-induced *MERVL* genes (GEO accession GSE166041). **o)** Representative 40kb Hi-C O/E interaction matrices of a *MERVL* loci (chr15:76,145,960-86,151,258) located at TAD boundaries are shown as heatmaps, along with 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 GSE166041).

Fig.5

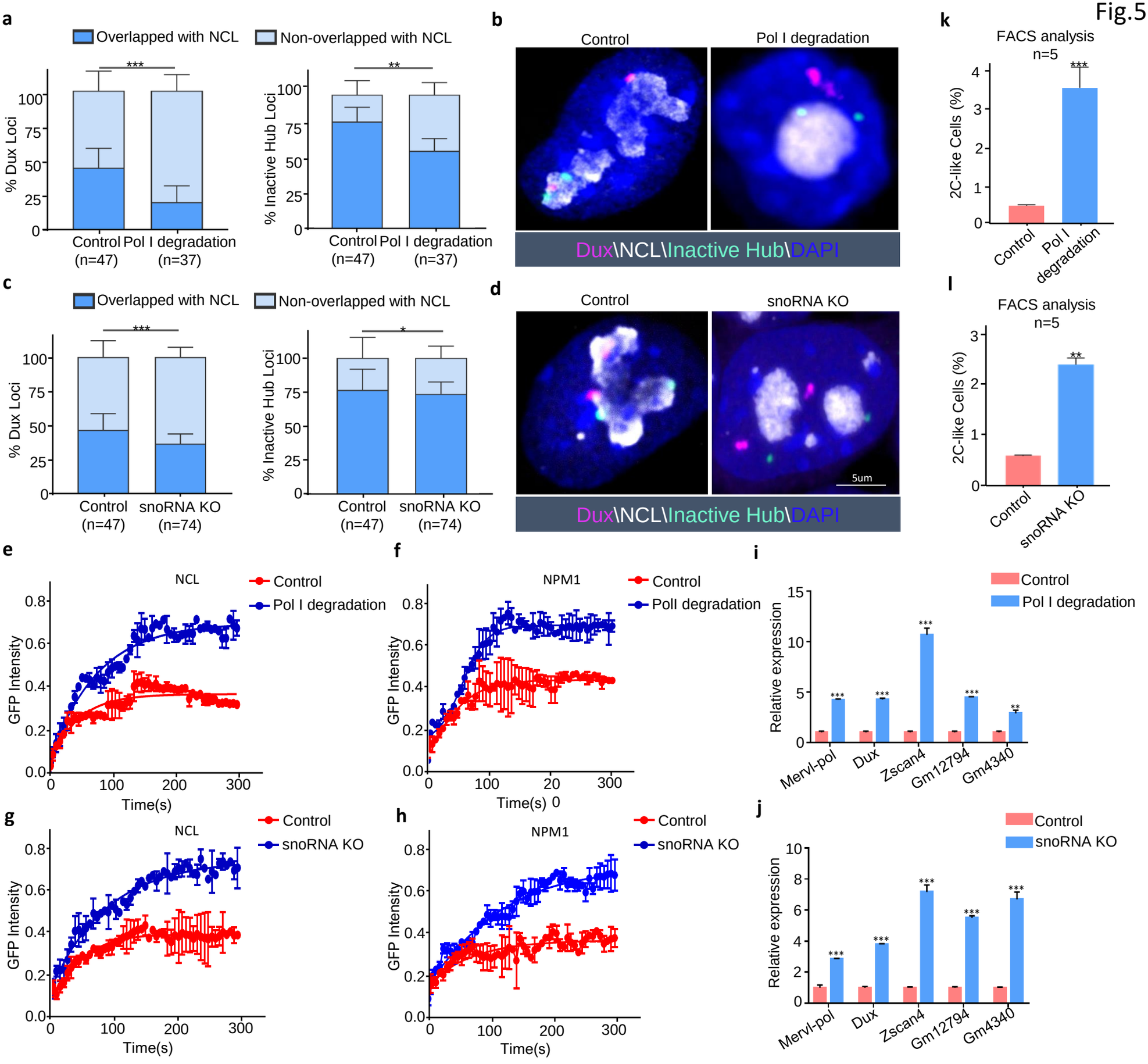
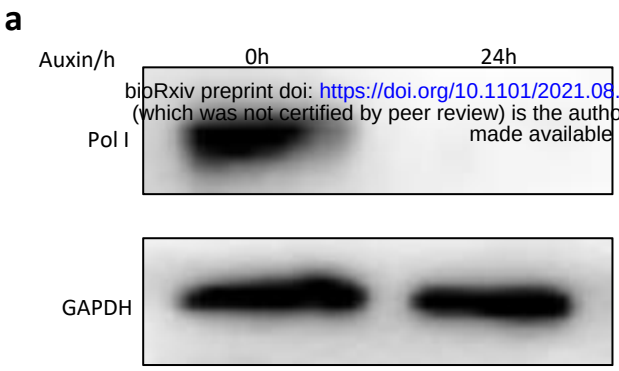
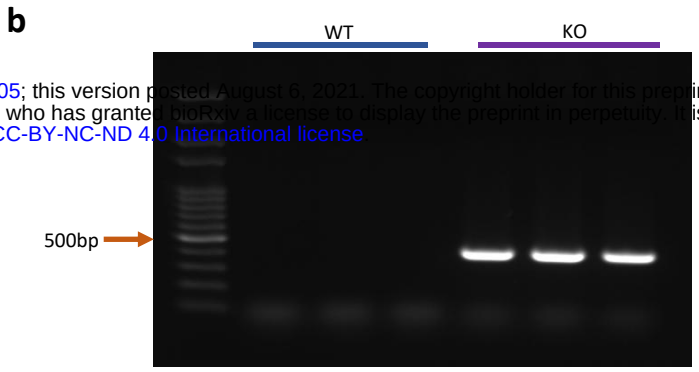
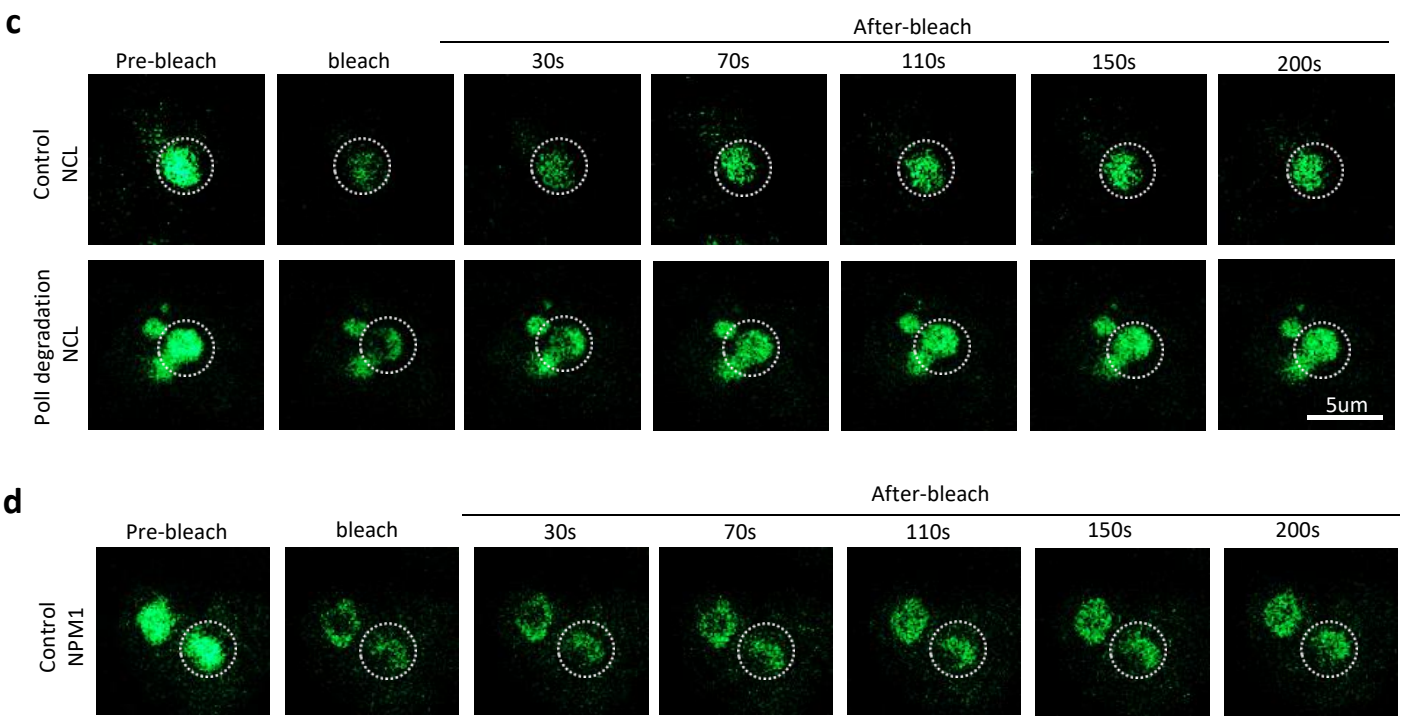
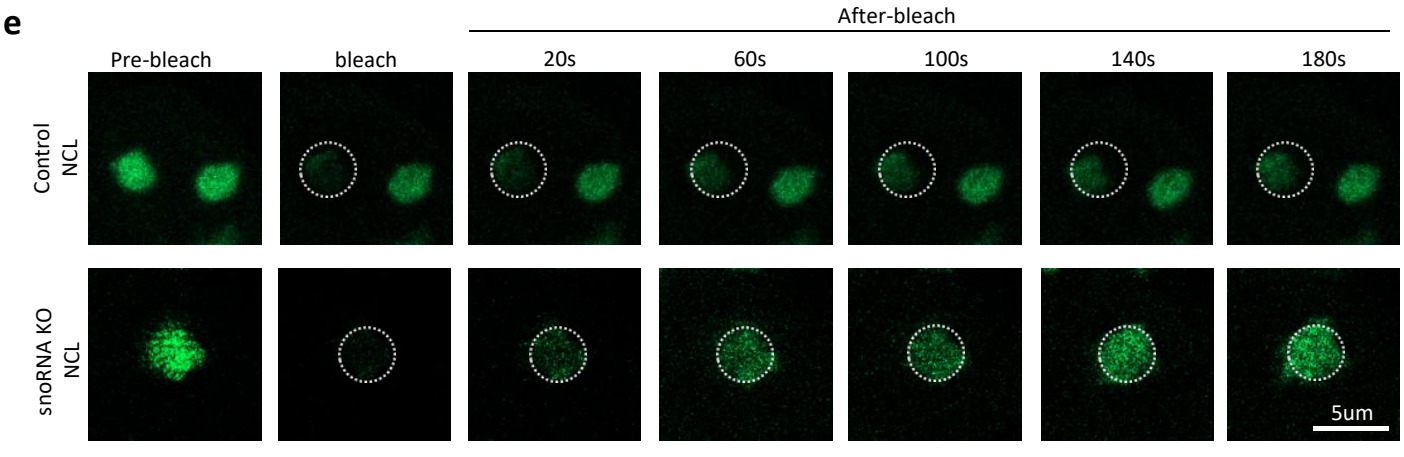

a
Overlapped with NCL
Non-overlapped with NCL
% Dux Loci
% Inactive Hub Loci
Control (n=47)
Pol I degradation (n=37)
b
Control
Pol I degradation
Dux\NCL\Inactive Hub\DAPI
k
FACS analysis
n=5
2C-like Cells (%)
Control
Pol I degradation
c
snoRNA KO (n=74)
d
snoRNA KO
5um
l
e
NCL
Control
Pol I degradation
GFP Intensity
Time(s)
f
NPM1
Poll degradation
i
Relative expression
Mervl-pol
Dux
Zscan4
Gm12794
Gm4340
g
snoRNA KO
h
j

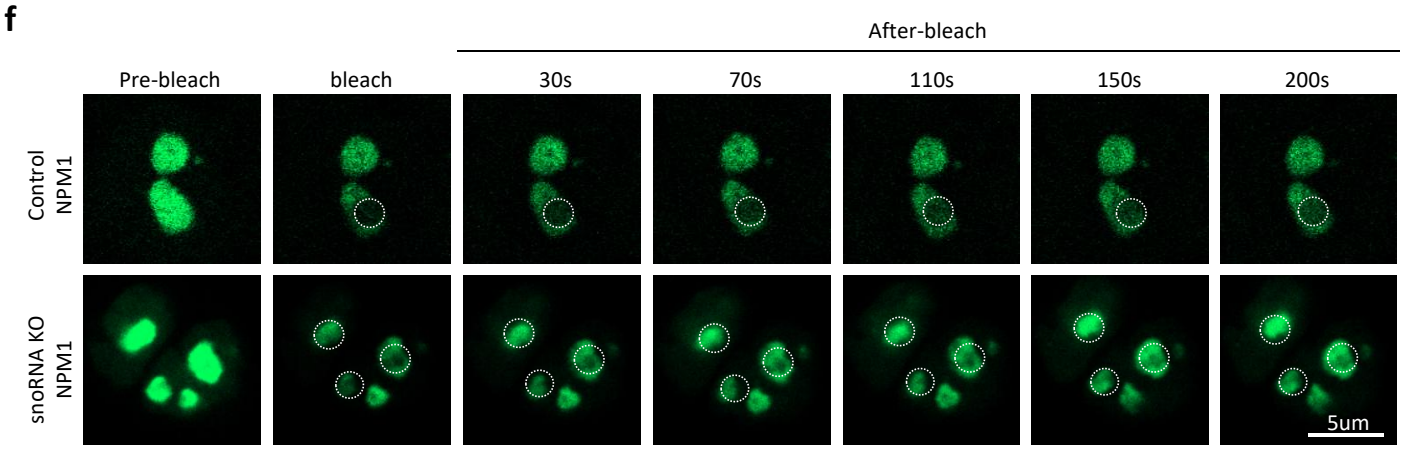
Fig.5: Genetic interferences of rRNA biogenesis recapitulate CX-5461-induced 2C-like molecular phenotypes. a) The percentage of Nucleolus-localized (overlapped with NCL) and Nucleoplasm-localized (nonoverlapped with NCL) of FISH signals in control and Pol I degradation mES cell lines; **: p<0.01, ***: p<0.001, two-way ANOVA, error bar: standard error of the mean, n denotes the number of observed mES cells. **b)** Representative images demonstrating DNA FISH analysis with a *Dux* locus probe and Inactive Hub locus probe, and co-immunostained with NCL protein in control and Pol I degradation mES cell lines. **c)** The percentage of Nucleolus-localized (overlapped with NCL) and Nucleoplasm-localized (nonoverlapped with NCL) of FISH signals in control and snoRNA knockout mES cell lines; *: p<0.05, ***: p<0.001, two-way ANOVA, error bar: standard error of the mean, n denotes the number of observed mES cells. **d)** Representative images demonstrating 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. **e)** FRAP analysis showing Pol I degradation causes accelerated recovery after photobleaching of NCL, the replicates of experiment n = 4. **f)** FRAP analysis showing Pol I degradation causes accelerated recovery after photobleaching of NPM1, the replicates of experiment n = 4. **g)** FRAP analysis showing snoRNA knockout causes accelerated recovery after photobleaching of NCL, the replicates of experiment n = 4. **h)** FRAP analysis showing snoRNA knockout causes accelerated recovery after photobleaching of NPM1, the replicates of experiment n = 4. **i)** qRT-PCR quantification of 2C marker gene expression in control mES cells and Pol I degraded mES cell lines; **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3; error bar: standard error of the mean. **j)** qRT-PCR quantification of 2C marker gene expression in control mES cells and snoRNA knockout mES cell lines; **: p<0.01, ***: p<0.001, two-way ANOVA, the replicates of experiment n=3; error bar: standard error of the mean. **k)** The percentage of 2C::tdTomato positive cells was quantified using FACS analysis in control mES cells and Pol I degraded mES cells; Data are means ± SD, SD: Standard Deviation, ***: p<0.001, two-way ANOVA, the replicates of experiment n=5. **l)** The percentage of 2C::tdTomato positive cells was quantified using FACS analysis in control mES cells and snoRNA knockout mES cells; Data are means ± SD, SD: Standard Deviation, **: p<0.01, two-way ANOVA, the replicates of experiment n=5.

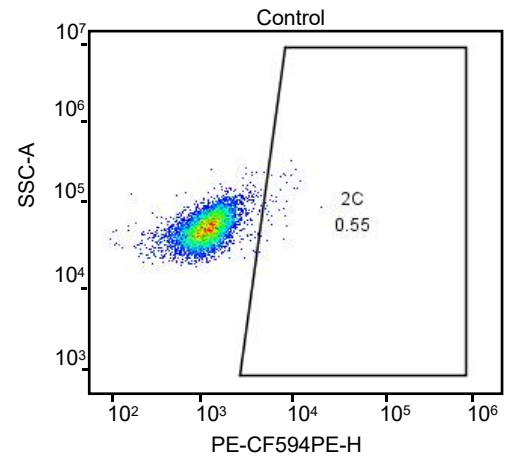


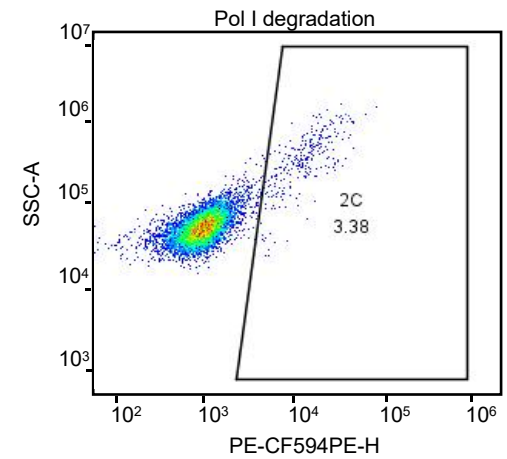


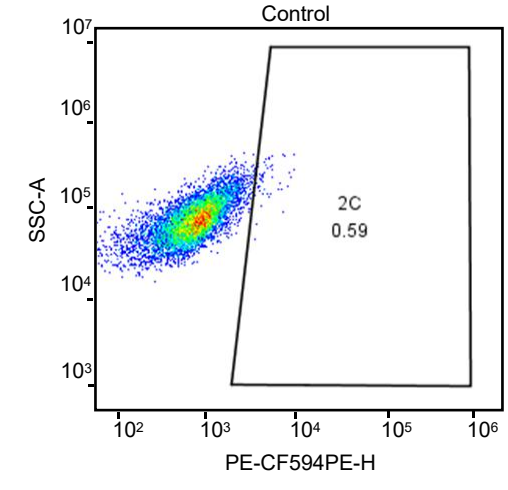


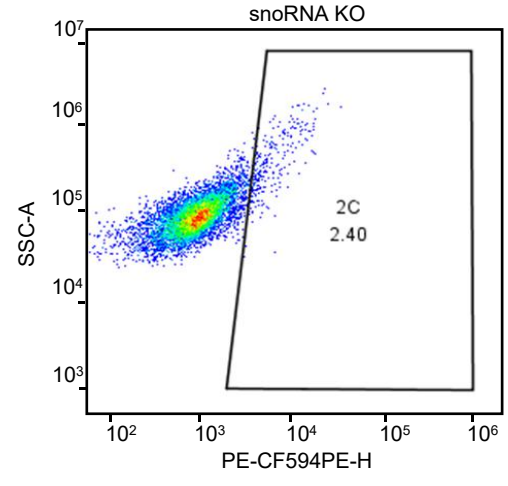












Extended Data Fig.S5: Genetic interferences of rRNA biogenesis recapitulate CX-5461-induced 2C-like molecular phenotypes, related to Fig.5. a) Western Blotting experiment showing the Pol I protein degradation after 24h of Auxin treatment. **b)** PCR experiment showing that a 400bp band was observed in the snoRNA KO mES cells, but not in the wild-type (WT) mES cells. As a band of 400bp was designed especially in the snoRNA KO mES cells, this result indicates that the homologs of human SNORD113-114 gene cluster was successfully knocked-out. **c)** Shown images are representative of 4 times of NCL FRAP experiments in control mES cells and Pol I degraded mES cells. **d)** Shown images are representative of 4 times of of NPM1 FRAP experiments in control mES cells and Pol I degraded mES cells. **e)** Shown images are representative of 4 times of NCL FRAP experiments in control mES cells and snoRNA knockout mES cells. **f)** Shown images are representative of 4 times of NPM1 FRAP experiments in control mES cells and snoRNA knockout mES cells. **g)** FACS analysis on 2C::tdTomato+ mES cells in Pol I degraded mES cell lines, showing the change of percentage of 2C-like cells**. h)** FACS analysis on 2C::tdTomato+ mES cells in snoRNA knockout mES cell lines, showing the change of percentage of 2C-like cells**.**

Fig.6

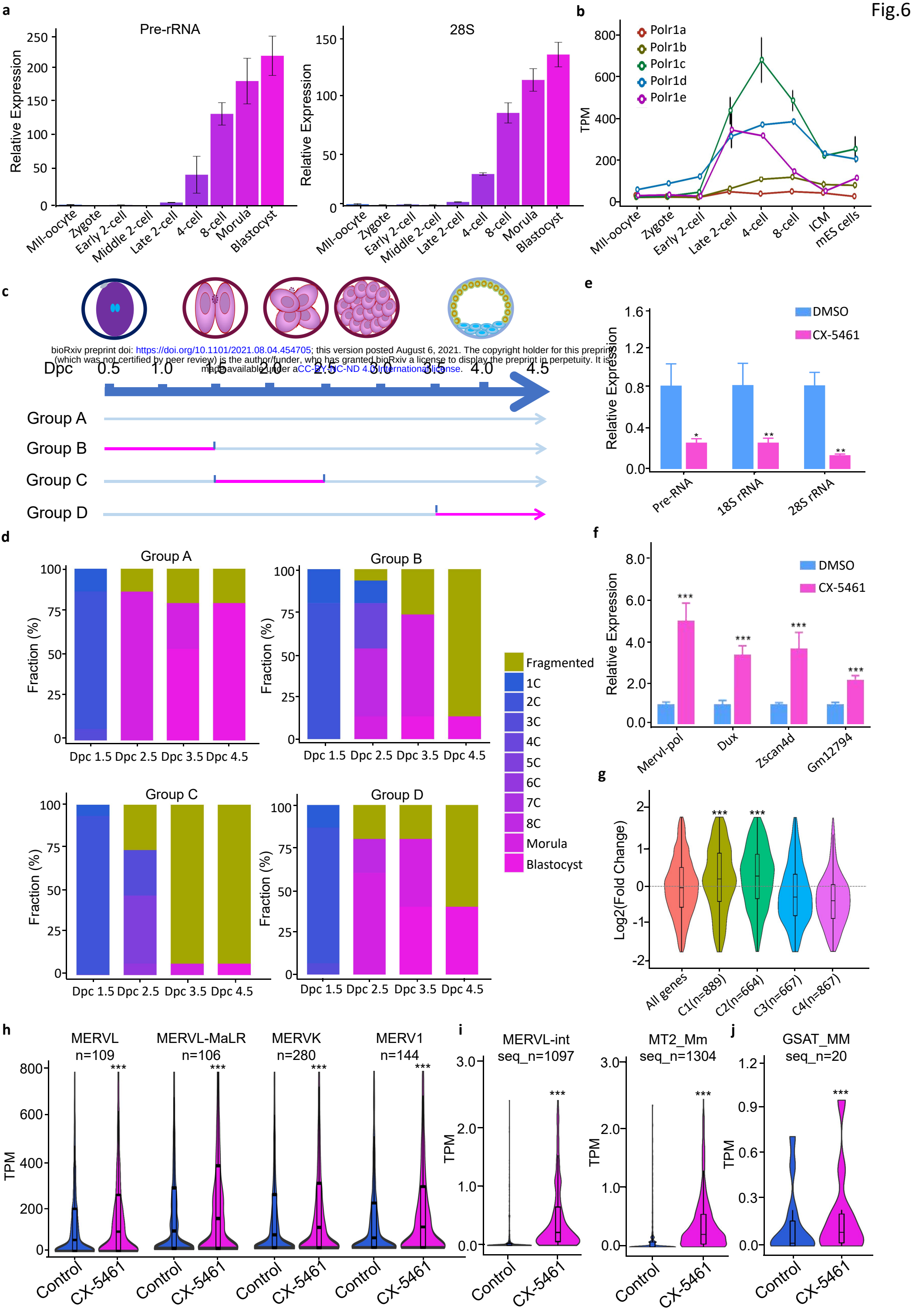
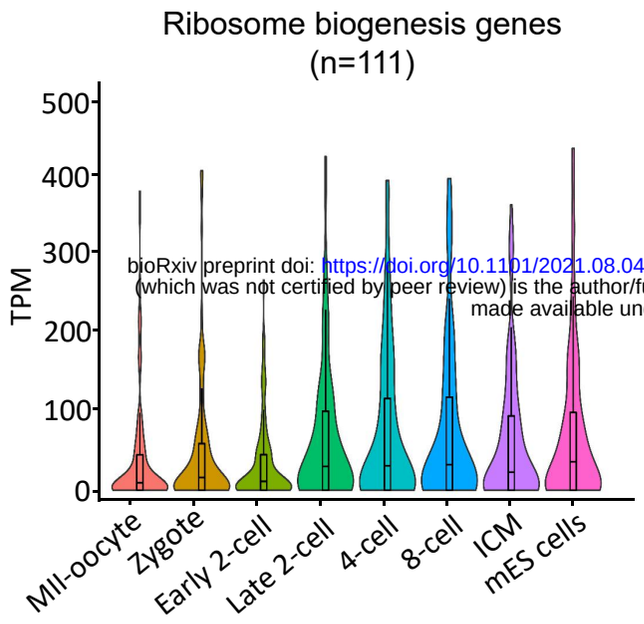
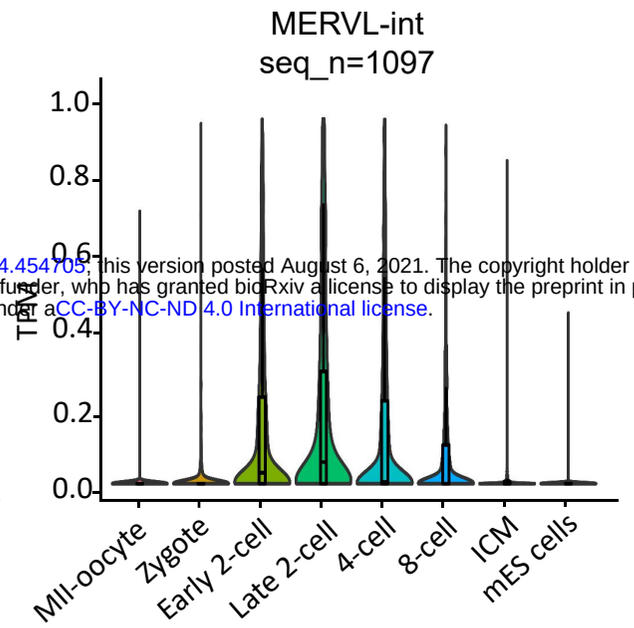


Fig.6: rRNA biogenesis is critically required at the 2-cell-to-4-cell stage transition during pre-implantation embryo development. a) Expression of pre-rRNA and 28S rRNA across different embryo developmental stages. **b)** Expression of different subunit genes of RNA polymerase I across different embryo developmental stages. **c)** Different schemes of treatment with CX-5461. The 24hrs 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. *: $p<0.05$, **: $p<0.01$, two-way ANOVA, the replicates of experiment n=3. **f)** qRT-PCR showing 2C marker gene expression level in blastocysts, after CX-5461 treatment of morula embryos followed by *in vitro* culture of the treated embryos. ***: $p<0.001$, two-way ANOVA, the replicates of experiment n=3. **g)** Violin plots demonstrating the expression level changes of stage-specific gene clusters of mouse pre-implantation embryos (as defined in Extended Data Fig S1) in CX-5461-treated and control blastocyst embryos, ***: $p<0.001$, Mann-Whitney U test (GEO accession GSE166041). **h)** Violin plots show expression levels of major ERV gene classes in control blastocyst embryos and CX-5461 treated blastocyst embryos; n denotes the number of sub-classes of ERV genes; ***: $p<0.001$, Wilcox signed rank test (GEO accession GSE166041). **i)** Violin plots show expression levels of ERV gene sub-classes of *MERVL-int* and *MT2_Mm* in control blastocyst embryos and CX-5461 treated blastocyst embryos; seq_n denotes the number of annotated *MERVL-int* and *MT2_Mm* sequences in the mouse mm10 reference genome; *: ***: $p<0.001$, Wilcox signed rank test (GEO accession GSE166041). **j)** 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 *GSAT_MM* sequences in the mouse mm10 reference genome; ***: $p<0.001$, Wilcox signed rank test (GEO accession GSE166041).

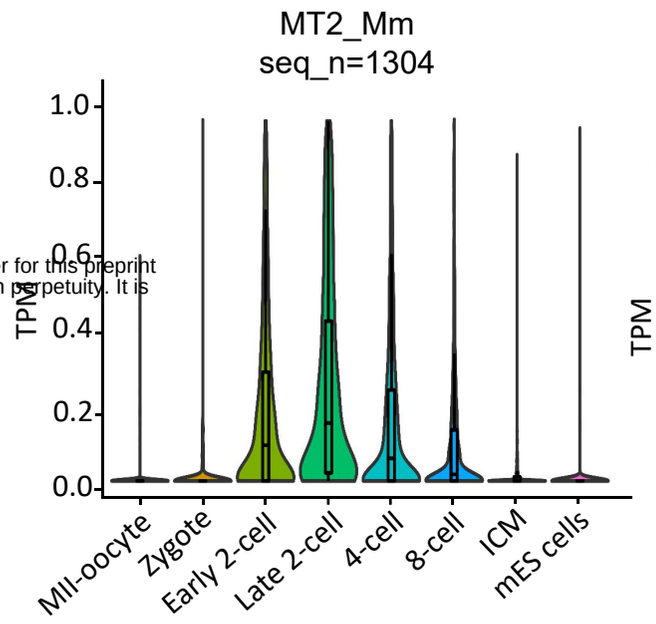
a



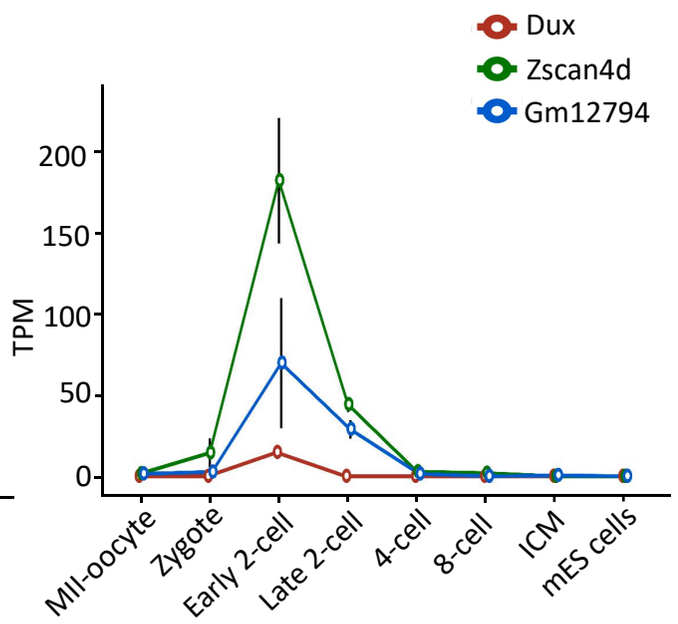
b



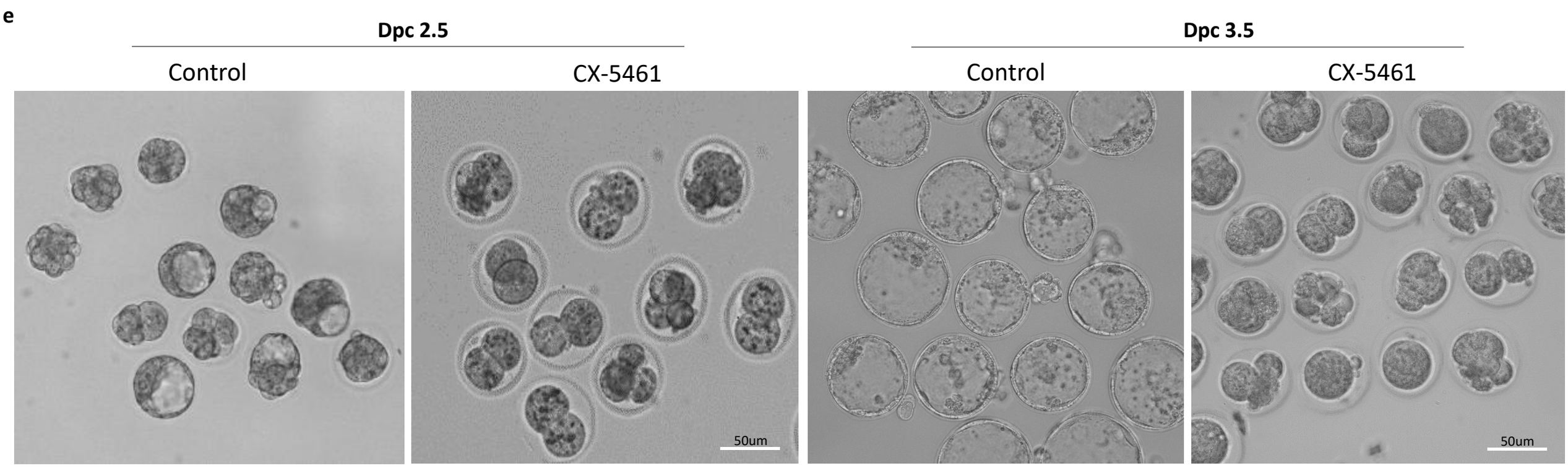
c



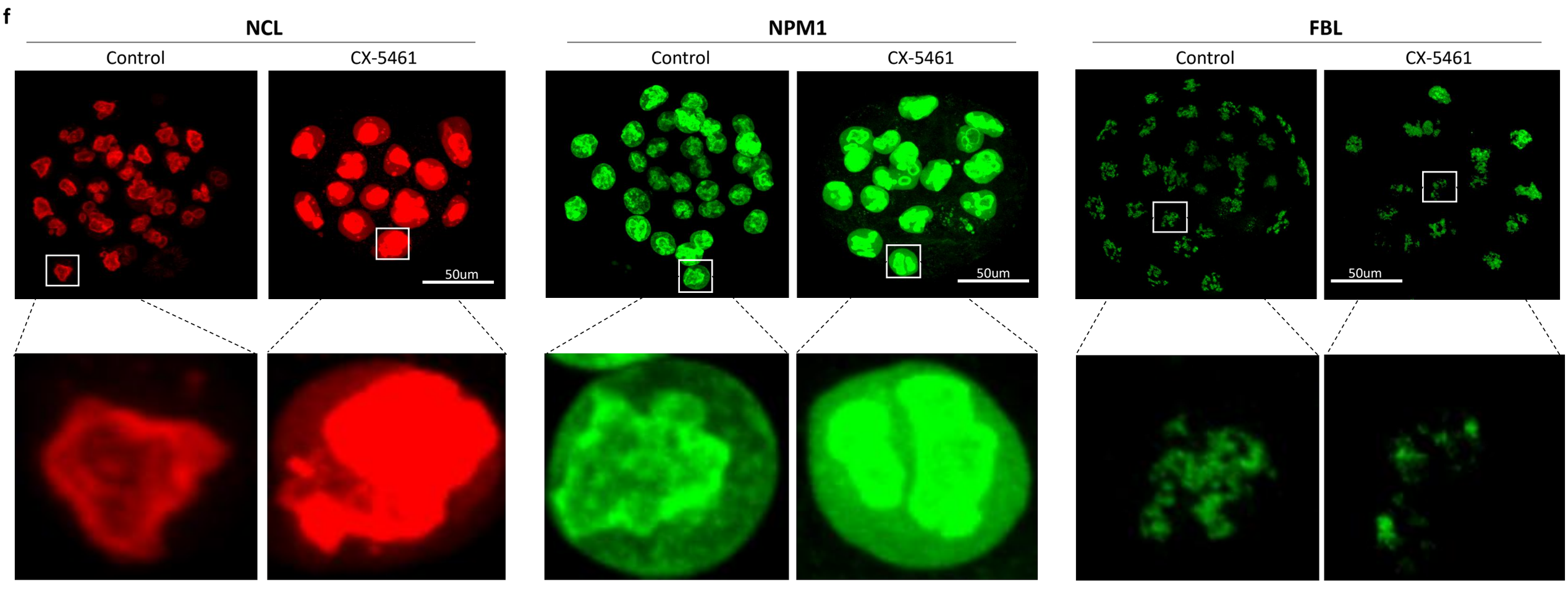
d



e



f



Extended Data Fig.S6: rRNA biogenesis is critically required at the 2-cell-to-4-cell stage transition during pre-implantation embryo development, related to Fig 6. a) Expression pattern of Ribosome biogenesis genes across different embryo developmental stages. n denotes the number of sub-classes of *MERVL* genes. **b)** Expression pattern of *MERVL-int* genes across different embryo developmental stages. seq_n denotes the number of annotated *MERVL-int* sequences in the mouse mm10 reference genome. **c)** Expression pattern of *MT2_Mm* genes across different embryo developmental stages. seq_n denotes the number of annotated *MT2_Mm* sequences in the mouse mm10 reference genome. **d)** Expression pattern of 2C marker genes, *Dux*, *Zscan4d* and *Gm12794*, across different embryo developmental stages. **e)** Representative images of mouse embryos produced from control and CX-5461 treatment during two different developmental stages; Dpc: Days post-coitum. This experiment was repeated three times independently with similar results. **f)** Immunofluorescence staining of NCL, NPM1 and FBL in control blastocyst embryos and CX-5461-treated blastocyst embryos.

Fig.7

rRNA

Dux/MERVL

ES program

2C-like program

mES state

2C-like state

WT Condition

Dux repressed

3D structure of PNH

3D structure of MERVL

Nucleolus

-Normal nucleolar phase separation
-Normal 3D chromatin of PNH and MERVL regions
-Dux repressed by PNH region

rRNA Inhibition

Dux activated

3D structure of PNH

3D structure of MERVL

Nucleolus

-Disrupted nucleolar phase separation
-Reshaped 3D structure of PNH and MERVL regions
-Dux released from PNH region

rRNA

PNH

FC

DFC

GC

NCL

TRIM28

rRNA biogenesis

3D chromatin structure
reorganization of PNH region

Blastocyst

2-cell

Fig.7: A mechanistic model for the role of rRNA biogenesis in regulating the 2C-like program and the homeostasis between 2C-like cells and mES cells. In the unperturbed mES cells, nucleolar integrity mediated by rRNA biogenesis maintains the normal the liquid-liquid phase separation (LLPS) of nucleolus and the formation of peri-nucleolar heterochromatin (PNH) containing *Dux*, and this normal nucleolar LLPS facilitated NCL/TRIM28 complex occupancy on the *Dux* locus to repress *Dux* expression. In contrast, in the rRNA biogenesis-inhibited mES cells, the liquid-like phase of nucleolus is disrupted, causing dissociation of the NCL/TRIM28 complex from the PNH and changes of epigenetic state and 3D structure of the PNH, which eventually leads to *Dux* to be released from the PNH, activation of 2C-like program and transition of mES cells to 2C-like cells.