SMARCAD1 and TOPBP1 contribute to heterochromatin maintenance at the transition from the 2C-like to the pluripotent state

Running title: The chromatin-bound proteome in 2C-like cells

Ruben Sebastian-Perez[1,#], Shoma Nakagawa[1,#], Xiaochuan Tu[1,#], Sergi Aranda[1], Martina Pesaresi[1], Pablo Aurelio Gomez-Garcia[1], Marc Alcoverro-Bertran[1], Jose Luis Gomez-Vazquez[1], Davide Carnevali[1], Eva Borràs[1,2], Eduard Sabidó[1,2], Laura Martin[1], Malka Nissim-Rafinia[3], Eran Meshorer[3,4], Maria Victoria Neguembor[1], Luciano Di Croce[1,2,5], Maria Pia Cosma[1,2,5,6,7,*]

Affiliations

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain.

[2]Universitat Pompeu Fabra (UPF), Dr. Aiguader 88, Barcelona 08003, Spain.

[3]Department of Genetics, The Alexander Silberman Institute of Life Sciences, Edmond J. Safra Campus, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.

[4]The Edmond and Lily Safra Center for Brain Sciences, Edmond J. Safra Campus, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.

[5]ICREA, Pg. Lluis Companys 23, Barcelona 08010, Spain.

[6]Bioland Laboratory, Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou 510005, China.

[7]CAS Key Laboratory of Regenerative Biology, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China.

Co-first authors

* Corresponding author. Email: pia.cosma@crg.es (M.P.C.)

Abstract

Chromocenters are established after the 2-cell (2C) stage during mouse embryonic development, but the factors that mediate chromocenter formation remain largely unknown. To identify regulators of 2C heterochromatin establishment, we generated an inducible system to convert embryonic stem cells (ESCs) to 2C-like cells. This conversion is marked by a global reorganization and dispersion of H3K9me3-heterochromatin foci, which are then reversibly formed upon re-entry into pluripotency. Profiling the chromatin-bound proteome (chromatome) by genome capture of ESCs transitioning to 2C-like cells, we uncover chromatin regulators involved in *de novo* heterochromatin formation. We identified TOPBP1 and investigated its binding partner SMARCAD1. SMARCAD1 and TOPBP1 associate with H3K9me3-heterochromatin in ESCs. Interestingly, the nuclear localization of SMARCAD1 is lost in 2C-like cells. SMARCAD1 or TOPBP1 depletion in mouse embryos lead to developmental arrest, reduction of H3K9me3 and remodeling of heterochromatin foci. Collectively, our findings contribute to comprehending the maintenance of chromocenters during early development.

Introduction

Early mammalian development is a dynamic process that involves large-scale chromatin reorganization (*1*). Blastomeres acquire a defined cell identity through the activation of a subset of genes and specific epigenetic modifications. Among the latter, there is a tailored control over histone H3 lysine 9 trimethylation (H3K9me3), a hallmark of the transcriptionally repressed constitutive heterochromatin (*2, 3*). During the first cleavage stages, constitutive heterochromatin reorganizes in the nucleus to form highly compacted chromocenters (*4, 5*). Systematic identification of the underlying factors involved in *de novo* heterochromatin establishment and maintenance - thus chromocenter compaction- is still lacking, mostly because of the minuscule amount of material available during embryogenesis.

Embryonic stem cells (ESCs) can fluctuate back to a 2-cell embryo-like (2C-like) state under defined culture conditions (*6*). Although ESCs can spontaneously revert their fate to resemble early embryogenesis, this process happens at a very low frequency (*6*). Recently, early mouse embryo development has been modelled with high efficiency after downregulation of chromatin assembly factors (*7*) or modulation of key developmentally regulated genes in ESCs (*8-11*). ESCs can efficiently be converted into 2C-like cells by the overexpression of a single murine transcription factor, Dux (*11*). Interestingly, decondensation of HP1α foci and of chromocenters in $2C^+$ were previously reported, suggesting that $2C^+$ cells can be used to study the remodeling of heterochromatin foci (*7, 12*). Here, using the Dux-dependent reprogramming system, we show that 2C-like cells can be used as a model system to investigate *de novo* chromocenter formation and dynamics. Using chromatin proteomics, we profiled the dynamic changes occurring in the chromatin-bound proteome (chromatome) during 2C-like cell reprogramming and identified factors potentially involved in chromocenter reorganization. H3K9me3-marked heterochromatin foci in 2C-like cells generated via Dux overexpression became larger and decreased in number during the reprogramming of ESCs to 2C-like cells. The chromocenters re-formed upon transition of 2C-like cells into ESC-like cells. We identified the DNA TOPoisomerase II Binding Protein 1 (TOPBP1) and the chromatin remodeler SWI/SNF-Related, Matrix-Associated Actin-Dependent Regulator Of Chromatin, Subfamily A, Containing DEAD/H Box 1 (SMARCAD1) to be associated with H3K9me3 in heterochromatin foci of ESCs. The association of SMARCAD1 was reduced upon entry of ESCs in the 2C-like state, although SMARCAD1 nuclear localization was recovered after 2C-like state exit. Depletion of SMARCAD1 and of TOPBP1 induced mouse embryo developmental arrest, which was accompanied by a remodeling of the heterochromatin foci. Our results suggest a contributing role of SMARCAD1 and of TOPBP1 activity in the maintenance of heterochromatin formation during early development.

Results

Chromatin-bound proteome profiling allows the identification of dynamic chromatome changes during 2C-like cell reprogramming

To explore the molecular driving events for the establishment of constitutive heterochromatin during embryo development, we generated stable ESC lines carrying doxycycline-inducible cassettes that drive expression of either Dux (Dux-codon altered, CA) or luciferase (control) (Fig. 1A). These ESC lines also carry an EGFP reporter under the control of the endogenous retroviral element MERVL long terminal repeat (2C::EGFP) (*7*). The EGFP reporter allows the purification of 2C-like cells (hereinafter named $2C^+$) and low *Dux* expressing cells, which are negative for MERVL reporter expression ($2C^-$) (Fig. 1A). $2C^-$ cells were reported to be an intermediate population generated during 2C-like reprogramming (*13, 14*). During the reprogramming process, in contrast to $2C^-$ cells, $2C^+$ cells do not show DAPI-dense chromocenters (*7, 11*). Therefore, we can study *de novo* chromocenter formation by following the transition of $2C^+$ cells toward an ESC-like state, thus modelling in culture the epigenetic reprogramming that occurs during mouse early development.

After culturing the Dux-CA line with doxycycline (Dox), the number of 2C-like cells increased to > 60 % as compared to luciferase control cells (fig. S1A-D). Dux overexpression resulted in the loss of DAPI-dense chromocenters and loss of the pluripotency transcription factor OCT4 (fig. S1E), in accordance with previous reports (*6, 7*). These changes were accompanied by an upregulation of specific genes of the 2-cell transcriptional program such as endogenous *Dux*, MERVL and major satellites (MajSat) (fig. S1F). Additionally, we looked at cell cycle progression in the heterogeneous population of cells generated after Dux overexpression since it has been previously shown that spontaneous 2C-like cells have an altered cell cycle (*15*). $2C^-$ cells displayed a cell cycle profile comparable to that of luciferase cells, whereas $2C^+$ cells accumulated in the G2/M cell cycle phase (fig. S1G) with a much-reduced S phase consistent in several clonal lines (fig. S1H). Overall, these data indicate that the $2C^+$ line we generated recapitulates known features of 2C-like cells.

Having characterized the Dux-CA line, we aimed to identify potential chromatin-associated factors involved in the *de novo* establishment of heterochromatin. For that, we performed DNA-mediated chromatin purification coupled to tandem mass spectrometry for the identification of proteins on total DNA (iPOTD) (*16, 17*). We captured the whole genome labelled with 5-ethynyl-2'-deoxyuridine (EdU) and identified candidate proteins differentially enriched in the 2C-like chromatin-bound (chromatome) fraction (Fig. 1A). We analyzed the chromatome of $2C^+$, $2C^-$ and luciferase (Luc) populations to characterize the chromatin-bound proteome profile of these distinct states. We first confirmed that we could enrich the iPOTD preparations for chromatin proteins, such as histone H3, and devoid them of cytoplasmic ones, such as vinculin (fig. S2A-C). We identified a total of 2396 proteins, suggesting an efficient pull-down of chromatin-associated factors (fig. S2D and Table S1).

Chromatin-resident proteins, such as core histones and histone variants, were comparably enriched in all +EdU replicates (fig. S2E and Table S1). Pearson's correlation coefficients (PCC) and principal component analysis (PCA) of independent replicates of $2C^+$, $2C^-$ and Luc samples showed consistent results regarding the abundance of the proteins detected (Fig. 1B-D and Table S1). Interestingly, Luc replicates clustered separately from $2C^+$ and $2C^-$ conditions, indicating significant changes in the chromatomes of these fractions (Fig. 1C, 1D).

We then ranked the identified chromatin-associated factors according to their fold change to interrogate the differences in protein-chromatin interactions in the $2C^+$, $2C^-$ and Luc chromatomes (Fig. 1E-G). Members of the ZSCAN4 (Zinc finger and SCAN domain containing 4) family of proteins, that are well-characterized markers of the 2C stage (*7, 11, 18, 19*), were identified among the top enriched factors in the $2C^+$ chromatome (Fig. 1E, 1F, 1H). ZSCAN4 family members, such as ZSCAN4F and ZSCAN4C, were found associated with chromatin already in the $2C^-$ chromatome (Fig. 1F), supporting previous findings (*14*). However, we identified regulators of 2C-like cells such as TET1, the non-canonical Polycomb (PcG) Repressor Complex 1 (PRC1) member PCGF6, the TGF-β regulator SMAD7 and the heterochromatic H4K20me3 methyltransferase SUV420H2 depleted from the $2C^+$ chromatome when compared to the $2C^-$ (*13, 20, 21*) (Fig. 1E, 1G, 1H). The pluripotency transcription factors NANOG, OCT4, STAT3, SOX2 and ESRRB were, as expected, exclusively enriched in the $2C^-$ and Luc chromatomes (Fig. 1H). We also identified several transcriptional regulators and epigenetic enzymes differentially enriched in the $2C^+$, $2C^-$ and Luc chromatomes (Fig. 1H). Of note, we identified marked differences in the enrichment for H3K9 histone methyltransferases SUV39H1 and SUV39H2 (Fig. 1H). SUV39H2 gradually increased its abundance on chromatin as Luc cells converted to $2C^-$ and, ultimately, to $2C^+$ (Fig. 1H). Contrarily, SUV39H1 was overall less abundantly chromatin-bound, although with a slight enrichment in the $2C^-$ chromatome (Fig. 1H). Altogether, these data indicate that ESC reprogramming toward 2C-like state correlates with a major reorganization of the chromatin-bound proteome.

We used the Significance Analysis of INTeractome (SAINT) algorithm (*22*) to further interrogate protein-chromatin interactions in the iPOTD datasets. To identify molecular drivers of chromocenter reorganization, we compared the enriched proteins in the $2C^+$, $2C^-$ and Luc chromatomes (Fig. 1I and Table S1). We identified a total of 397 proteins shared by the $2C^-$ and Luc chromatomes that were not enriched in the $2C^+$ chromatome (Fig. 1I and Table S1). We focused on analyzing this cluster since chromocenters are present in $2C^-$ and Luc cells. This protein cluster included gene ontology (GO) terms associated with RNA and chromatin binding, active remodeling activity (e.g. ATPase activity), repressive chromatin (e.g. heterochromatin condensed chromosome, negative regulation of gene expression), and pluripotent stem cell identity (e.g. response to LIF, stem cell maintenance, blastocyst growth) (Fig. 1J). To identify putative factors responsible for chromocenter reorganization, we ranked the commonly identified proteins included in $2C^-$ and Luc chromatomes according to their fold change (Fig. 1K). Notably, this protein cluster included known transcriptional regulators such as the DNA methyltransferase DNMT3L, the bromodomain-

containing protein BRD2, the core pluripotency factor OCT4 and the DNA topoisomerase 2-binding protein 1, TOPBP1 (Fig. 1K). We focused our attention on TOPBP1, since topoisomerases control genome structure and folding (*23*). We asked if the lack of topoisomerase activity could promote $2C^+$ cell induction. Thus, we treated ESCs with camptothecin (CPT) and ICRF-193, which are inhibitors of DNA topoisomerases I and II, respectively (*24, 25*). Inhibition of topoisomerase II alone increased the number of $2C^+$ cells 1.5-fold (fig. S3A) and triggered a prominent cell cycle arrest in the G2/M phase (*25, 26*) (fig. S3B). Simultaneous inhibition of topoisomerases I and II resulted in an enhanced effect, leading to a 2.4-fold increase in the fraction of $2C^+$ cells (fig. S3A, S3C). These results motivated us to further investigate TOPBP1 network. TOPBP1 has been shown to interact with chromatin remodelers such as the SWI/SNF-like remodeler SMARCAD1 in yeast and human cells (*27, 28*) (Fig. 1L). We then investigated TOPBP1 and SMARCAD1 as potential candidate factors controlling the remodeling of chromocenters.

Entry in the 2C-like state is characterized by the remodeling of H3K9me3 heterochromatic regions

To study remodeling of the chromocenters, we asked about the reorganization of heterochromatic regions upon reprogramming of ESCs into $2C^+$ cells. H3K9me3 is a well-known pericentric heterochromatin histone modification that prominently associates with constitutive heterochromatin (*2, 3, 29, 30*). H3K9me3 can therefore be used as marker for chromocenters. H3K9me3 foci in $2C^+$ cells were morphologically distinct from those of $2C^-$ cells (Fig. 2A). They were 2.3-fold fewer (3.89 ± 0.19 foci/nucleus) (Fig. 2B), and occupied 2.4-fold larger area (4.76 ± 0.33 μm^2) in $2C^+$ as compared with both ESCs (8.88 ± 0.30 foci/nucleus; 1.99 ± 0.07 μm^2) and $2C^-$ cells (8.72 ± 0.25 foci/nucleus; 1.93 ± 0.08 μm^2) (Fig. 2C). These results suggest that H3K9me3 heterochromatin undergoes massive spatial reorganization, during the reprogramming of ESCs into 2C-like state. Importantly, the levels of H3K9me3 remain unchanged among ESCs, $2C^-$ and $2C^+$ cells, indicating that the remodeling of chromocenters was not due to loss of H3K9me3 (fig. S3D). The increased size of the H3K9me3 foci and the reduction in the number of H3K9me3 foci per nucleus might be due to the decompaction or fusion of several chromocenters.

We then imaged global DNA organization with Stochastic Optical Reconstruction super-resolution Microscopy (STORM). DNA was labelled using the nucleotide analogue 5-ethynyl-2'-deoxycytidine (EdC) (*31, 32*). DNA images were quantified by Voronoi tessellation analysis (*33, 34*) which can precisely determine the DNA density based on the number of localizations in each Voronoi tessel (see Methods). Voronoi analysis showed a marked decrease in the localization density of the chromatin in $2C^+$ cells (Fig. 2D). Furthermore, Voronoi analysis confirmed the decreased DNA density as a function of the GFP intensity in $2C^+$ cells (Fig. 2E). Interestingly, $2C^-$ cells were heterogeneous with respect to DNA density, with the majority of them showing low DNA density as compared with $2C^+$ cells, suggesting that DNA might undergo decompaction prior to GFP activation (Fig.

2E). Overall, the DNA decompaction of the chromatin fibers in $2C^+$ cells is in agreement with the chromatin landscape of early/late 2-cell embryos, which was reported to be more accessible by Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq) (*11, 35*).

H3K9me3 heterochromatin becomes rapidly formed following exit from the 2C-like state

We then asked whether $2C^+$ cells could undergo the reverse transition, exiting the 2C-like state and subsequently re-entering pluripotency, thereby becoming ESC-like cells. To answer this question about the kinetics of the reverse transition, we followed the expression of EGFP in FACS-sorted $2C^+$ cells 24 h, 48 h, 72 h and one week after sorting (Fig. 2F). Strikingly, over 60 % of the $2C^+$ cells in culture lost the expression of the MERVL reporter 24 h after sorting. Moreover, 48 h after sorting, only 6 % of the cells still expressed the reporter, suggesting rapid repression of the 2C program, and quick re-establishment of the pluripotency network (Fig. 2F). 72 h and 7 days (7 d) after sorting, EGFP expression levels were comparable to those derived from the endogenous fluctuation ("*steady state*") of ESC cultures (Fig. 2F). The decay in EGFP levels was accompanied by a downregulation of MERVL, endogenous *Dux*, *Zscan4*, *Nelfa*, *Zfp352* and *Eif1a-like* gene expression (Fig. 2G). These results indicate that 2C-like cells could revert their fate back to pluripotency after Dux overexpression, and that such transition occurs rapidly, as early as 24 h after sorting.

We then quantified the number and area of H3K9me3 foci during the $2C^+$ to ESC-like transition (Fig. 2H-J). Our results indicate that chromocenters underwent rapid re-formation and increased in number (24 h: 9.67 ± 0.50; 48 h: 7.07 ± 0.46; 72 h: 8.82 ± 0.44 foci/nucleus) as compared to $2C^+$ cells (3.89 ± 0.19 foci/nucleus), concomitantly to the loss of EGFP expression and to the exit from the 2C-like state (Fig. 2H, 2I). The areas of chromocenters in ESC-like cells were similar across the different time-points analyzed (24 h: 1.54 ± 0.15; 48 h: 1.77 ± 0.14; 72 h: 1.75 ± 0.10 μm^2) and smaller of those of $2C^+$ cells (4.76 ± 0.33 μm^2) (Fig. 2H, 2J). These results suggest that the *in vitro* transition of the $2C^+$ cells toward ESC-like state can be used as a model system to study chromocenter formation and chromatin reorganization occurring during early development.

SMARCAD1 and TOPBP1 associate with H3K9me3 in ESCs and can maintain heterochromatin foci

The results of the iPOTD revealed TOPBP1 as a potential regulator of chromocenter reorganization. SMARCAD1 has been shown to interact with TOPBP1 in yeast and human cells (*27*) (Fig. 1K, 1L). Interestingly, SMARCAD1, a SWI/SNF-like chromatin remodeler, is known to promote heterochromatin maintenance during DNA replication in terminally differentiated cells and silencing of endogenous retroviruses in ESCs (*36, 37*). Nonetheless,

it is not known whether SMARCAD1 plays a role in 2C-like fate transition and early embryo development.

We, therefore, decided to investigate SMARCAD1 and TOPBP1 in $2C^+$ cells undergoing the transition to ESC-like cells where chromocenters are formed *de novo*. First, we evaluated whether SMARCAD1 was associated with TOPBP1 and/or H3K9me3 in ESCs, since TOPBP1 enriched in the $2C^-$ and Luc chromatomes. We observed that both SMARCAD1 and TOPBP1 were associated with H3K9me3 (Fig. 3A). However, we could not detect direct association between TOPBP1 and SMARCAD1, likely because of the suboptimal efficiency of available antibodies. Indeed, we cannot exclude their interaction as was shown previously in yeast and human cells (*27*). These data suggest SMARCAD1 and TOPBP1 to be potential regulators of H3K9me3 heterochromatin. Since SMARCAD1 is a chromatin remodeler, we then asked whether it localized at chromocenters in ESCs. We found that SMARCAD1 co-localized with H3K9me3 in heterochromatin foci of chromocenters in both ESCs and $2C^-$ cells (Fig. 3B-D). In contrast, the expression of SMARCAD1 decreased in $2C^+$ cells, where foci were much reduced in number (Fig. 3C, 3D). We then asked whether SMARCAD1 depletion would increase the fraction of $2C^+$ cells. We depleted *Smarcad1* using two independent sgRNAs, as confirmed comparing to a control sgRNA targeting luciferase (fig. S3E). SMARCAD1 depletion resulted in no major impact in the $2C^+$ conversion either in the endogenous fluctuation or in Dux-induced cells when inspected at the steady state (Fig. 3E). We then investigated *Smarcad1*-depleted cells 24 h, 48 h, 72 h after the $2C^+$ exit. Control KO cells (sgLuc) followed comparable exit kinetics as compared to non-transfected (NT) $2C^+$ cells (fig. S3F). However, SMARCAD1 depletion resulted into a tendency to increased percentage of $2C^+$ cells at all time point after the exit (Fig. 3F). Accordingly, the nuclear distribution of SMARCAD1 during exit from the 2C-like state changed. We first observed a diminution in SMARCAD1 signal as ESCs started to express the MERVL reporter, attaining severe reduction of SMARCAD1 in $2C^+$ at the 24 h time point (Fig. 3H, 3I). SMARCAD1 nuclear signal was then gradually recovered in the heterochromatin foci as $2C^+$ cells were converted in ESC-like cells up to the 72 h from the exit, indicating reversibility of foci formation (Fig. 3H, 3I). Surprisingly, the fraction of cells that repressed retroelements within 24 h from the $2C^+$ exit (ESC-like at 24 h) already showed SMARCAD1 enriched foci (Fig. 3H). Altogether, these results suggest that SMARCAD1 was severely reduced from chromatin as ESCs progress to the 2C-like state and, later, SMARCAD1 nuclear distribution was reverted during the $2C^+$ exit.

We then used published single-cell RNA-seq (scRNA-seq) data (*38*) and found *Smarcad1* expression starting at the 2-cell stage, but increasing at the 4-cell stage embryo, which is the time when chromocenters compact during mouse embryo development (Fig. 3J). Notably, *Topbp1* showed a similar expression profile during preimplantation development (Fig. 3J). Similar to what observed for *Smarcad1*-depleted cells, *Topbp1*-depleted cells showed a tendency to increased percentage of $2C^+$ cells at 24 h, 48 h and 72 h after $2C^+$ exit (Fig. 3G). To further confirm these results and to investigate the role of TOPBP1 in the regulation of heterochromatin foci, we generated knocked-down ESC clones carrying shSmarcad1 or shTopbp1 (fig. S4A). We observed that the number of foci decreased and their area become

larger after either knocking down *Smarcad1* or *Topbp1*, with respect to scramble controls (fig. S4B, S4C). Moreover, larger and fewer chromatin foci were visible in $2C^+$ cells when compared to ESCs and $2C^-$ cells (fig. S4D). We confirmed these results investigating *Topbp1*-depleted cells 24 h, 48 h, 72 h after the $2C^+$ exit. We observed a decreased number of foci and their larger area at all time points analyzed after the exit (fig. S4E-G). These data suggest that heterochromatin foci are maintained in ESCs by SMARCAD1 and TOPBP1 and the depletion of both of these two proteins leads to a remodeling of the H3K9me3 foci.

Next, we asked about the functional interaction of SMARCAD1 and TOPBP1 and thus we evaluated the localization of SMARCAD1 after knocking down *Topbp1*. We found a significant reduction of SMARCAD1 co-localization with H3K9me3 in heterochromatin foci in *Topbp1*-depleted cells (Fig. 3K, 3L), suggesting that SMARCAD1 and TOPBP1 might work as complex in the maintenance of heterochromatin foci.

SMARCAD1 and TOPBP1 are necessary for early embryo development

Collectively, our findings suggested that both SMARCAD1 and TOPBP1 could be potential regulators of H3K9me3 heterochromatin in the $2C^+$ transition. With this in mind, we aimed at investigating their function in preimplantation embryos. We injected zygote-stage (E0.5) embryos with morpholino antisense oligos (MO) targeting *Smarcad1* or *Topbp1* along with a scrambled control morpholino (Ctrl MO) (Fig. 4A and fig. S5A). As expected from MO, which acts by blocking translation, SMARCAD1 was degraded from the 2-cell stage, and a reduction in its levels was observed up to the 8-cell stage, in *Smarcad1* MO-injected embryos (fig. S5B, S5C). We could not image the degradation of TOPBP1 since available anti-TOPBP1 antibodies provide unspecific signal in immunofluorescence experiments. It is noteworthy that SMARCAD1 localizes exclusively in the nucleus of preimplantation embryos (fig. S5B). We observed that embryos developed slower than normal when *Smarcad1* was silenced (Fig. 4A, 4B). Indeed, they did not show the formation nor expansion of a blastocoel cavity at the early blastocyst stage, indicating a severe developmental delay (Fig. 4A, 4B). Notably, 68 % of the embryos deficient for *Smarcad1* arrested and did not develop until the late blastocyst stage (Fig. 4A, 4B). In the case of *Topbp1* silencing, we observed an even a more severe phenotype. All the embryos, 100 % of the *Topbp1* MO-injected ones, did not develop and arrest at 4-cell stage (Fig. 4A, 4B).

Since we observed that both SMARCAD1 and TOPBP1 were necessary for embryo developmental progression, we decided next to image H3K9me3 upon depletion of SMARCAD1 or of TOPBP1 (fig. S5A). H3K9me3 signal was significantly reduced in the embryos injected with *Smarcad1* MO already at the 8-cell stage (E2.5), almost one day earlier than early blastocyst (E3.5), when the developmental delay was morphologically visible (Fig. 4C, 4D and fig. S5A). In *Topbp1* MO embryos, we did not observe decreased intensity of the H3K9me3 signal since the developmental arrest was present already at 4-cell stage and variation in this histone mark might be clearly measurable only starting from

morula stage (fig. S5A, S5D, S5E). On the other hand, we analyzed HP1b, a major component of constitutive heterochromatin which binds to both DNA and to H3K9me3 (*39, 40*). We observed a major remodeling of heterochromatin in both 2-cell and 4-cell Topbp1 MO arrested embryos, as indicated by the spreading and increased signal of HP1b (fig. S5F-S5I).

Finally, given that we observed SMARCAD1 reduction in heterochromatin foci in *Topbp1*-depleted cells (Fig. 3K, 3L), we investigated SMARCAD1 level in *Topbp1* MO in 2-cell and 4-cell arrested embryos. We observed a severe reduction of SMARCAD1 that was even more pronounced when analyzing the pool of 2-cell arrested embryos (Fig 4E-4H and S5J-S5L).

Collectively, these results confirm the functional interaction between SMARCAD1 and TOPBP1showing that *Smarcad1* or *Topbp1* knockdown impair mouse embryo development and that their role in the maintenance of H3K9me3 heterochromatin foci might contribute to the developmental arrest. Overall, our results suggest that both SMARCAD1 and TOPBP1 contribute to proper early embryo development.

Discussion

Heterochromatin formation during early embryogenesis is a fundamental aspect of development (*4*). Here, we have reported that the transition from the 2C-like to the pluripotent state is a robust *in vitro* model system to study heterochromatin foci establishment and their reorganization in early embryo development. During the 2C-like to pluripotency transition, we found that heterochromatin foci are re-formed along with the DNA compaction of the chromatin fibers. Unlike previous reports that focused exclusively on transcriptional changes (*13, 14, 41*), our study exploited chromatin proteomics by genome capture to unravel an additional layer of information and complexity in the 2C-like system. Thus, we provided a detailed characterization of the stepwise chromatome dynamics occurring during the 2C-like state transition. Remarkably, we identified the chromatin remodeler factor SMARCAD1 and TOPBP1, a binding protein interacting with topoisomerase to contribute to embryo development. Depletion of SMARCAD1 or TOPBP1 in preimplantation embryos led to severe developmental arrest and to a substantial remodeling of H3K9me3 heterochromatin foci. These findings have important implications because the establishment and maintenance of heterochromatin foci during embryo development is a key step in the embryonic totipotent program of the 2-cell stage toward pluripotency (*1, 42*).

Endogenous retroviruses (ERVs) are transposable elements flanked by long terminal direct repeats (LTRs) (*43, 44*). Tight control of ERVs and their transposable activity is essential for genome integrity and play an important role in early development and pluripotency (*43, 44*). H3K9me3 has been associated with retrotransposons through the KRAB-associated protein 1, KAP1 (*45*). KAP1 led to the silencing of ERVs in ESCs by inducing H3K9me3

heterochromatin formation via the recruitment of the H3K9 histone methyltransferase SETDB1 (*45-47*). SMARCAD1 was discovered recently to directly interact with KAP1 and therefore be an important regulator of the KAP1-SETDB1 silencing complex in ESCs (*36, 48*). SMARCAD1 is also a key factor for ERV silencing in ESCs (*36*), where it remodels nucleosomes (*49*). Of note, although SMARCAD1 is highly expressed in ESCs, its depletion does not affect pluripotency (*36, 49, 50*).

SMARCAD1 has been described in ESCs, yet its function in 2C-like cells has not been explored. Our observation that SMARCAD1 enriches in H3K9me3 heterochromatin foci during the transition from the 2C-like state to pluripotency and that it contributes to early mouse embryo development is aligned with the observations previously reported in ESCs. It will be interesting in the future to study whether SMARCAD1 can tether the KAP1-SETDB1 to directly induce the formation of H3K9me3 heterochromatin foci at the exit of the 2-cell stage in the embryos. Recently, the H3K9 histone methyltransferase SUV39H2 has been reported to catalyze *de novo* H3K9me3 in the paternal pronucleus after fertilization (*51*). Yet, *Suv39h2* downregulation in zygote-stage embryos did not translate on appreciable changes in H3K9me3 levels on the maternal chromatin. This opens up the possibility that different methyltransferases, and their regulators like SMARCAD1, could be responsible for H3K9me3 acquisition in this early developmental stage.

Topoisomerases likely cooperate with the chromatin remodeling factor SMARCAD1 in yeast (*27*). We observed that SMARCAD1 and TOPBP1 both interact with H3K9me3 suggesting they might work as a complex, this result is also confirmed by the observed reduction of SMARCAD1 in heterochromatin foci in both *Topbp1*-depleted cells and *Topbp1*-depleted embryos. We showed that topoisomerase inhibition led to an increase in the fraction of 2C-like cells and cell cycle arrest in the G2/M phase. Moreover, the knock-down of TOPBP1 leads to a severe developmental arrest. Thus, it is also tempting to speculate that cell cycle progression, especially since we observed that $2C^+$ cells might be arrested in the G2/M phase, has a role in regulating SMARCAD1 recruitment and/or function on chromatin during the $2C^+$ exit. In the *in vivo* scenario, this prolonged G2/M phase might be necessary to rewire specific epigenetic modifications in the 2-cell blastomeres to allow heterochromatin formation. TOPBP1, being a DNA topoisomerase 2-binding protein, might have a role in this process. This is a key step before the blastomeres can embark into the correct developmental process, as proposed for early Drosophila embryos (*52*).

By using chromatin proteomics, we have provided additional data that will help to elucidate the molecular intricacies of the 2C-like state and early mammalian development. In the current study, we focused on heterochromatin establishment and we identified SMARCAD1 and TOPBP1, which both interact with H3K9me3. SMARCAD1 might act in the complex as the remodeler factor that, by regulating methyltransferases, can facilitate H3K9me3 deposition at the exit of the totipotent 2-cell stage when heterochromatin is established *de novo*. Although we could not collect robust data on the alteration of the 2C program, we have indication of its prolonged activity when either SMARCAD1 or TOPBP1 are knocked down, in line with a role in regulating early development in the maintenance of heterochromatin and the regulation of the 2C program.

Materials and Methods

Cell lines and culture conditions

E14Tg2a mouse ESCs were cultured in gelatinized plates in high glucose DMEM supplemented with 15 % FBS (Sigma), GlutaMAX, sodium pyruvate, non-essential amino acids, penicillin/streptomycin, 100 µM 2-mercaptoethanol, 1000 U/ml mouse leukemia inhibitory factor (mLIF) (Millipore), 1 µM PD0325901 and 3 µM CHIR99021. After viral infection, ESCs were selected and maintained with ES medium containing the appropriate combination of selection drugs (250 µg/ml Geneticin (G418, Life Technologies), 0.5 µg/ml Puromycin (Life Technologies)). ESCs were treated with 2 µg/ml doxycycline (D9891, Sigma) for 24 h to induce Dux expression.

Lentivirus production and ESC infection

Lentiviral particles were produced following the RNA interference Consortium (TRC) instructions for viral production and cell infection (http://www.broadinstitute.org/rnai/public/). HEK293T cells were co-transfected with the lentiviral plasmid of interest (pCW57.1-Luciferase or pCW57.1-mDux-CA) and the viral packing vectors (pCMV-ΔR8.9 and pCMV-VSV-G) using the CalPhos mammalian transfection kit (631312, Clontech). pCW57.1-Luciferase and pCW57.1-mDux-CA were a gift from Stephen Tapscott (Addgene plasmids #99283 and #99284). Short hairpins targeting *Smarcad1* (shSmarcad1), *Topbp1* (shTopbp1 #1 and shTopbp1 #2) and a scramble control sequence (shScbl) were cloned into the pLKO.1-Hygro lentiviral vector (Addgene plasmid #24150). The lentiviral-containing medium was harvested from HEK293T cells at 48 h and 72 h after transfection, filtered and used for ESC infection. Two days after the last round of infection, ESCs were selected with the indicated concentration of the selection drug (see Cell culture).

shRNA Name	Top oligo (5' to 3')
shScbl	GTCACGATAAGACAATGAT
shSmarcad1	CCTCCCTTCTAAACCAAAGTT
shTopbp1 #1	CCTGAATTTGAATCACTGGTT
shTopbp1 #2	GCTCTTAGAAACTGCGAGAAT

Table 1 – List of top oligos used for cloning shRNAs

Fluorescence-activated cell sorting (FACS)

Quantification of GFP positive cells and cell cycle analysis was performed with a LSR II Analyzer (BD Biosciences). For cell sorting, an Influx Cell Sorter (BD Biosciences) was used to sort the specified populations in each experiment.

Cell cycle analysis by flow cytometry

For cell cycle analysis of live cells, 5 $\times 10^4$ ESCs were plated per well in gelatin-coated 6-well plates one day before starting the experiment. At the moment of the assay, ESCs were trypsinized, collected and washed with PBS before incubation with ES medium supplemented with 10 µg/ml Hoechst 33342 (H1399, Thermo Fisher) for 30 min at 37 ºC. Propidium iodide (PI) (1 µg/ml; P4864, Sigma) was added to stain dead cells. All flow cytometry data were processed and analyzed with FlowJo (v10).

Inhibition of DNA topoisomerases

To inhibit DNA topoisomerases, ESCs were treated with 500 nM of the topoisomerase I inhibitor camptothecin (CPT; ab120115, Abcam) and/or with 5 µM of the topoisomerase II inhibitor ICRF-193 (I4659, Sigma) for 12 h.

Immunostaining, image processing and quantification

Immunofluorescence staining of ESCs

ESCs were plated at a concentration of 56.000 cells/cm^2 in gelatin-coated borosilicate glass bottom Nunc Lab-Tek (155411, Thermo Fisher) or µ-Slide (80827, Ibidi) 8-well chambers. Cells were fixed with 4 % paraformaldehyde (PFA) for 10 min and were then washed three times with PBS. Cells were permeabilized and blocked (10 % GS, 2.5 % BSA, 0.4 % Triton X-100) for 30 min at room temperature (RT). Incubation with the corresponding primary antibodies at the indicated dilutions lasted 3 h at 37 ºC. Cells were then washed and incubated with Alexa Fluor (Molecular Probes, Invitrogen) secondary antibodies for 1 h at RT. For H3K9me3 and SMARCAD1 co-staining, cells were washed three times with PBS after secondary antibody incubation. Then cells were incubated with second primary antibody and the corresponding secondary antibody as indicated above. Finally, cells were washed three times with PBS containing DAPI for nuclear counterstain. Images were acquired on a Leica TCS SP5 confocal microscope equipped with a 63x oil objective.

The following antibodies were used: chicken anti-GFP (1:500; ab13970, Abcam), mouse anti-Oct-3/4 (1:200; sc-5279, Santa Cruz), rabbit anti-histone H3K9me3 (1:500; ab8898, Abcam), mouse anti-SMARCAD1 (1:500; ab67548, Abcam), goat anti-chicken Alexa Fluor 488, goat anti-mouse Alexa Fluor 568, goat anti-rabbit Alexa Fluor 568, goat anti-mouse Alexa Fluor 647. All secondary antibodies were provided by Molecular Probes (Invitrogen).

EdC incorporation and DNA labelling

To label DNA, a 14 h incorporation pulse of 5-ethynyl-2’-deoxycytidine (EdC; T511307, Sigma) at 2.5 µM was performed in ESCs, in parallel to doxycycline treatment. Cells were plated in gelatin-coated borosilicate glass bottom chambers at a concentration of 56.000 cells/cm^2 in ES medium supplemented with EdC for 14 hours. At the end of EdC incorporation, ESCs were fixed with PFA 4 % (43368, Thermo Fisher Alfa Aesar) and

permeabilized with 0.4 % Triton X-100. Click chemistry reaction was performed by incubating cells for 30 min at RT in click chemistry buffer: 100 mM Hepes pH 8.2, 50 mM Amino Guanidine (396494, Sigma), 25 mM Ascorbic Acid (A92902, Sigma), 1 mM $CuSO_4$, 2 % Glucose (G8270, Sigma), 0.1 % Glox solution [0.5 mg/ml glucose oxidase, 40 mg/ml catalase (G2133 and C100, Sigma)] and 10 mM Alexa Fluor 647 Azide (A-10277, Thermo Fisher) (*31, 32, 53*). After washing the samples three times with PBS, we directly proceeded to perform STORM imaging.

STORM imaging

Stochastic Optical Reconstruction Microscopy (STORM) imaging was performed on a N-STORM 4.0 microscope (Nikon) equipped with a CFI HP Apochromat TIRF 100x 1.49 oil objective and a iXon Ultra 897 camera (Andor) with a pixel size of 16 µm. This objective/camera combination provides an effective pixel size of 160 nm. STORM images were acquired with 10 msec exposure time for 60000 frames using highly inclined (HILO) illumination. An activator/reporter pair strategy was used with AF405 and AF647 fluorophores, respectively. Continuous imaging acquisition was performed with simultaneous 405 nm and 647 nm illumination. 647 nm laser was used at constant ~2 kW/cm^2 power density. 405 nm laser was used at low laser power and gradually increased during the imaging to enhance fluorophore reactivation and to maintain the density of localizations per frame constant. Before STORM imaging, we acquired conventional fluorescence images of GFP for each nucleus to discriminate between $2C^-$ and $2C^+$ cells. Imaging buffer composition for STORM imaging was 100 mM Cysteamine MEA (30070, Sigma), 1 % Glox Solution and 5 % Glucose (G8270, Sigma) in PBS.

STORM images were analyzed and rendered in Insight3 as previously described (*54, 55*). Localizations were identified based on an intensity threshold and the intensity distribution of their corresponding Point Spread Functions (PSFs) fit with a 2D Gaussian to determine the x-y positions of their centers with high accuracy (~20 nm).

Voronoi Tesselation analysis

For Voronoi Tesselation analysis, we used the list of localization from STORM (*33, 34*) and then we used a previously developed custom-made Matlab script (*31*). X-y coordinates of the localizations were used to generate the Voronoi polygons. Local densities were defined as the inverse value of the area of each Voronoi polygon. For visualization, we color-coded each Voronoi polygon based on their area, from yellow for the smallest polygons (density $> 0.01\ nm^{-2}$), to blue for larger polygons (density $< 0.0001\ nm^{-2}$). Finally, the largest 0.5 % of polygons were set to black. For each nucleus, we computed the mean Voronoi density (nm^{-2}) as a measure of global DNA compaction.

For the GFP intensity score, we quantified the GFP conventional images (488 nm channel) with lower intensities in order to assign a GFP intensity score to each nucleus. We summed the fluorescence intensity ADU counts inside each nucleus and divided it by the total number of pixels to obtain the average GFP intensity. Then, we used the distribution of GFP

intensities from the different nuclei to normalize the values, obtaining a GFP intensity score ranging from 0 (less bright) to 1 (most bright). We then performed a cell-by-cell analysis of the relation between GFP intensity score and global chromatin compaction obtained from Voronoi Tesselation analysis.

Immunofluorescence of preimplantation embryos

Preimplantation embryos at E1.5 and E2.5 stages were fixed with 2 % PFA for 10 min at RT, permeabilized (0.25 % Triton X-100) for 10 min, and then blocked (3 % BSA) for 1 h at 37 ºC. Incubation with the corresponding primary antibodies at the indicated dilutions in 1 % BSA lasted one overnight at 4 ºC. After washing, embryos were incubated with Alexa Fluor (Molecular Probes, Invitrogen) secondary antibodies diluted in 1 % BSA for 1 h at 37 ºC. Finally, embryos were washed and transferred to an imaging buffer containing DRAQ5 (1:500; 62251, Thermo Fisher) for DNA staining. Images were acquired on a Leica TCS SP8 STED3X confocal microscope equipped with a 63x oil objective.

The following antibodies were used: rabbit anti-histone H3K9me3 (1:500; ab8898, Abcam), mouse anti-SMARCAD1 (1:250; ab67548, Abcam), rabbit anti-HP1β (1:200; ab10478, Abcam), goat anti-rabbit Alexa Fluor 488 and goat anti-mouse Alexa Fluor 488. All secondary antibodies were provided by Molecular Probes (Invitrogen).

Image processing and quantification

Immunofluorescence images were processed and analyzed with the ImageJ software (https://imagej.net/download/). All immunofluorescence images were acquired with z-stacks. Z-stacks were projected using the *maximum intensity* z-projection type. For SMARCAD1 nuclear signal analysis, manual selection of nuclear area was performed and integrated intensity was measured. For SMARCAD1-H3K9me3 co-immunofluorescence images, a Gaussian blur filtering ($\sigma = 0.5$) was applied to the SMARCAD1 channel. Fluorescence intensities of H3K9me3 foci were analyzed using the *3D Object Counter* function (https://imagej.net/3D_Objects_Counter, ImageJ). Co-localization analysis was done using the *JACoP* plugin (https://imagej.net/JaCoP, ImageJ). Manders' coefficient was calculated with the *JACoP* plugin. Manders' coefficient was used as a co-localization indicator because of its independence of the intensity of the overlapping pixels. For the quantification of H3K9me3 and SMARCAD1 fluorescence intensities in preimplantation embryos, manual selection of the nuclear area was performed for each blastomere. Fluorescent signals were measured and then normalized by the average cytoplasmic signal (background) in each condition. For the normalization step, the fluorescence intensity of a squared shape of equal size was taken for each individual blastomere.

RNA extraction and quantitative real-time PCR (qRT-PCR)

RNA was extracted from pelleted or sorted ESCs using the RNA isolation RNeasy Mini kit (QIAGEN), according to the manufacturer protocol. RNA was reverse-transcribed with

iScript cDNA Synthesis kit (Bio-Rad). qRT-PCR reactions were performed using LightCycler 480 SYBR Green I Master (Roche) in a LightCycler 480 (Roche) instrument, according to the manufacturer recommendations. The oligos used are listed in Table 1. qRT-PCR data was normalized to *Gapdh* or *β-actin* expression. For each sample, we had at least a technical duplicate.

Gene Name	Forward (5' to 3')	Reverse (5' to 3')
Dux	GGAGAAGAGATACCTGAGCTTCAA	AATCTGAGACCCCCATTCG
MERVL	CTCTACCCACTTGGACCATATGAC	GAGGCTCCAAACAGCATCTCTA
MajSat	GCACACTGAAGGACCTGGAATATG	GATTTCGTCATTTTTCAAGTCGTC
Zscan4	GAGATTCATGGAGAGTCTGACTGATGAGTG	GCTGTTGTTTCAAAAGCTTGATGACTTC
Nelfa	TGCTAGTGGACACAGTGTTCGA	TTGAAGCGTGTCCACTGGCC
Zfp352	CCAGGACCCTGCAATACACA	TACAGGTTGTCTCCTGTGTGC
Eif1a-like	AACAGGCGCAGAGGTAAAAA	CTTATATGGCACAGCCTCCT
Smarcad1	AAATTCAGCAAAGACACAGTGATT	CAGAAGGAAGGTCATGGGATT
Topbp1	GCGCCACCAGCAATGTG	TGTACAGGATACAGTTACGTCAGACATTA
Gapdh	TCAAGAAGGTGGTGAAGCAGG	ACCAGGAAATGAGCTTGACAAA
β-actin	GCTGTATTCCCCTCCATCGTG	CACGGTTGGCCTTAGGGTTCAG

Table 2 – List of oligos used for qRT-PCR

Chromatin-bound proteome profiling by genome capture (iPOTD)

ESCs were plated at a concentration of 34.000 cells/cm^2 in gelatin-coated 150-mm dishes. Then, ESCs were pulsed for 24 h with 0.1 µM 5-ethynyl-2'-deoxyuridine (EdU; T511285, Sigma), in parallel to doxycycline treatment. Sorted luciferase (± EdU), 2C$^-$ +EdU and 2C$^+$ +EdU cells were fixed with 1 % PFA, quenched with 0.125 mM glycine (pH 7) and harvested immediately after sorting. Of note, ~ 10^7 cells were sorted per replicate and condition. Cells were later processed as described previously to extract the chromatin-bound proteins (*16, 17*).

Mass spectrometry analysis

Sample preparation

Eluted proteins were reduced with dithiothreitol (37 ºC, 60 min) and alkylated in the dark with iodoacetamide (25 ºC, 20 min) prior to sequential digestion with endoproteinase LysC (1:10 w:w, 37 ºC, overnight; 129-02541, Wako) and trypsin (1:10 w:w, 37 ºC, 8 h) according to filter-aided sample preparation procedure (*56*). After digestion, the peptide mixtures were acidified with formic acid and desalted with a MicroSpin C18 column (The Nest Group, Inc) prior to LC-MS/MS analysis.

Chromatographic and mass spectrometric analysis

Samples were analyzed using a LTQ-Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) coupled to an EASY-nLC 1200 (Thermo Fisher Scientific (Proxeon), Odense, Denmark). Peptides were loaded directly onto the analytical column and were separated by reversed-phase chromatography using a 50-cm column with an inner diameter of 75 µm, packed with 2 µm C18 particles spectrometer (Thermo Scientific, San Jose, CA, USA).

Chromatographic gradients started at 95 % buffer A and 5 % buffer B with a flow rate of 300 nl/min for 5 minutes and gradually increased to 22 % buffer B and 78 % A in 79 min and then to 35 % buffer B and 65 % A in 11 min. After each analysis, the column was washed for 10 min with 10 % buffer A and 90 % buffer B. Buffer A was 0.1 % formic acid in water and buffer B was 0.1 % formic acid in acetonitrile.

The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 1.9 kV and source temperature at 275 °C. Ultramark 1621 was used for external calibration of the FT mass analyzer prior the analyses, and an internal calibration was performed using the background polysiloxane ion signal at m/z 445.1200. The acquisition was performed in data-dependent acquisition (DDA) mode and full MS scans with 1 micro scans at resolution of 120,000 were used over a mass range of m/z 350-1500 with detection in the Orbitrap mass analyzer. Auto gain control (AGC) was set to 1E5 and charge state filtering disqualifying singly charged peptides was activated. In each cycle of data-dependent acquisition analysis, following each survey scan, the most intense ions above a threshold ion count of 10000 were selected for fragmentation. The number of selected precursor ions for fragmentation was determined by the "Top Speed" acquisition algorithm and a dynamic exclusion of 60 seconds. Fragment ion spectra were produced via high-energy collision dissociation (HCD) at normalized collision energy of 28 % and they were acquired in the ion trap mass analyzer. AGC was set to 1E4, and an isolation window of 1.6 m/z and maximum injection time of 200 ms were used. All data were acquired with Xcalibur software.

Digested bovine serum albumin (P8108S, NEB) was analyzed between each sample to avoid sample carryover and to assure stability of the instrument and QCloud has been used to control instrument longitudinal performance during the project (*57*).

Data analysis

Acquired spectra were analyzed using the Proteome Discoverer software suite (v2.3, Thermo Fisher Scientific) and the Mascot search engine (*58*) (v2.6, Matrix Science). The data were searched against a Swiss-Prot mouse database (as in October 2019) plus a list of common contaminants and all the corresponding decoy entries (*22*). For peptide identification a precursor ion mass tolerance of 7 ppm was used for MS1 level, trypsin was chosen as enzyme, and up to three missed cleavages were allowed. The fragment ion mass tolerance was set to 0.5 Da for MS2 spectra. Oxidation of methionine and N-terminal protein acetylation were used as variable modifications whereas carbamidomethylation on cysteines

was set as a fixed modification. False discovery rate (FDR) in peptide identification was set to a maximum of 5 %. The analysis of specific chromatin interactors was carried out with SAINT (v2, Significance Analysis of INTeractome) as previously described (*22, 58*). Replicate 2 of the $2C^+$ condition was excluded for SAINT analysis due to abnormal lower peptide-spectrum matches (PSM) observed in this run. Hierarchical clustering of all the chromatome replicates was computed and visualized using Instant Clue (*59*) v0.5.2 (https://www.instantclue.uni-koeln.de/). Pearson's correlation coefficients were calculated using the Prism software (v9.0, GraphPad, San Diego, CA). To identify proteins shared by the $2C^-$ and Luc chromatomes and not enriched in the $2C^+$ chromatome, an average enrichment value was computed from the respective pairwise comparisons (i.e., Luc vs $2C^-$; Luc vs $2C^+$; $2C^-$ vs Luc; $2C^-$ vs $2C^+$) and then selecting those hits that were more commonly enriched among the $2C^-$ and Luc chromatomes (FC $\geq$ 2). Gene ontology (GO) term enrichment was performed with GO Enrichment Analysis using the PANTHER tool (*60, 61*) (https://geneontology.org/). Protein interaction data were retrieved from the STRING database v11.0 (*62*) and visualized with Cytoscape v3.8.2 (*63*).

Western blot (WB) analysis

Protein extracts were boiled in Laemmli buffer, run in precast protein gel (Mini-PROTEAN TGX; 4561084, Bio-Rad) and then transferred to immuo-blot polyvinylidene difluoride membranes (162-0177, Bio-Rad). The membranes were blocked and incubated with the indicated primary antibodies overnight at 4 °C [rabbit anti-histone H3K9me3 (1:500; ab8898, Abcam), mouse anti-OCT4 (1:500; sc-5279, Santa Cruz), rabbit anti-TOPBP1 (1:1000; ab2402, Abcam) and mouse anti-SMARCAD1 (1:500; ab67548, Abcam)].

After washing, membranes were incubated with specific peroxidase-conjugated secondary antibodies [sheep anti-mouse IgG HRP-linked (1:1000; NA931, GE Healthcare) and donkey anti-rabbit IgG HRP-linked (1:2000; NA934, GE Healthcare)] and visualized on an Amersham Imager 600 (29083461, GE Healthcare Life Sciences).

Dot blot analysis

Samples were spotted in triplicates in 1 µl dots onto a nitrocellulose membrane (0.2 µM, Amersham Protan), air-dried, and detected following standard blotting procedures with the corresponding antibodies (rabbit anti-histone H3 (1:1000; ab1791, Abcam), mouse anti-vinculin (1:1000; V9131, Merck)). Quantification of dot blots was performed by Image Studio Lite software (v5.2, LI-COR, Biosciences). For quantification, each protein was normalized to its background signal.

Co-immunoprecipitation (co-IP)

ESCs were washed with PBS, trypsinized and pelleted. Freshly collected pellets were used for co-IP experiments. ESCs were lysed for 10 min at 4 ºC (10 mM tris-HCl (pH 7.4), 10 mM KCl, 15 mM $MgCl_2$) in the presence of Protease Inhibitor Cocktail (PIC; 5056489001,

Roche). Nuclei were pelleted and lysed for 1 h at 4 ºC (300 mM NaCl, 50 mM Hepes (pH 7.5), 0.5 % NP-40, 2.5 mM $MgCl_2$) in the presence of benzonase endonuclease (100 U/ml; 71205, Novagen). Benzonase removes DNA and RNA, thereby releasing DNA-binding proteins. After nuclear lysis, samples were pelleted and supernatants containing protein extracts were harvested. Protein concentrations were determined using the Pierce BCA protein assay (23225, Thermo Fisher) according to the manufacturer's protocol. For endogenous IPs, an aliquot of protein extract (0.8-1 mg) diluted in dilution buffer (300 mM NaCl, 50 mM Hepes (pH 7.5), 0.5 % NP-40, 5 mM EDTA + PIC) was incubated with 25 µl of pre-washed Dynabeads Protein G (10004D, Thermo Fisher) and with 4 µg of the antibody of interest overnight at 4 ºC. For mock IP, ChromPure Mouse (015-000-003, Jackson Immunoresearch) or Rabbit (011-000-003, Jackson Immunoresearch) IgGs were used. Samples were then washed three times. Elution was performed by incubating the dried beads with 30 µl of 2x Laemmli buffer (1610747, Bio-Rad) supplemented with 50 mM of 2-mercaptoethanol for 10 min at 95 °C. After eluting the samples, we directly proceeded to perform western blotting.

The following antibodies were used: rabbit anti-histone H3K9me3 (ab8898, Abcam) and rabbit anti-TOPBP1 (ab2402, Abcam).

CRISPR-Cas9 plasmid generation and delivery

Single guide RNAs (sgRNAs) targeting each of the specific target genes were retrieved from the Mouse CRISPR Knockout Pooled Library (Addgene #73632). Two sgRNA sequences were selected per gene of interest (for sgRNAs sequences, see Table 2). The sgRNAs with the highest on-target activity score (Rule Set 2) were selected for assembly into the CRISPR-Cas9 vector. An sgRNA targeting the luciferase sequence was also included as control. Primers containing sequences for the sgRNAs were annealed in the presence of T4 ligation buffer (Thermo Fisher) and T4 PNK (NEB) in a heat block (30 ºC for 30 min, 95 ºC for 5 min and slow cool down to RT). Annealed primers were then cloned into the pU6-(BbsI)_CBh-Cas9-T2A-mCherry plasmid following a one-step cloning reaction. pU6-(BbsI)_CBh-Cas9-T2A-mCherry was a gift from Ralf Kuehn (Addgene plasmid #64324).

To generate CRISPR-Cas9-targeted ESCs, cells were nucleofected with 4 µg of the sgRNA-containing plasmid individually following the Amaxa Mouse ES cell Nucleofector kit recommendations (VPH-1001, Lonza). Later, ESCs were FACS-sorted 48 h after nucleofection to enrich for the modified cells.

sgRNA Name	Genomic Sequence	Strand	sgRNA Target Sequence	PAM
sgSmarcad1 #1	NC_000072.6	antisense	AACAGAGCACATTTAAACTG	GGG
sgSmarcad1 #2	NC_000072.6	sense	AGTCTGTAAAACAGCCGCGA	GGG
sgTopbp1 #1	NC_000075.6	sense	GAAGCAGAGTGAGCTCAATG	GGG
sgTopbp1 #2	NC_000075.6	antisense	GTGATTTGCTAAGAATACCA	AGG
sgLuciferase			ACAACTTTACCGACCGCGCC	

Table 3 – List of top oligos used for cloning sgRNAs

Zygote collection and culture

Embryos were collected at E0.5 from 6 to 10 weeks BDF1 female mice (Charles River Laboratories) following 5 IU pregnant mare's serum gonadotrophin (PMSG) and 5 IU human chorionic gonadotropin (hCG) injections at 48 hours intervals. Female mice were mated with BDF1 male mice immediately after hCG injection. Embryos were collected from the oviducts 24 hours post-hCG and were briefly cultured in M2 medium supplemented with 0.2 mg/ml hyaluronidase (H3506, Sigma) to remove cumulus cells. Cumulus-free embryos were washed and cultured with Advanced KSOM medium (MR-101-D, Millipore) at 37 ºC until microinjection.

Microinjection of morpholino antisense oligos

Morpholino antisense oligos (MO) for *Smarcad1* and *Topbp1* and non-targeting control were designed and produced by Gene Tools (Gene Tools, LLC). MOs were microinjected into the cytoplasm of E0.5 embryos using a Narishige micromanipulator system mounted on an Olympus IX71 inverted microscope. Embryos were immobilized using a holding pipette and MOs were then microinjected using a Narishige pneumatic microinjector (IM-300, Narishige). After microinjection, embryos were cultured in Advanced KSOM medium in low oxygen conditions (5 % CO_2, 5 % O_2) at 37 ºC for 5 days (until E5.5). Preimplantation development was examined every 24 hours using an AMG EVOS microscope.

The following MO sequences were used:

Control MO: TCCAGGTCCCCCGCATCCCGGATCC;

Smarcad1 MO: ATATTGGGAGGAACCACCACCCTGA;

Topbp1 MO: ACGGCTCTTGGTCATTTCTGGACAT;

All morpholino sequences are written from 5' to 3' and they are complementary to the translation-blocking target.

All animal experiments were approved and performed in accordance with institutional guidelines [Parc de Recerca Biomèdica de Barcelona (PRBB), Barcelona, Spain] and in accordance with the Ethical Committee for Animal Experimentation (CEEA) number PC-17-0019-PI, approved by La Comissió d'Experimentació Animal, Departament de Territori i Sostenibilitat, Direcció General de Polítiques Ambientals i Medi Natural, Generalitat de Catalunya.

Statistical analysis

As specified in the figure legends, data are presented either as scatter dot plots with line at mean ± SD or at median ± interquartile range, bar graphs showing mean ± SD, min to max boxplots with line at median, or as violin plots showing median and quartiles. All statistical tests and graphs were generated using the Prism software (v9.0, GraphPad, San Diego, CA), unless otherwise indicated. Depending on the experimental setup, we used unpaired two-tailed Student's *t*-test, multiple *t*-test, Fisher's exact test, Mann-Whitney test, one-way

ANOVA or two-way ANOVA with the indicated post-comparison test. In all cases, a *P* value $p \leq 0.05$ was considered significant ($p \leq 0.05$*; $p \leq 0.01$**; $p \leq 0.001$***; $p \leq 0.0001$****; $p > 0.05^{ns}$, not significant).

Acknowledgments

We would like to thank M.-E. Torres-Padilla (Helmholtz Zentrum München), B.R. Cairns (Huntsman Cancer Institute) and B. Huang (UCSF) for sharing E14 ESCs containing the 2C::EGFP reporter, an independent E14 ESC clone containing the Dux-CA cassette and Insight3 software. We are grateful to S. Sdelci for critical reading of the manuscript. We also thank the CRG/UPF Flow Cytometry Unit, the CRG Advanced Light Microscopy Unit, and the PRBB animal facility (PRBB, Barcelona).

Funding: This work was supported by the European Union's Horizon 2020 Research and Innovation Programme (CellViewer No 686637 to M.P.C.); Ministerio de Ciencia e Innovación, grant BFU2017-86760-P (AEI/FEDER, UE); and an AGAUR grant from Secretaria d'Universitats i Recerca del Departament d'Empresa I Coneixement de la Generalitat de Catalunya (2017 SGR 689 to M.P.C.); National Natural Science Foundation of China (No 31971177 to M.P.C.); Spanish Ministry of Economy, Industry and Competitiveness (MEIC) (PID2019-108322GB-100 to L.D.C.), and from AGAUR (L.D.C.). We acknowledge the support of the Spanish Ministry of Science and Innovation to the EMBL partnership, the Centro de Excelencia Severo Ochoa and the CERCA Programme. The CRG/UPF Proteomics Unit is part of the Spanish Infrastructure for Omics Technologies (ICTS OmicsTech) and it is a member of the ProteoRed PRB3 consortium which is supported by grant PT17/0019 of the PE I+D+i 2013-2016 from the Instituto de Salud Carlos III (ISCIII) and ERDF. R.S.-P. was supported by a FI-AGAUR PhD fellowship from the Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya and the co-finance of Fondo Social Europeo (2018FI_B_00637 and FSE). X.T. is supported by a FPI PhD fellowship from the Ministerio de Ciencia e Innovación (PRE2018-085107). S.A. is funded by the Ramon y Cajal program of the Ministerio de Ciencia, Innovación y Universidades and the European Social Fund under the reference number RYC-2018-025002-I, and the Instituto de Salud Carlos III-FEDER (PI19/01814). L.M. is supported by a grant for the recruitment of early-stage research staff FI-2020 (Operational Program of Catalonia 2014-2020 CCI grant no. 2014ES05SFOP007 of the European Social Fund) and La Caixa Foundation fellowship (LCF/BQ/DR20/11790016). M.P. was supported by a Severo Ochoa PhD fellowship from the Subprograma Estatal de Formación del Ministerio de Economía y Competitividad (BES-2015-072802). M.V.N. is funded by FP7/2007–2013 under an REA grant (608959) and Juan de la Cierva-Incorporación 2017.

Author contributions: R.S.-P. and M.P.C. conceptualized this work. R.S.-P. designed and performed most of the experiments with contributions from S.N., X.T., S.A., M.P., M.A.-

B., J.L.G.-V, E.B., M.N.-R. Data were primarily analyzed by R.S.-P. with contributions from S.N., X.T., S.A., P.A.G.-G., D.C., E.B., E.S., L.M., M.N.-R., E.M.
R.S.-P. and M.P.C. wrote the manuscript with input from S.N., X.T., M.V.N., S.A. and L.D.C. M.P.C., M.V.N., and L.D.C. supervised the project.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE (*64*) partner repository with the dataset identifier PXD019703. All other data needed to evaluate the conclusions in this study are present in the paper and/or the Supplementary Materials. Additional materials generated in this study are available from the corresponding author upon request.

References

1. A. Burton, M. E. Torres-Padilla, Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat Rev Mol Cell Biol* **15**, 723-734 (2014).
2. J. Nakayama, J. C. Rice, B. D. Strahl, C. D. Allis, S. I. Grewal, Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110-113 (2001).
3. A. H. Peters, D. O'Carroll, H. Scherthan, K. Mechtler, S. Sauer, C. Schofer, K. Weipoltshammer, M. Pagani, M. Lachner, A. Kohlmaier, S. Opravil, M. Doyle, M. Sibilia, T. Jenuwein, Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323-337 (2001).
4. A. V. Probst, G. Almouzni, Heterochromatin establishment in the context of genome-wide epigenetic reprogramming. *Trends Genet* **27**, 177-185 (2011).
5. K. W. Jones, Chromosomal and nuclear location of mouse satellite DNA in individual cells. *Nature* **225**, 912-915 (1970).
6. T. S. Macfarlan, W. D. Gifford, S. Driscoll, K. Lettieri, H. M. Rowe, D. Bonanomi, A. Firth, O. Singer, D. Trono, S. L. Pfaff, Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57-63 (2012).
7. T. Ishiuchi, R. Enriquez-Gasca, E. Mizutani, A. Boskovic, C. Ziegler-Birling, D. Rodriguez-Terrones, T. Wakayama, J. M. Vaquerizas, M. E. Torres-Padilla, Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat Struct Mol Biol* **22**, 662-671 (2015).
8. Z. Hu, D. E. K. Tan, G. Chia, H. Tan, H. F. Leong, B. J. Chen, M. S. Lau, K. Y. S. Tan, X. Bi, D. Yang, Y. S. Ho, B. Wu, S. Bao, E. S. M. Wong, W. W. Tee, Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells. *Nat Cell Biol* **22**, 175-186 (2020).
9. M. Eckersley-Maslin, C. Alda-Catalinas, M. Blotenburg, E. Kreibich, C. Krueger, W. Reik, Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev* **33**, 194-208 (2019).
10. A. De Iaco, A. Coudray, J. Duc, D. Trono, DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. *EMBO Rep* **20**, (2019).
11. P. G. Hendrickson, J. A. Dorais, E. J. Grow, J. L. Whiddon, J. W. Lim, C. L. Wike, B. D. Weaver, C. Pflueger, B. R. Emery, A. L. Wilcox, D. A. Nix, C. M. Peterson, S. J. Tapscott, D. T. Carrell, B. R. Cairns, Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* **49**, 925-934 (2017).
12. T. Akiyama, L. Xin, M. Oda, A. A. Sharov, M. Amano, Y. Piao, J. S. Cadet, D. B. Dudekula, Y. Qian, W. Wang, S. B. Ko, M. S. Ko, Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells. *DNA Res* **22**, 307-318 (2015).
13. X. Fu, X. Wu, M. N. Djekidel, Y. Zhang, Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells. *Nat Cell Biol* **21**, 835-844 (2019).
14. D. Rodriguez-Terrones, X. Gaume, T. Ishiuchi, A. Weiss, A. Kopp, K. Kruse, A. Penning, J. M. Vaquerizas, L. Brino, M. E. Torres-Padilla, A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat Genet* **50**, 106-119 (2018).
15. M. A. Eckersley-Maslin, V. Svensson, C. Krueger, T. M. Stubbs, P. Giehr, F. Krueger, R. J. Miragaia, C. Kyriakopoulos, R. V. Berrens, I. Milagre, J. Walter, S. A. Teichmann, W. Reik, MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. *Cell Rep* **17**, 179-192 (2016).
16. S. Aranda, E. Borras, E. Sabido, L. Di Croce, Chromatin-Bound Proteome Profiling by Genome Capture. *STAR Protoc* **1**, 100014 (2020).

17. S. Aranda, A. Alcaine-Colet, E. Blanco, E. Borras, C. Caillot, E. Sabido, L. Di Croce, Chromatin capture links the metabolic enzyme AHCY to stem cell proliferation. *Sci Adv* **5**, eaav2448 (2019).
18. A. De Iaco, E. Planet, A. Coluccio, S. Verp, J. Duc, D. Trono, DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* **49**, 941-945 (2017).
19. G. Falco, S. L. Lee, I. Stanghellini, U. C. Bassey, T. Hamatani, M. S. Ko, Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. *Dev Biol* **307**, 539-550 (2007).
20. F. Lu, Y. Liu, L. Jiang, S. Yamaguchi, Y. Zhang, Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev* **28**, 2103-2119 (2014).
21. L. Aloia, B. Di Stefano, L. Di Croce, Polycomb complexes in stem cells and embryonic development. *Development* **140**, 2525-2534 (2013).
22. H. Choi, B. Larsen, Z. Y. Lin, A. Breitkreutz, D. Mellacheruvu, D. Fermin, Z. S. Qin, M. Tyers, A. C. Gingras, A. I. Nesvizhskii, SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods* **8**, 70-73 (2011).
23. J. C. Wang, Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol* **3**, 430-440 (2002).
24. Y. Pommier, Topoisomerase I inhibitors: camptothecins and beyond. *Nat Rev Cancer* **6**, 789-802 (2006).
25. C. S. Downes, D. J. Clarke, A. M. Mullinger, J. F. Gimenez-Abian, A. M. Creighton, R. T. Johnson, A topoisomerase II-dependent G2 cycle checkpoint in mammalian cells. *Nature* **372**, 467-470 (1994).
26. H. M. Robinson, S. Bratlie-Thoresen, R. Brown, D. A. Gillespie, Chk1 is required for G2/M checkpoint response induced by the catalytic topoisomerase II inhibitor ICRF-193. *Cell Cycle* **6**, 1265-1267 (2007).
27. S. C. Bantele, P. Ferreira, D. Gritenaite, D. Boos, B. Pfander, Targeting of the Fun30 nucleosome remodeller by the Dpb11 scaffold facilitates cell cycle-regulated DNA end resection. *Elife* **6**, (2017).
28. K. Liu, Y. Luo, F. T. Lin, W. C. Lin, TopBP1 recruits Brg1/Brm to repress E2F1-induced apoptosis, a novel pRb-independent and E2F1-specific control for cell survival. *Genes Dev* **18**, 673-686 (2004).
29. A. H. Peters, S. Kubicek, K. Mechtler, R. J. O'Sullivan, A. A. Derijck, L. Perez-Burgos, A. Kohlmaier, S. Opravil, M. Tachibana, Y. Shinkai, J. H. Martens, T. Jenuwein, Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol Cell* **12**, 1577-1589 (2003).
30. S. Rea, F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z. W. Sun, M. Schmid, S. Opravil, K. Mechtler, C. P. Ponting, C. D. Allis, T. Jenuwein, Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593-599 (2000).
31. J. Otterstrom, A. Castells-Garcia, C. Vicario, P. A. Gomez-Garcia, M. P. Cosma, M. Lakadamyali, Super-resolution microscopy reveals how histone tail acetylation affects DNA compaction within nucleosomes in vivo. *Nucleic Acids Res* **47**, 8470-8484 (2019).
32. P. J. Zessin, K. Finan, M. Heilemann, Super-resolution fluorescence imaging of chromosomal DNA. *J Struct Biol* **177**, 344-348 (2012).
33. L. Andronov, I. Orlov, Y. Lutz, J. L. Vonesch, B. P. Klaholz, ClusterViSu, a method for clustering of protein complexes by Voronoi tessellation in super-resolution microscopy. *Sci Rep* **6**, 24084 (2016).
34. F. Levet, E. Hosy, A. Kechkar, C. Butler, A. Beghin, D. Choquet, J. B. Sibarita, SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data. *Nat Methods* **12**, 1065-1071 (2015).

35. J. Wu, B. Huang, H. Chen, Q. Yin, Y. Liu, Y. Xiang, B. Zhang, B. Liu, Q. Wang, W. Xia, W. Li, Y. Li, J. Ma, X. Peng, H. Zheng, J. Ming, W. Zhang, J. Zhang, G. Tian, F. Xu, Z. Chang, J. Na, X. Yang, W. Xie, The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652-657 (2016).
36. P. Sachs, D. Ding, P. Bergmaier, B. Lamp, C. Schlagheck, F. Finkernagel, A. Nist, T. Stiewe, J. E. Mermoud, SMARCAD1 ATPase activity is required to silence endogenous retroviruses in embryonic stem cells. *Nat Commun* **10**, 1335 (2019).
37. S. P. Rowbotham, L. Barki, A. Neves-Costa, F. Santos, W. Dean, N. Hawkes, P. Choudhary, W. R. Will, J. Webster, D. Oxley, C. M. Green, P. Varga-Weisz, J. E. Mermoud, Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. *Mol Cell* **42**, 285-296 (2011).
38. Q. Deng, D. Ramskold, B. Reinius, R. Sandberg, Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193-196 (2014).
39. M. Lachner, D. O'Carroll, S. Rea, K. Mechtler, T. Jenuwein, Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116-120 (2001).
40. T. Zhao, T. Heyduk, C. D. Allis, J. C. Eissenberg, Heterochromatin protein 1 binds to nucleosomes and DNA in vitro. *J Biol Chem* **275**, 28332-28338 (2000).
41. X. Fu, M. N. Djekidel, Y. Zhang, A transcriptional roadmap for 2C-like-to-pluripotent state transition. *Sci Adv* **6**, eaay5181 (2020).
42. M. Zernicka-Goetz, S. A. Morris, A. W. Bruce, Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat Rev Genet* **10**, 467-477 (2009).
43. D. Rodriguez-Terrones, M. E. Torres-Padilla, Nimble and Ready to Mingle: Transposon Outbursts of Early Development. *Trends Genet* **34**, 806-820 (2018).
44. M. Friedli, D. Trono, The developmental control of transposable elements and the evolution of higher species. *Annu Rev Cell Dev Biol* **31**, 429-451 (2015).
45. H. M. Rowe, J. Jakobsson, D. Mesnard, J. Rougemont, S. Reynard, T. Aktas, P. V. Maillard, H. Layard-Liesching, S. Verp, J. Marquis, F. Spitz, D. B. Constam, D. Trono, KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237-240 (2010).
46. T. Matsui, D. Leung, H. Miyashita, I. A. Maksakova, H. Miyachi, H. Kimura, M. Tachibana, M. C. Lorincz, Y. Shinkai, Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927-931 (2010).
47. D. Wolf, S. P. Goff, TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131**, 46-57 (2007).
48. D. Ding, P. Bergmaier, P. Sachs, M. Klangwart, T. Ruckert, N. Bartels, J. Demmers, M. Dekker, R. A. Poot, J. E. Mermoud, The CUE1 domain of the SNF2-like chromatin remodeler SMARCAD1 mediates its association with KRAB-associated protein 1 (KAP1) and KAP1 target genes. *J Biol Chem* **293**, 2711-2724 (2018).
49. C. Navarro, J. Lyu, A. M. Katsori, R. Caridha, S. J. Elsasser, An embryonic stem cell-specific heterochromatin state promotes core histone exchange in the absence of DNA accessibility. *Nat Commun* **11**, 5095 (2020).
50. S. Xiao, J. Lu, B. Sridhar, X. Cao, P. Yu, T. Zhao, C. C. Chen, D. McDee, L. Sloofman, Y. Wang, M. Rivas-Astroza, B. Telugu, D. Levasseur, K. Zhang, H. Liang, J. C. Zhao, T. S. Tanaka, G. Stormo, S. Zhong, SMARCAD1 Contributes to the Regulation of Naive Pluripotency by Interacting with Histone Citrullination. *Cell Rep* **18**, 3117-3128 (2017).
51. A. Burton, V. Brochard, C. Galan, E. R. Ruiz-Morales, Q. Rovira, D. Rodriguez-Terrones, K. Kruse, S. Le Gras, V. S. Udayakumar, H. G. Chin, A. Eid, X. Liu, C. Wang, S. Gao, S. Pradhan, J. M. Vaquerizas, N. Beaujean, T. Jenuwein, M. E. Torres-Padilla, Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. *Nat Cell Biol* **22**, 767-778 (2020).

52. C. A. Seller, C. Y. Cho, P. H. O'Farrell, Rapid embryonic cell cycles defer the establishment of heterochromatin by Eggless/SetDB1 in Drosophila. *Genes Dev* **33**, 403-417 (2019).
53. A. Raulf, C. K. Spahn, P. J. Zessin, K. Finan, S. Bernhardt, A. Heckel, M. Heilemann, Click chemistry facilitates direct labelling and super-resolution imaging of nucleic acids and proteinsdaggerElectronic supplementary information (ESI) available. See DOI: 10.1039/c4ra01027bClick here for additional data file. *RSC Adv* **4**, 30462-30466 (2014).
54. M. Bates, B. Huang, G. T. Dempsey, X. Zhuang, Multicolor super-resolution imaging with photo-switchable fluorescent probes. *Science* **317**, 1749-1753 (2007).
55. M. J. Rust, M. Bates, X. Zhuang, Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods* **3**, 793-795 (2006).
56. J. R. Wisniewski, A. Zougman, N. Nagaraj, M. Mann, Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009).
57. C. Chiva, R. Olivella, E. Borras, G. Espadas, O. Pastor, A. Sole, E. Sabido, QCloud: A cloud-based quality control system for mass spectrometry-based proteomics laboratories. *PLoS One* **13**, e0189209 (2018).
58. D. N. Perkins, D. J. Pappin, D. M. Creasy, J. S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).
59. H. Nolte, T. D. MacVicar, F. Tellkamp, M. Kruger, Instant Clue: A Software Suite for Interactive Data Visualization and Analysis. *Sci Rep* **8**, 12648 (2018).
60. H. Mi, A. Muruganujan, D. Ebert, X. Huang, P. D. Thomas, PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* **47**, D419-D426 (2019).
61. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).
62. D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, L. J. Jensen, C. von Mering, The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* **45**, D362-D368 (2017).
63. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
64. Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Perez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, J. A. Vizcaino, The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-D450 (2019).

Figures

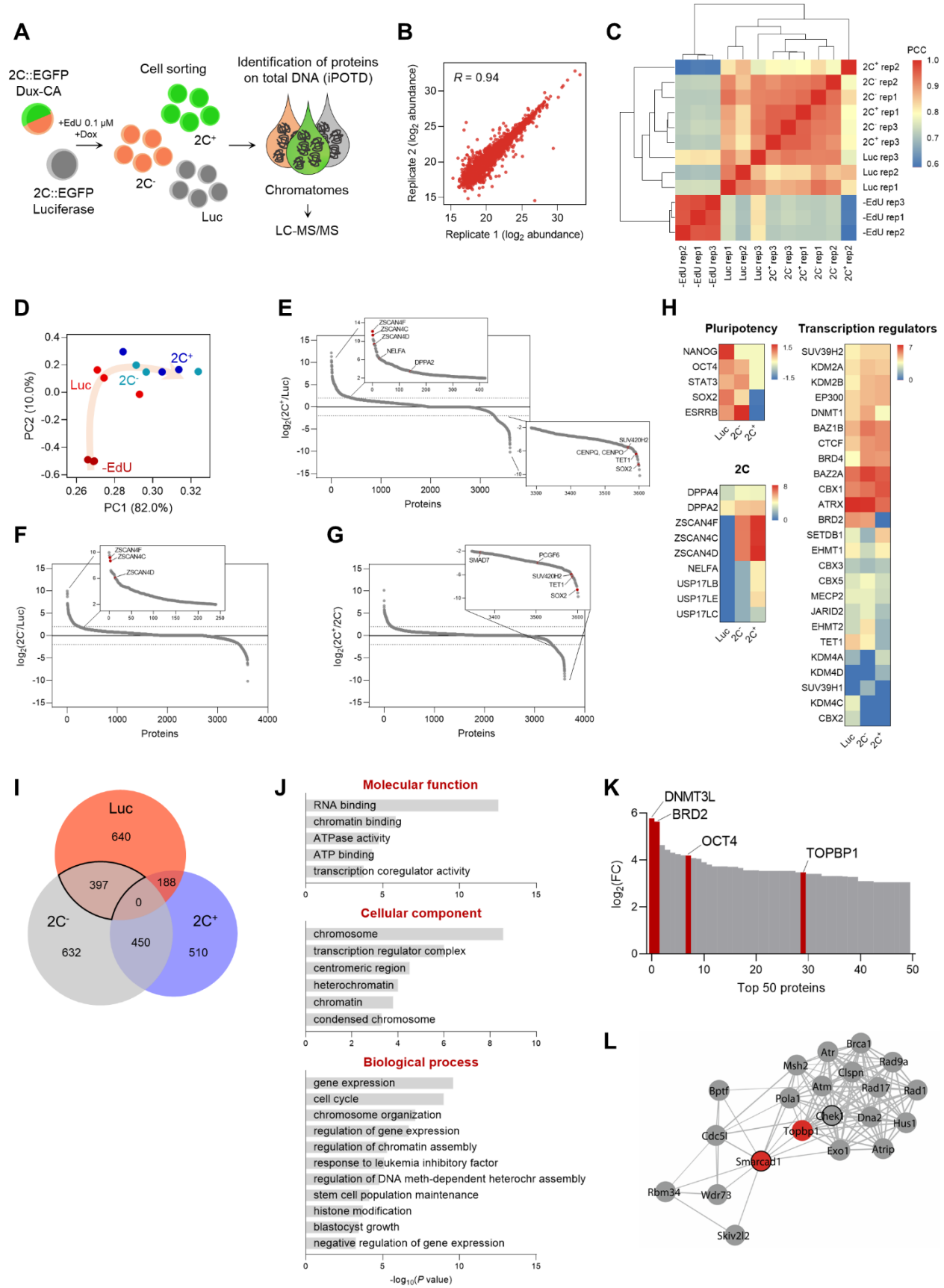
A
2C::EGFP
Dux-CA
+EdU 0.1 μM
+Dox
2C::EGFP
Luciferase
Cell sorting
2C+
2C-
Luc
Identification of proteins
on total DNA (iPOTD)
Chromatomes
LC-MS/MS
B
R = 0.94
Replicate 2 (log2 abundance)
Replicate 1 (log2 abundance)
C
PCC
2C+ rep2
2C- rep2
2C- rep1
2C+ rep1
2C- rep3
2C+ rep3
Luc rep3
Luc rep2
Luc rep1
-EdU rep3
-EdU rep1
-EdU rep2
D
PC2 (10.0%)
PC1 (82.0%)
Luc
2C-
2C+
-EdU
E
log2(2C+/Luc)
Proteins
ZSCAN4F
ZSCAN4C
ZSCAN4D
NELFA
DPPA2
SUV420H2
CENPQ, CENPO
TET1
SOX2
F
log2(2C-/Luc)
Proteins
ZSCAN4F
ZSCAN4C
ZSCAN4D
G
log2(2C+/2C-)
Proteins
SMAD7
PCGF6
SUV420H2
TET1
SOX2
H
Pluripotency
Transcription regulators
NANOG
OCT4
STAT3
SOX2
ESRRB
2C
DPPA4
DPPA2
ZSCAN4F
ZSCAN4C
ZSCAN4D
NELFA
USP17LB
USP17LE
USP17LC
SUV39H2
KDM2A
KDM2B
EP300
DNMT1
BAZ1B
CTCF
BRD4
BAZ2A
CBX1
ATRX
BRD2
SETDB1
EHMT1
CBX3
CBX5
MECP2
JARID2
EHMT2
TET1
KDM4A
KDM4D
SUV39H1
KDM4C
CBX2
Luc
2C-
2C+
I
Luc
640
397
188
0
2C-
632
450
2C+
510
J
Molecular function
RNA binding
chromatin binding
ATPase activity
ATP binding
transcription coregulator activity
Cellular component
chromosome
transcription regulator complex
centromeric region
heterochromatin
chromatin
condensed chromosome
Biological process
gene expression
cell cycle
chromosome organization
regulation of gene expression
regulation of chromatin assembly
response to leukemia inhibitory factor
regulation of DNA meth-dependent heterochr assembly
stem cell population maintenance
histone modification
blastocyst growth
negative regulation of gene expression
-log10(P value)
K
DNMT3L
BRD2
OCT4
TOPBP1
log2(FC)
Top 50 proteins
L
Atr
Brca1
Msh2
Rad9a
Clspn
Bptf
Atm
Rad17
Rad1
Pola1
Chek1
Dna2
Hus1
Cdc5l
Topbp1
Exo1
Atrip
Smarcad1
Rbm34
Wdr73
Skiv2l2

Fig. 1. Chromatin-bound proteome profiling allows the identification of dynamic chromatome changes during 2C-like cell reprogramming. (**A**) Schematic representation of the samples collected to perform the identification of protein on total DNA (iPOTD) workflow. LC-MS/MS, liquid chromatography-tandem mass spectrometry. (**B**) Reproducibility between replicates of independent chromatome preparations. Correlation between replicate 1 and replicate 2 from the $2C^-$ condition is shown. *R* indicates Pearson's R. (**C**) Correlation matrix showing reproducibility among independent replicates of -EdU, Luc, $2C^-$ and $2C^+$ protein abundances. Hierarchical clustering analysis of the individual replicates is also shown. PCC, Pearson's correlation coefficient. (**D**) Principal component analysis (PCA) of the chromatome dataset. Each point corresponds to a single replicate. Beige arrow indicates the 2C-like reprogramming trajectory. (**E**) Protein enrichment analysis of the comparison between $2C^+$ and Luc chromatomes. Red dots indicate known regulators of the 2C-like state, which were found enriched in the $2C^+$ chromatome (upper panel), and novel factors which were found depleted from the $2C^+$ chromatome (lower panel). Dashed lines indicate $\log_2$ FC ± 2. (**F**) Protein enrichment analysis of the comparison between $2C^-$ and Luc chromatomes. Red dots indicate known regulators of the 2C-like state, which were found enriched in the $2C^-$ chromatome. Dashed lines indicate $\log_2$ FC ± 2. (**G**) Protein enrichment analysis of the comparison between $2C^+$ and $2C^-$chromatomes. Red dots indicate novel factors which were found depleted from the $2C^+$ chromatome. Dashed lines indicate $\log_2$ FC ± 2. (**H**) Heat map representation of the chromatin-bound abundance of pluripotency transcription factors, 2-cell-specific factors and transcriptional regulators in Luc, $2C^-$ and $2C^+$ cells. Data are presented as $\log_2$ FC of PSM values to -EdU. PSM, peptide spectrum match. (**I**) Venn diagram indicating the overlap between the identified proteins enriched in Luc, $2C^-$ and $2C^+$ chromatomes after SAINT analysis. Specifically, an average enrichment value was computed from the respective pairwise comparisons (i.e., Luc vs $2C^-$; Luc vs $2C^+$; $2C^-$ vs Luc; $2C^-$ vs $2C^+$), and proteins were selected on the basis of a minimum of FC ≥ 2. The solid line highlights the 397 proteins enriched in the Luc and $2C^-$ chromatomes that were not enriched in the $2C^+$ chromatome. (**J**) Gene ontology analysis of the commonly enriched chromatin-bound proteins identified in (I). (**K**) Distribution of the top 50 chromatin-bound proteins identified in the Luc and $2C^-$ chromatomes ranked by $\log_2$ FC. (**L**) Functional protein network of TOPBP1 interactors. TOPBP1 and SMARCAD1 nodes are colored in red. The black node border indicates chromatin remodeling function. Network edges indicate the degree of confidence prediction of the interaction. Protein interaction data were retrieved from the STRING database (Szklarczyk et al., 2017).

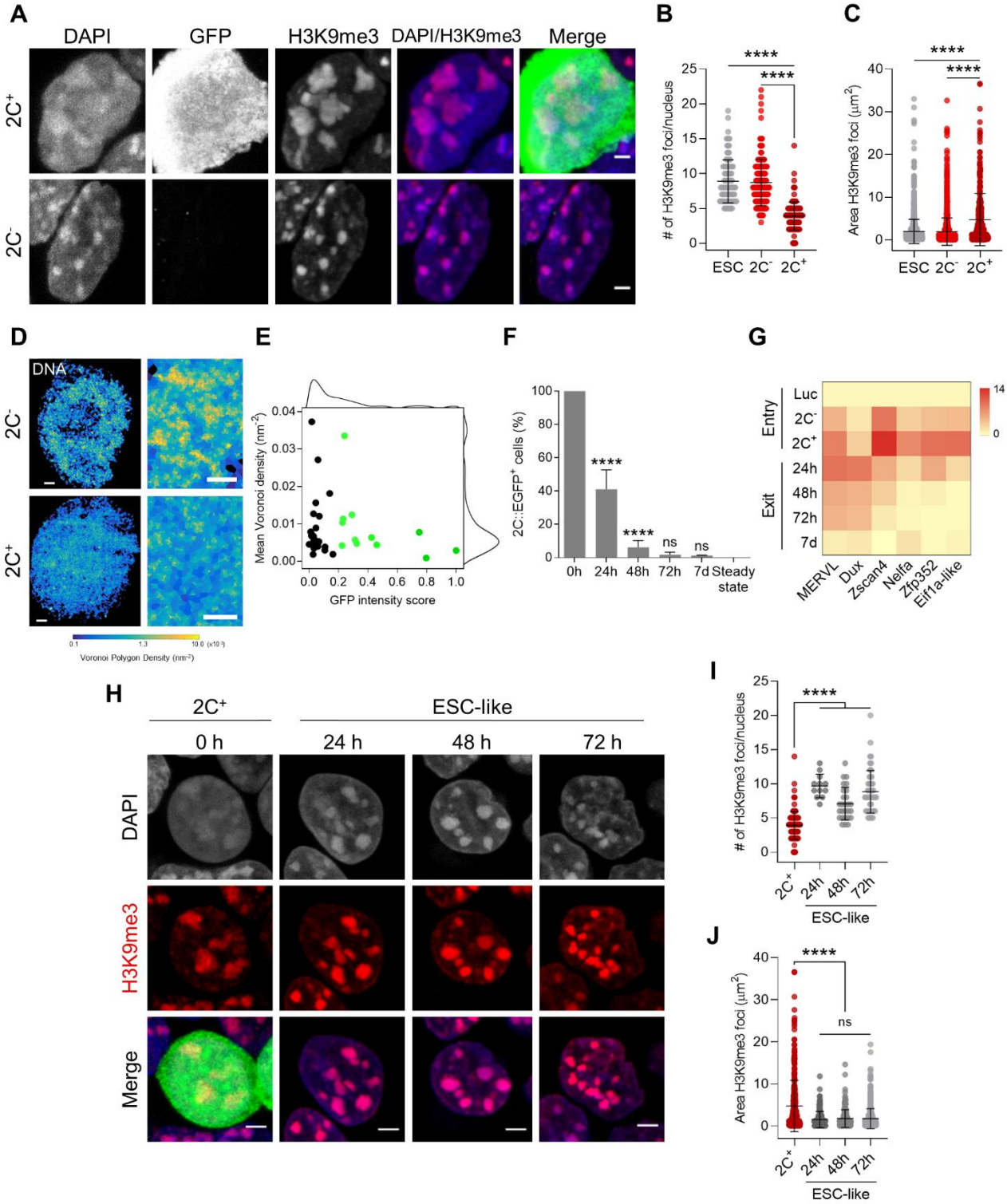
A
DAPI
GFP
H3K9me3
DAPI/H3K9me3
Merge
2C+
2C-
B
# of H3K9me3 foci/nucleus
ESC
2C-
2C+
C
Area H3K9me3 foci (μm2)
D
DNA
Voronoi Polygon Density (nm-2)
E
Mean Voronoi density (nm-2)
GFP intensity score
F
2C::EGFP+ cells (%)
0h
24h
48h
72h
7d
Steady state
ns
G
Entry
Exit
Luc
24h
48h
72h
7d
MERVL
Dux
Zscan4
Nelfa
Zfp352
Elf1a-like
H
2C+
ESC-like
0 h
24 h
48 h
72 h
DAPI
H3K9me3
Merge
I
J

Fig. 2. Entry in the 2C-like state is characterized by the remodeling of H3K9me3 heterochromatin, which is reverted upon 2C$^+$ exit. (**A**) Representative immunofluorescence images of the 2C::EGFP reporter and H3K9me3 in 2C$^-$ and 2C$^+$ cells. Scale bar, 2 µm. (**B**) Quantification of the number of H3K9me3 foci in ESCs, 2C$^-$ and 2C$^+$ cells. Data are presented as scatter dot plots with line at mean ± SD (n > 3 independent cultures, ESCs = 103 cells, 2C$^-$ = 170 cells and 2C$^+$ = 119 cells). $P < 0.0001$**** by one-way ANOVA (Tukey's multiple comparisons test). (**C**) Quantification of H3K9me3 foci area in ESCs, 2C$^-$ and 2C$^+$ cells. Data are presented as scatter dot plots with line at mean ± SD (n > 3 independent cultures, ESCs = 1712 foci, 2C$^-$ = 1445 foci and 2C$^+$ = 340 foci). $P < 0.0001$**** by one-way ANOVA (Tukey's multiple comparisons test). (**D**) Voronoi tessellation rendering of super-resolution images of DNA in 2C$^-$ and 2C$^+$ cells. Full nuclei (left; scale bar, 1 µm) and zoomed images (right; scale bar, 400 nm) are shown. (**E**) Biaxial density plot showing mean Voronoi density of DNA (inverse of the polygon area) as a measure of chromatin compaction and GFP intensity score in 2C$^-$ and 2C$^+$ cells. Cells with a GFP intensity score > 0.2 are colored in green. Black dots indicate 2C$^-$ cells and green dots indicate 2C$^+$ cells. Each dot represents a single-cell (2C$^-$ = 23 cells and 2C$^+$ = 12 cells). (**F**) Quantification of the percentage of 2C-like cells 24 h, 48 h, 72 h and 7 days after 2C$^+$ cell sorting. The endogenous 2C-like fluctuation was used as the steady-state condition. Data are presented as mean ± SD (n = 3 independent experiments). $P = 0.7656^{ns}$, $P < 0.0001$**** by one-way ANOVA (Tukey's multiple comparisons test). (**G**) Heat map representation of *MERVL*, *Dux*, *Zscan4*, *Nelfa*, *Zfp352* and *Eif1a-like* expression in luciferase (Luc), 2C$^-$ and 2C$^+$ sorted cells (entry) and in ESC-like cells at 24 h, 48 h, 72 h and 7 days (7d) after 2C$^+$ sorting. Data are presented as $\log_2$ fold change (FC) values to luciferase detected by qRT-PCR. (**H**) Representative immunofluorescence images of H3K9me3 at 0 h (2C$^+$ before exit), 24 h, 48 h and 72 h after 2C-like state exit. Scale bar, 3 µm. (**I**) Quantification of the number of H3K9me3 foci in 2C$^+$ cells and at 24 h, 48 h and 72 h after 2C-like state exit. Data are presented as scatter dot plots with line at mean ± SD (n = 2 independent cultures, 2C$^+$ = 119 cells, same dataset plotted in Fig. 2B; ESC-like 24 h = 12 cells; ESC-like 48 h = 27 cells; ESC-like 72 h = 49 cells). $P < 0.0001$**** by one-way ANOVA (Tukey's multiple comparisons test). (**J**) Quantification of H3K9me3 foci area in 2C$^+$ cells and at 24 h, 48 h and 72 h after 2C-like state exit. Data are presented as scatter dot plots with line at mean ± SD (n = 2 independent cultures, 2C$^+$ = 340 foci, same dataset plotted in Fig. 2C; ESC-like 24 h = 168 foci; ESC-like 48 h = 238 foci; ESC-like 72 h = 605 foci). $P > 0.05^{ns}$, $P < 0.0001$**** by one-way ANOVA (Tukey's multiple comparisons test).

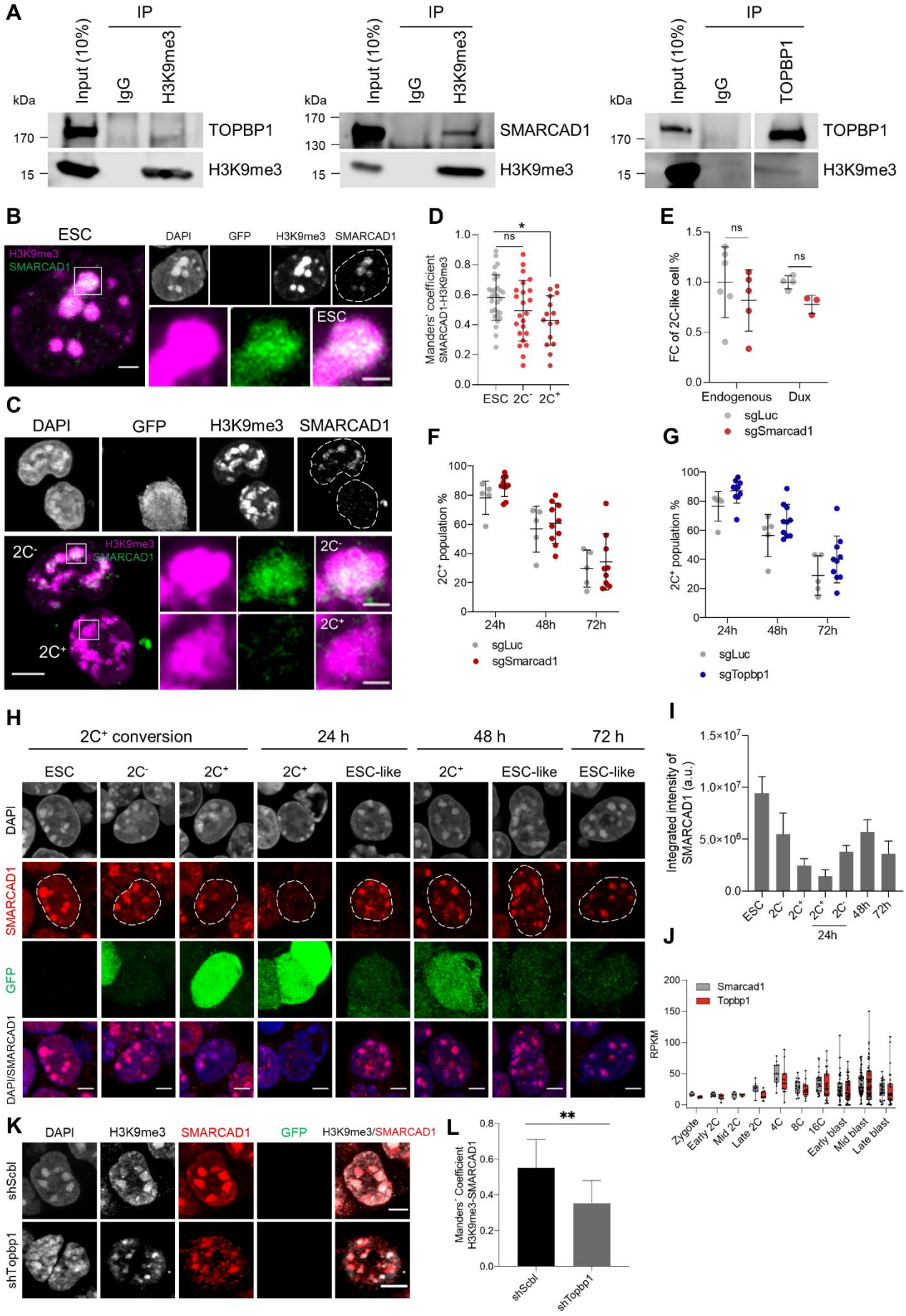
A
IP
Input (10%)
IgG
H3K9me3
kDa
170
15
TOPBP1
H3K9me3
130
SMARCAD1
TOPBP1
B
ESC
DAPI
GFP
H3K9me3
SMARCAD1
C
2C-
2C+
D
Manders' coefficient SMARCAD1-H3K9me3
ns
*
ESC
2C-
2C+
E
FC of 2C-like cell %
Endogenous
Dux
sgLuc
sgSmarcad1
F
2C+ population %
24h
48h
72h
G
sgTopbp1
H
2C+ conversion
ESC-like
I
Integrated intensity of SMARCAD1 (a.u.)
J
Smarcad1
Topbp1
RPKM
Zygote
Early 2C
Mid 2C
Late 2C
4C
8C
16C
Early blast
Mid blast
Late blast
K
shScbl
shTopbp1
L
Manders' Coefficient H3K9me3-SMARCAD1
**

Fig. 3. SMARCAD1 associates with H3K9me3 in ESCs and its nuclear localization is reduced in the 2C-like state. (**A**) Co-immunoprecipitation (co-IP) experiment in ESCs using antibodies against H3K9me3 and TOPBP1. Western blots of H3K9me3, SMARCAD1 and TOPBP1 are shown. (**B**) Representative immunofluorescence images of H3K9me3 and SMARCAD1 in ESCs. Dashed lines indicate nuclei contour. Scale bar, 2 µm. Zoomed images of H3K9me3 and SMARCAD1 foci are shown for comparisons. Scale bar, 1 µm. (**C**) Representative immunofluorescence images of H3K9me3 and SMARCAD1 in $2C^-$ and $2C^+$ cells. Dashed lines indicate nuclei contour. Scale bar, 5 µm. Zoomed images of H3K9me3 and SMARCAD1 foci are shown for comparisons. Scale bar, 1 µm. (**D**) Co-localization analysis showing Manders' coefficient between SMARCAD1 and H3K9me3 in ESCs, $2C^-$ and $2C^+$ cells. Data are presented as scatter dot plots with line at mean ± SD from ESC (n = 30), $2C^-$ (n = 23), $2C^+$ (n = 15) SMARCAD1-H3K9me3 foci. $P > 0.05^{ns}$, $P = 0.0124$* by one-way ANOVA (Dunnett's multiple comparisons test). (**E**) Impact of targeting *Smarcad1* (sgSmarcad1) on the endogenous fluctuation and the Dux-induced 2C-like conversion. Data are presented as scatter dot plots with line at mean ± SD (n ≥ 3 independent CRISPR-Cas9 KO rounds). $P = 0.4286^{ns}$, $P = 0.0571^{ns}$ by Mann-Whitney test. (**F**) Impact of targeting *Smarcad1* (sgSmarcad1) on the 2C-like cell percentage during the $2C^+$ exit (24 h, 48 h and 72 h). Data are presented as scatter dot plots with line at mean ± SD (n = 5 independent CRISPR-Cas9 KO rounds). Individual points indicate scores of technical replicates. $P = 0.1174^{ns}$ at 24 h, $P = 0.6158^{ns}$ at 48 h, $P = 0.6441^{ns}$ at 72 h by multiple *t*-test. (**G**) Impact of targeting *Topbp1* (sgTopbp1) on the 2C-like cell percentage during the $2C^+$ exit (24 h, 48 h and 72 h). Data are presented as scatter dot plots with line at mean ± SD (n = 5 independent CRISPR-Cas9 KO rounds). Individual points indicate scores of technical replicates. $P = 0.0503^{ns}$ at 24 h, $P = 0.1589^{ns}$ at 48 h, $P = 0.2166^{ns}$ at 72 h by multiple *t*-test. (**H**) Representative immunofluorescence images of SMARCAD1 and the 2C::EGFP reporter along the ESCs to $2C^+$ reprogramming and during the $2C^+$ exit (24 h, 48 h and 72 h). Dashed lines indicate nuclei contour. Scale bar, 4 µm. (**I**) SMARCAD1 integrated intensity analysis along the conversion of ESCs into $2C^+$ cells and during the $2C^+$ exit (24 h, 48 h and 72 h). Data are presented as mean ± SD. (**J**) Single-cell RNA-seq (scRNA-seq) expression profile of *Smarcad1* and *Topbp1* in pre-implantation mouse embryos. Data are presented as min-max boxplots with line at median. Each dot represents a single-cell. scRNA-seq data was obtained from ref. (*38*). RPKM, reads per kilobase of transcript per million mapped reads. (**K**) Representative immunofluorescence images of H3K9me3, SMARCAD1 and the 2C::EGFP reporter in *Topbp1* knockdown (shTopbp1) and control scramble (shScbl) cells. Scale bar, 5 µm. (**L**) Co-localization analysis showing Manders' coefficient between H3K9me3 and SMARCAD1 in *Topbp1* knockdown (shTopbp1) and control scramble (shScbl) cells. Data are presented as mean ± SD (n = 2 independent cultures). $P = 0.0066$** by unpaired two-tailed Student's *t*-test.

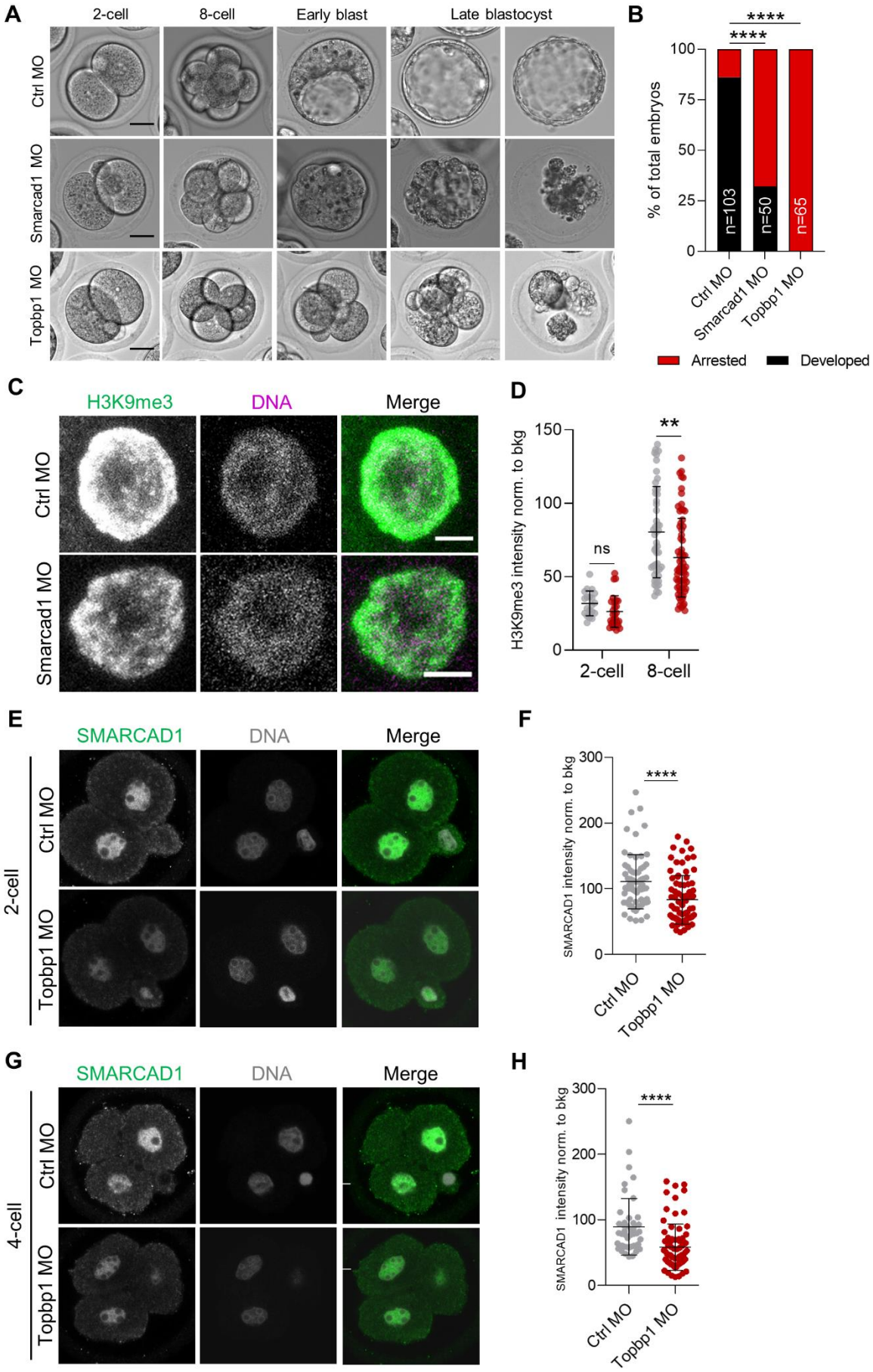
A
2-cell
8-cell
Early blast
Late blastocyst
Ctrl MO
Smarcad1 MO
Topbp1 MO
B
****
****
% of total embryos
100
75
50
25
0
n=103
n=50
n=65
Ctrl MO
Smarcad1 MO
Topbp1 MO
Arrested
Developed
C
H3K9me3
DNA
Merge
Ctrl MO
Smarcad1 MO
D
H3K9me3 intensity norm. to bkg
150
100
50
0
ns
**
2-cell
8-cell
E
SMARCAD1
DNA
Merge
2-cell
Ctrl MO
Topbp1 MO
F
SMARCAD1 intensity norm. to bkg
300
200
100
0
****
Ctrl MO
Topbp1 MO
G
SMARCAD1
DNA
Merge
4-cell
Ctrl MO
Topbp1 MO
H
SMARCAD1 intensity norm. to bkg
300
200
100
0
****
Ctrl MO
Topbp1 MO

Fig. 4. SMARCAD1 and TOPBP1 downregulation impairs embryo development (**A**) Representative embryos from control (Ctrl), *Smarcad1* and *Topbp1* morpholino-injected (MO) groups from 2-cell (E1.5) to late blastocyst stage (E5.5). Scale bar, 20 µm. (**B**) Quantification of the percentage of arrested or fully developed embryos at late blastocyst stage (E4.5). $P < 0.0001$**** by Fisher's exact test (Ctrl MO = 103 embryos, *Smarcad1* MO = 50 embryos, *Topbp1* MO = 65 embryos). (**C**) Representative immunofluorescence images of H3K9me3 in Ctrl and *Smarcad1* MO embryos at 8-cell stage (E2.5) embryos. Representative blastomere nuclei are shown. Scale bar, 5 µm. (**D**) Quantification of H3K9me3 mean fluorescence intensity in control (Ctrl, grey dots) and *Smarcad1* MO (red dots) embryos at 2-cell (E1.5) and 8-cell stage (E2.5). Data are presented as scatter dot plots with line at mean ± SD (2-cell: Ctrl MO = 12 embryos, *Smarcad1* MO = 15 embryos; 8-cell: Ctrl MO = 16 embryos, *Smarcad1* MO = 20 embryos). H3K9me3 signal was normalized to the average background signal. $P = 0.0618^{ns}$ and $P = 0.0016$** by unpaired two-tailed Student's *t*-test. (**E**) Representative immunofluorescence images of SMARCAD1 in Ctrl and *Topbp1* MO embryos at 2-cell stage (E1.5) embryos. Representative blastomere nuclei are shown. Scale bar, 10 µm. (**F**) Quantification of SMARCAD1 mean fluorescence intensity in Ctrl and *Topbp1* MO embryos at 2-cell (E1.5). Data are presented as scatter dot plots with line at mean ± SD (Ctrl MO = 38 embryos, *Topbp1* MO = 44 embryos). SMARCAD1 signal was normalized to the average background signal. $P < 0.0001$**** by unpaired two-tailed Student's *t*-test. (**G**) Representative immunofluorescence images of SMARCAD1 in Ctrl and *Topbp1* MO embryos at 4-cell stage (E2.0) embryos. Representative blastomere nuclei are shown. Scale bar, 10 µm. (**H**) Quantification of SMARCAD1 mean fluorescence intensity in Ctrl and *Topbp1* MO embryos arrested at 4-cell. Data are presented as scatter dot plots with line at mean ± SD (Ctrl MO = 20 embryos, *Topbp1* MO = 31 embryos). SMARCAD1 signal was normalized to the average background signal. $P < 0.0001$**** by unpaired two-tailed Student's *t*-test.

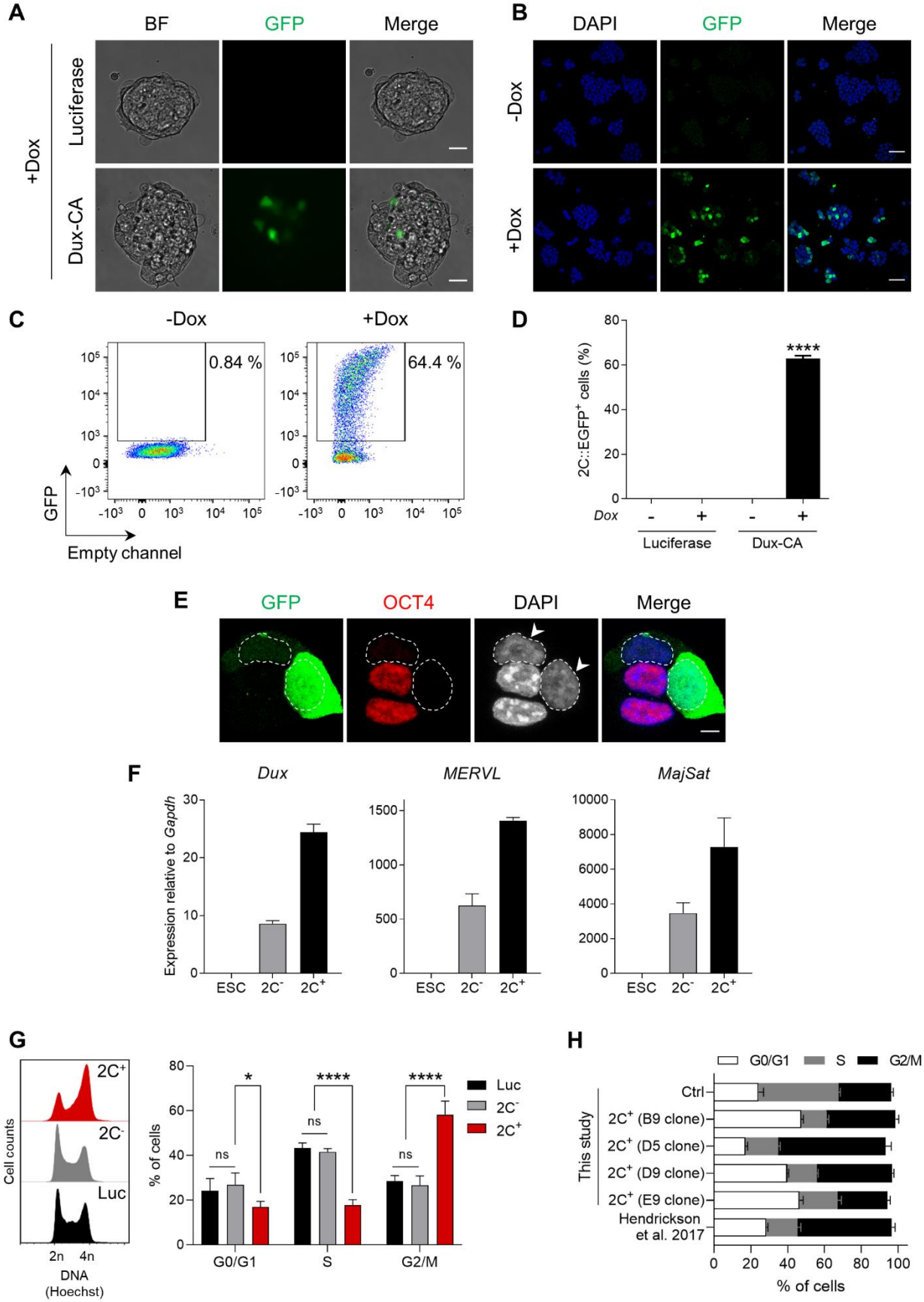
A
BF
GFP
Merge
+Dox
Luciferase
Dux-CA
B
DAPI
GFP
Merge
-Dox
+Dox
C
-Dox
+Dox
0.84 %
64.4 %
GFP
Empty channel
D
2C::EGFP+ cells (%)
****
Dox
Luciferase
Dux-CA
E
GFP
OCT4
DAPI
Merge
F
Dux
MERVL
MajSat
Expression relative to Gapdh
ESC
2C-
2C+
G
2C+
2C-
Luc
Cell counts
2n
4n
DNA
(Hoechst)
% of cells
ns
*
****
Luc
2C-
2C+
G0/G1
S
G2/M
H
G0/G1
S
G2/M
Ctrl
2C+ (B9 clone)
2C+ (D5 clone)
2C+ (D9 clone)
2C+ (E9 clone)
Hendrickson et al. 2017
This study
% of cells

Fig. S1. Characterization of Dux-derived 2C-like cells. (**A**) Representative live-cell images of stable luciferase and Dux-CA ESC lines upon doxycycline (Dox) induction. Scale bar, 20 µm. (**B**) Representative immunofluorescence images of the 2C::EGFP reporter in the Dux-CA line in control (-Dox) and Dux overexpressing (+Dox) conditions showing activation of the 2C::EGFP reporter after 24 h of Dox administration. Scale bar, 50 µm. (**C**) Representative FACS plots showing GFP^+ cells in the Dux-CA line without (-Dox) and after 24 h of Dox treatment (+Dox). (**D**) Effect of Dux overexpression on the activation of the 2C::EGFP reporter by flow cytometry. Data are presented as mean ± SD (n = 3 independent cultures). $P < 0.0001$**** by one-way ANOVA (Tukey's multiple comparisons test). (**E**) Representative immunofluorescence images out of two experiments of the 2C::EGFP reporter and the endogenous pluripotency transcription factor OCT4 after 24 h of Dox induction. Arrowheads indicate $2C^+$ cells. Dashed lines indicate nuclei contour. Scale bar, 5 µm. (**F**) qRT-PCR of *Dux*, MERVL and major satellites (MajSat) in ESCs, $2C^-$ and $2C^+$ sorted cells. Data are presented as mean ± SE (n ≥ 2, $2C^-$ and $2C^+$ samples are technical replicates). (**G**) Cell cycle profile of non-induced luciferase (Luc) control ESCs, $2C^-$ and $2C^+$ cells (left). Quantification of the percentage of ESCs, $2C^-$ and $2C^+$ cells in different phases of the cell cycle (right). Data are presented as mean ± SD (n > 3 independent cultures). $P > 0.05^{ns}$, $P = 0.0289$*, $P < 0.0001$**** by two-way ANOVA (Dunnett's multiple comparisons test). (**H**) Quantification of the percentage of control ESCs (Ctrl) and $2C^+$ cells in different phases of the cell cycle. $2C^+$ cells induced from several Dux overexpressing clonal lines generated in our laboratory were analyzed. An independent Dux-CA clonal line was included for comparison, ref. Hendrickson et al., 2017. Data are presented as mean ± SE (n ≥ 3, technical replicates).

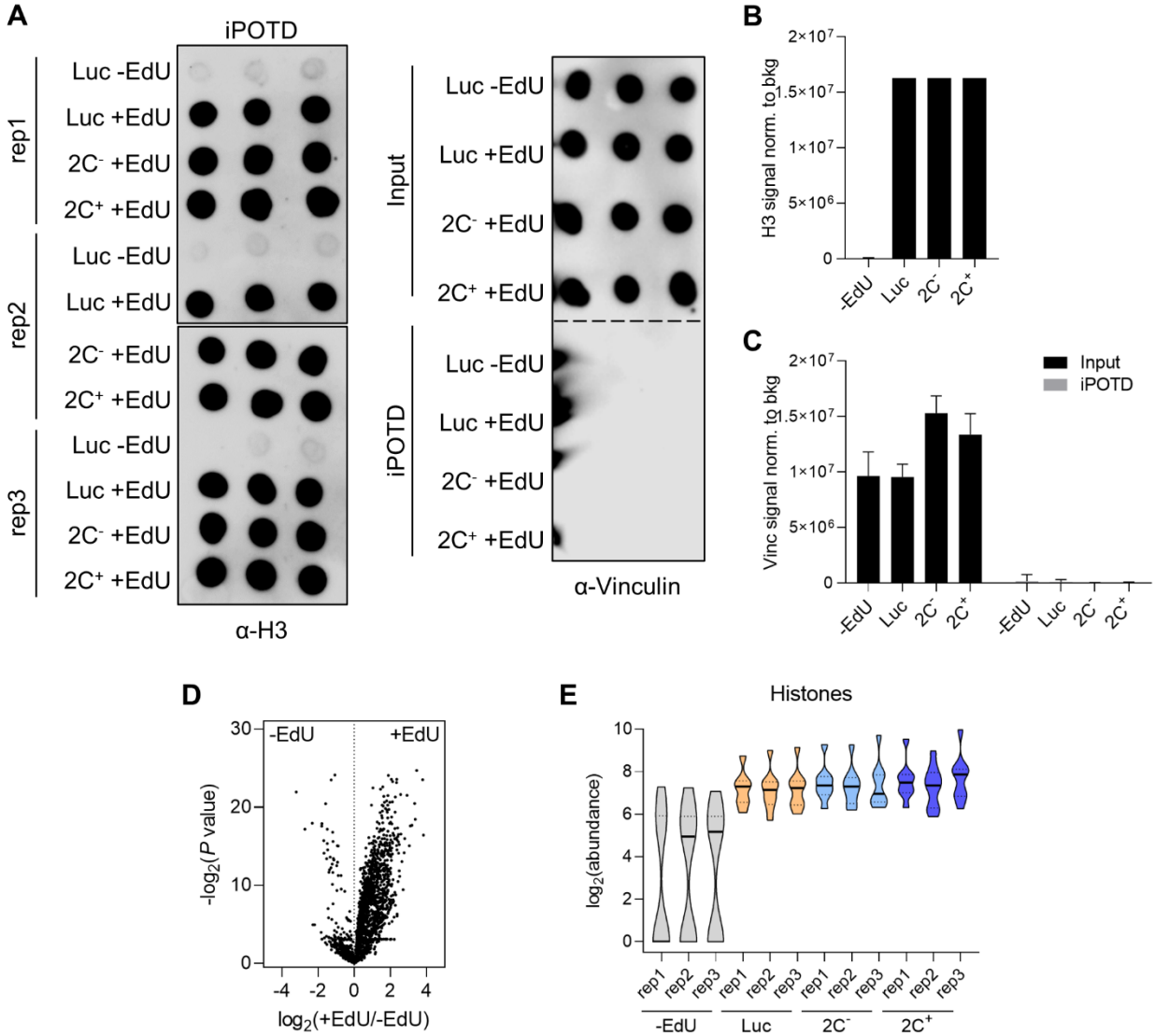
A
iPOTD
rep1
rep2
rep3
Luc -EdU
Luc +EdU
2C⁻ +EdU
2C⁺ +EdU
α-H3
Input
iPOTD
α-Vinculin
B
H3 signal norm. to bkg
C
Vinc signal norm. to bkg
Input
iPOTD
-EdU
Luc
2C⁻
2C⁺
D
-EdU
+EdU
-log2(P value)
log2(+EdU/-EdU)
E
Histones
log2(abundance)
rep1
rep2
rep3

Fig. S2. Chromatin proteomics of 2C-like cells. (**A**) Independent DNA-mediated chromatin pull-down (iPOTD) eluates from sorted luciferase, $2C^-$ and $2C^+$ replicates in the absence or presence of EdU (±EdU) were analyzed by dot blot with an anti-H3 antibody (left). Input and eluates from equivalent preparations were incubated with an anti-vinculin antibody (right). Each condition was spotted in triplicates. (**B**) Quantification of histone H3 signal detected by dot blot in the absence or presence of EdU. Data are presented as mean ± SD of H3 signal normalized to the background. (**C**) Quantification of vinculin (Vinc) signal detected by dot blot in input or iPOTD samples. Data are presented as mean ± SD of vinculin signal normalized to the background. (**D**) Volcano plot of proteins identified by mass-spectrometry after DNA-mediated chromatin pull-down in Luc and –EdU conditions. 2396 proteins were enriched in the Luc (+EdU) chromatome compared with the control –EdU condition (fold change > 1). (**E**) Abundance of histones in the individual replicates from –EdU, Luc, $2C^-$ and $2C^+$ conditions. The following histones were included in the analysis: core histones H2A, H2B and H4, macro-H2A.1, macro-H2A.2, H2A.V, H2A.X, H3.3 and CENP-A. Violin plot shows median with a solid line and quartiles with dashed lines.

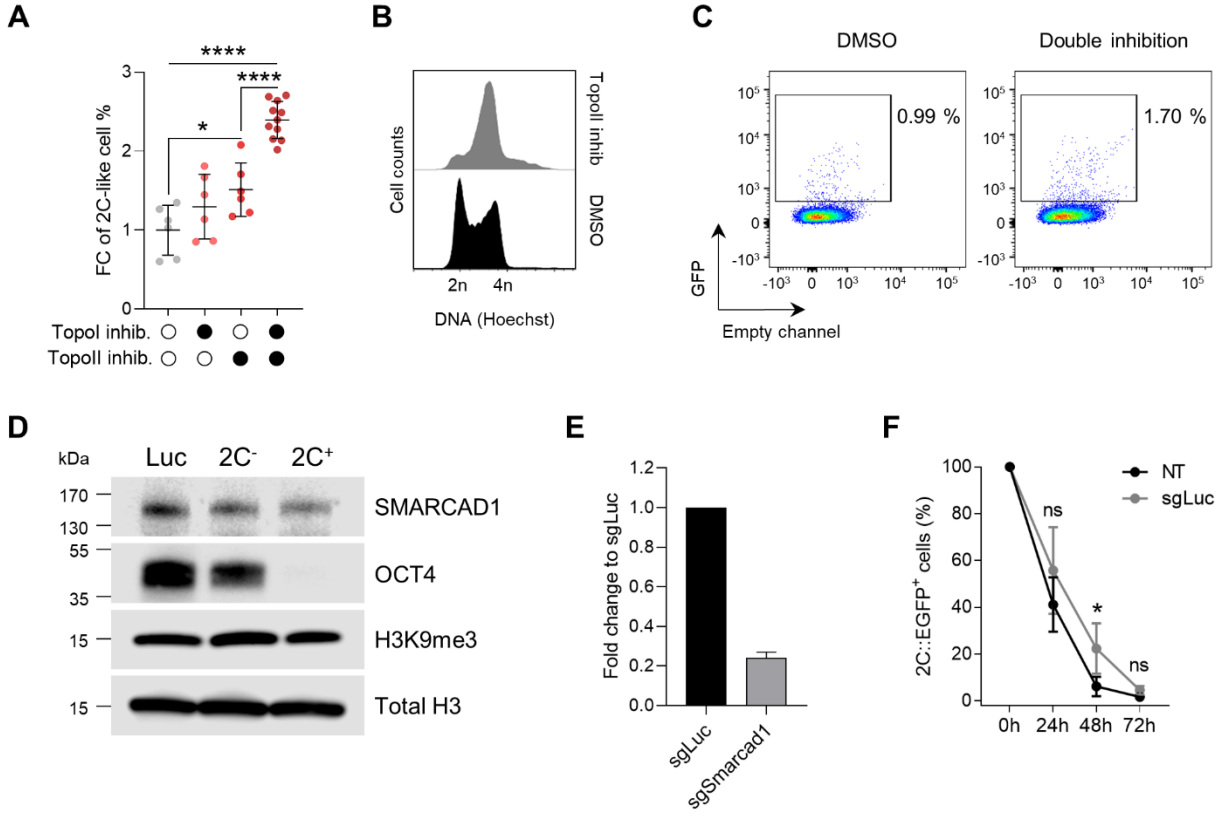
A
FC of 2C-like cell %
****
****
*
Topol inhib.
TopoII inhib.
B
Cell counts
TopoII inhib
DMSO
2n
4n
DNA (Hoechst)
C
DMSO
Double inhibition
0.99 %
1.70 %
GFP
Empty channel
D
kDa
Luc
2C-
2C+
170
130
55
35
15
15
SMARCAD1
OCT4
H3K9me3
Total H3
E
Fold change to sgLuc
sgLuc
sgSmarcad1
F
2C::EGFP+ cells (%)
NT
sgLuc
ns
*
ns
0h
24h
48h
72h

Fig. S3. Pharmacological and genetic perturbations in 2C-like cells. (**A**) Quantification of the percentage of $2C^+$ cells after inhibition of DNA topoisomerase I (TopoI inhib.), DNA topoisomerase II (TopoII inhib.) or its combined inhibition. Data are presented as scatter dot plots with line at mean ± SD ($n \geq 3$ independent experiments). $P = 0.0421$*, $P < 0.0001$**** by one-way ANOVA (Tukey's multiple comparisons test). (**B**) Cell cycle profile of DMSO and DNA topoisomerase II inhibited ESCs. (**C**) Representative FACS plots showing GFP^+ cells in DMSO and double DNA topoisomerase inhibition conditions. (**D**) Representative Western Blots for luciferase (Luc), $2C^-$ and $2C^+$ cells. SMARCAD1, OCT4, H3K9me3 and total histone H3 blots are shown. (**E**) qRT-PCR of *Smarcad1* in control luciferase (sgLuc) and *Smarcad1*-targeted (sgSmarcad1) $mCherry^+$ sorted ESCs 48 h after CRISPR-Cas9 sgRNA delivery. Data are presented as mean ± SD from two replicates transfected in independent rounds of sgRNA delivery. Independent sgRNAs targeting the same target gene were used in each round. (**F**) Quantification of the percentage of 2C-like cells 24 h, 48 h and 72 h after $2C^+$ cell sorting in non-transfected (NT) and luciferase-targeted (sgLuc) ESCs. Data are presented as mean ± SD. $P > 0.05^{ns}$, $P = 0.0463$* by two-way ANOVA (Sidak's multiple comparisons test).

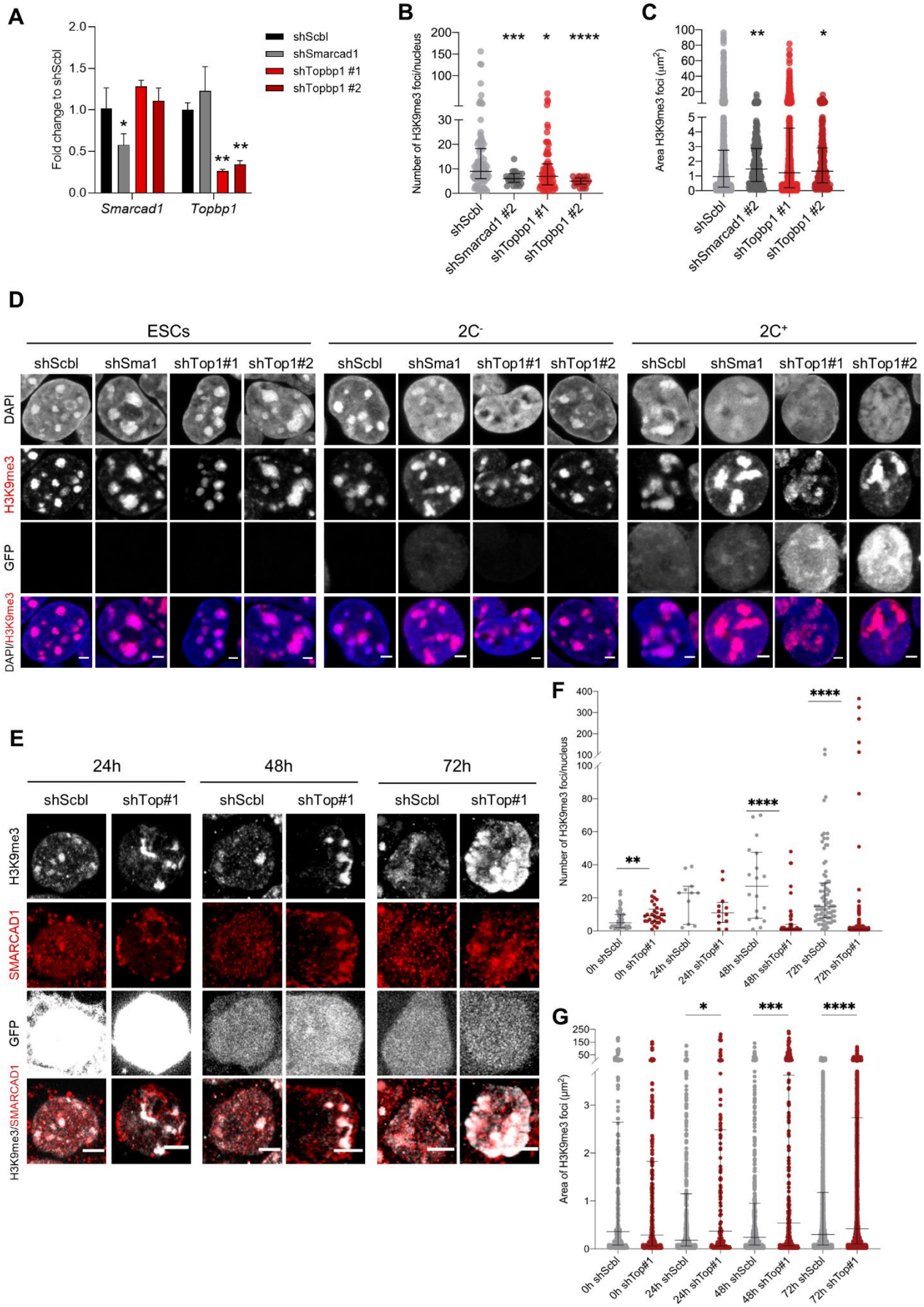
A
Fold change to shScbl
shScbl
shSmarcad1
shTopbp1 #1
shTopbp1 #2
Smarcad1
Topbp1
B
Number of H3K9me3 foci/nucleus
shScbl
shSmarcad1 #2
shTopbp1 #1
shTopbp1 #2
C
Area H3K9me3 foci (µm²)
D
ESCs
2C-
2C+
shScbl
shSma1
shTop1#1
shTop1#2
DAPI
H3K9me3
GFP
DAPI/H3K9me3
E
24h
48h
72h
shScbl
shTop#1
H3K9me3
SMARCAD1
GFP
H3K9me3/SMARCAD1
F
Number of H3K9me3 foci/nucleus
0h shScbl
0h shTop#1
24h shScbl
24h shTop#1
48h shScbl
48h sshTop#1
72h shScbl
72h shTop#1
G
Area of H3K9me3 foci (µm²)

Fig. S4. H3K9me3 foci analysis in *Smarcad1* and *Topbp1* knockdown ESCs, 2C$^-$ and 2C$^+$ cells. (**A**) qRT-PCR of *Smarcad1* and *Topbp1* in *Smarcad1* knockdown (shSmarcad1), *Topbp1* knockdown (shTopbp1 #1 and shTopbp1 #2) and control scramble (shScbl) ESCs. Data are presented as mean ± SD (n = 3 independent experiments). *P* = 0.0268*, *P* = 0.0010**, *P* = 0.0020** by one-way ANOVA (Dunnett's multiple comparisons test). (**B**) Quantification of the number of H3K9me3 foci in shScbl, shSmarcad1, shTopbp1 #1 and shTopbp1 #2 ESCs. Data are presented as scatter dot plots with line at median ± interquartile range (n > 3 independent cultures; shScbl = 90 cells, shSmarcad1 = 25 cells, shTopbp1 #1 = 81 cells and shTopbp1 #2 = 22 cells). *P* = 0.0009***, *P* = 0.0217*, *P* < 0.0001**** by Mann-Whitney test. (**C**) Quantification of H3K9me3 foci area in shScbl, shSmarcad1, shTopbp1 #1 and shTopbp1 #2 ESCs. Data are presented as scatter dot plots with line at median ± interquartile range (n > 3 independent cultures; shScbl = 1241 foci, shSmarcad1 = 281 foci, shTopbp1 #1 = 1004 foci and shTopbp1 #2 = 225 foci). *P* = 0.0016**, *P* = 0.2260^{ns}, *P* = 0.0103* by Mann-Whitney test. (**D**) Representative immunofluorescence images of H3K9me3 and the 2C::EGFP reporter during the ESCs to 2C$^+$ reprogramming in *Smarcad1* knockdown (shSmarcad1), *Topbp1* knockdown (shTopbp1 #1 and shTopbp1 #2) and control scramble (shScbl) cells. Scale bar, 2 μm. (**E**) Representative immunofluorescence images of H3K9me3, SMARCAD1 and the 2C::EGFP reporter during the 2C$^+$ exit (24 h, 48 h and 72 h) in *Topbp1* knockdown (shTopbp1 #1) and control scramble (shScbl) cells. Scale bar, 5 μm. (**F**) Quantification of the number of H3K9me3 foci in 2C$^+$ cells in shScbl and shTopbp1 #1 (shTop#1) samples at 0 h, 24 h, 48 h and 72 h after 2C-like state exit. Data are presented as scatter dot plots with line at median ± interquartile range (n = 2 independent cultures; shScbl 0 h = 49 cells, shTop#1 0 h = 30 cells, shScbl 24 h = 13 cells, shTop#1 24 h = 12 cells, shScbl 48 h = 18 cells, shTop#1 48 h = 48 cells, shScbl 72 h = 83 cells, shTop#1 72 h = 515 cells). *P* = 0.0086**, *P* = 0.1546^{ns}, *P* < 0.0001****, *P* < 0.0001**** by Mann-Whitney test. (**G)** Quantification of H3K9me3 foci area in 2C$^+$ cells in shScbl and shTopbp1 #1 (shTop#1) samples at 0 h, 24 h, 48 h and 72 h after 2C-like state exit. Data are presented as scatter dot plots with line at median ± interquartile range (n = 2 independent cultures; shScbl 0 h = 342 foci, shTop#1 0 h = 304 foci, shScbl 24 h = 481 foci, shTop#1 24 h = 158 foci, shScbl 48 h = 537 foci, shTop#1 48 h = 271 foci, shScbl 72 h = 2634 foci, shTop#1 72 h = 3477 foci). *P* = 0.1037^{ns}, *P* = 0.0163*, *P* = 0.0002***, *P* < 0.0001**** by Mann-Whitney test.

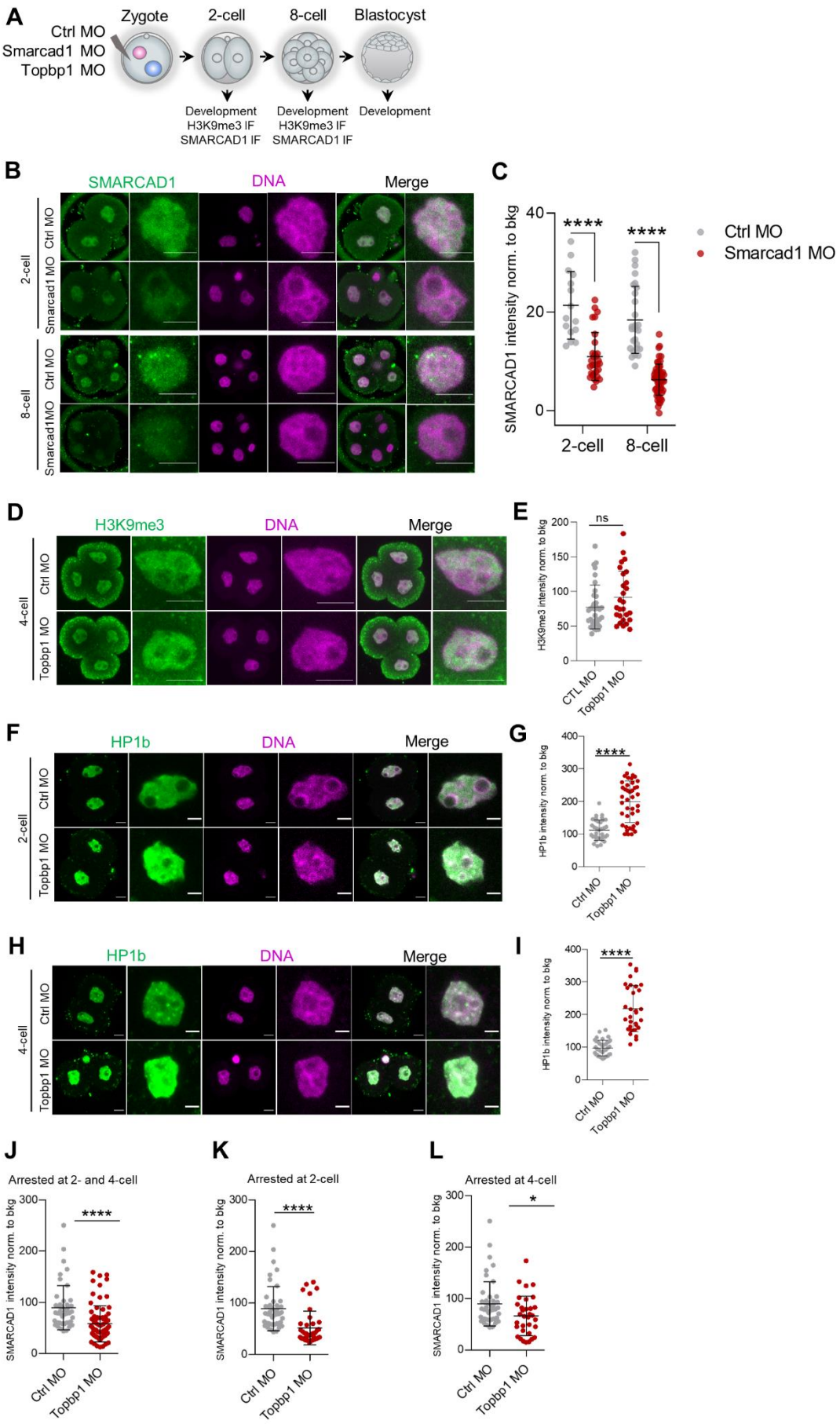
A
Ctrl MO
Smarcad1 MO
Topbp1 MO
Zygote
2-cell
8-cell
Blastocyst
Development
H3K9me3 IF
SMARCAD1 IF
Development
H3K9me3 IF
SMARCAD1 IF
Development
B
SMARCAD1
DNA
Merge
2-cell
8-cell
Ctrl MO
Smarcad1 MO
Smarcad1MO
C
SMARCAD1 intensity norm. to bkg
****
****
Ctrl MO
Smarcad1 MO
2-cell
8-cell
D
H3K9me3
DNA
Merge
4-cell
Ctrl MO
Topbp1 MO
E
H3K9me3 intensity norm. to bkg
ns
CTL MO
Topbp1 MO
F
HP1b
DNA
Merge
2-cell
Ctrl MO
Topbp1 MO
G
HP1b intensity norm. to bkg
****
Ctrl MO
Topbp1 MO
H
HP1b
DNA
Merge
4-cell
Ctrl MO
Topbp1 MO
I
HP1b intensity norm. to bkg
****
Ctrl MO
Topbp1 MO
J
Arrested at 2- and 4-cell
SMARCAD1 intensity norm. to bkg
****
Ctrl MO
Topbp1 MO
K
Arrested at 2-cell
SMARCAD1 intensity norm. to bkg
****
Ctrl MO
Topbp1 MO
L
Arrested at 4-cell
SMARCAD1 intensity norm. to bkg
*
Ctrl MO
Topbp1 MO

Fig. S5. SMARCAD1 downregulation in mouse embryos. (**A**) Schematic representation of the experimental design to assess SMARCAD1 and TOPBP1 function in early mouse embryo development. Morpholino antisense oligos (MO) targeting *Smarcad1*, *Topbp1* and a scrambled control (Ctrl) sequence were microinjected into the cytoplasm of zygotes (E0.5 embryos). Embryo development was monitored daily from the 2-cell stage (E1.5) until the late blastocyst stage (E5.5). (**B**) Representative immunofluorescence images of SMARCAD1 in Ctrl and *Smarcad1* MO embryos at 2-cell (E1.5) and 8-cell stage (E2.5) embryos. Representative blastomere nuclei are shown. Scale bar, 5 µm. (**C**) Quantification of SMARCAD1 mean fluorescence intensity in Ctrl and *Smarcad1*MO embryos at 2-cell (E1.5) and 8-cell stage (E2.5). Data are presented as scatter dot plots with line at mean ± SD (2-cell: Ctrl MO = 8 embryos, *Smarcad1* MO = 15 embryos; 8-cell: Ctrl MO = 11 embryos, *Smarcad1* MO = 15 embryos). SMARCAD1 signal was normalized to the average background signal. $P < 0.0001$**** by unpaired two-tailed Student's *t*-test. (**D**) Representative immunofluorescence images of H3K9me3 in Ctrl and *Topbp1* MO embryos at 4-cell (E2.0) embryos. Representative blastomere nuclei are shown. Scale bar, 5 µm. (**E**) Quantification of H3K9me3 mean fluorescence intensity in Ctrl and *Topbp1* MO embryos at 4-cell (E 2.0) embryos. Data are presented as scatter dot plots with line at mean ± SD (Ctrl MO = 35 embryos, *Topbp1* MO = 27 embryos). H3K9me3 signal was normalized to the average background signal. (**F**) Representative immunofluorescence images of HP1b in Ctrl and *Topbp1* MO embryos at 2-cell (E1.5) embryos. Representative blastomere nuclei are shown. Scale bar, 10 µm. (**G**) Quantification of HP1b mean fluorescence intensity in Ctrl and *Topbp1* MO embryos at 2-cell (E1.5). Data are presented as scatter dot plots with line at mean ± SD (Ctrl MO = 20 embryos, *Topbp1* MO = 21 embryos). HP1b signal was normalized to the average background signal. $P < 0.0001$**** by unpaired two-tailed Student's *t*-test. (**H**) Representative immunofluorescence images of HP1b in Ctrl and *Topbp1* MO embryos at 4-cell (E2.0) embryos. Representative blastomere nuclei are shown. Scale bar, 10 µm. (**I**) Quantification of HP1b mean fluorescence intensity in Ctrl and *Topbp1* MO embryos at 4-cell (E2.0). Data are presented as scatter dot plots with line at mean ± SD (Ctrl MO = 16 embryos, *Topbp1* MO = 13 embryos). HP1b signal was normalized to the average background signal. $P < 0.0001$**** by unpaired two-tailed Student's *t*-test. (**J**) Quantification of SMARCAD1 mean fluorescence intensity in Ctrl and *Topbp1* MO embryos arrested at 2-cell and 4-cell stage (E2.0). Data are presented as scatter dot plots with line at mean ± SD (Ctrl MO = 20 embryos, *Topbp1* MO = 31 embryos). SMARCAD1 signal was normalized to the average background signal. $P < 0.0001$**** by unpaired two-tailed Student's *t*-test. (**K**) Quantification of SMARCAD1 mean fluorescence intensity in Ctrl and *Topbp1* MO embryos arrested at 2-cell stage. Data are presented as scatter dot plots with line at mean ± SD (Ctrl MO = 20 embryos, *Topbp1* MO = 21 embryos). SMARCAD1 signal was normalized to the average background signal. $P < 0.0001$**** by unpaired two-tailed Student's *t*-test. (**L**) Quantification of SMARCAD1 mean fluorescence intensity in Ctrl and *Topbp1* MO embryos arrested at 4-cell stage. Data are presented as scatter dot plots with line at mean ± SD (Ctrl MO = 20 embryos, *Topbp1* MO = 11 embryos). SMARCAD1 signal was normalized to the average background signal. $P < 0.0174$* by unpaired two-tailed Student's *t*-test.