


RESEARCH

Open Access



# Elevated retrotransposon activity and genomic instability in primed pluripotent stem cells

Haifeng Fu<sup>1,2†</sup>, Weiyu Zhang<sup>1,4†</sup>, Niannian Li<sup>1,2†</sup>, Jiao Yang<sup>1,2†</sup>, Xiaoying Ye<sup>1,2</sup>, Chenglei Tian<sup>1,2</sup>, Xinyi Lu<sup>1,4</sup> and Lin Liu<sup>1,2,3\*</sup> 

\* Correspondence: [liulin@nankai.edu.cn](mailto:liulin@nankai.edu.cn)

Haifeng Fu, Weiyu Zhang, Niannian Li and Jiao Yang are Co-first authors

<sup>1</sup>State Key Laboratory of Medicinal Chemical Biology, Nankai University, Tianjin, China

<sup>2</sup>Department of Cell Biology and Genetics, College of Life Sciences, Nankai University, Tianjin, China

Full list of author information is available at the end of the article

## Abstract

**Background:** Naïve and primed pluripotent stem cells (PSCs) represent two different pluripotent states. Primed PSCs following in vitro culture exhibit lower developmental potency as evidenced by failure in germline chimera assays, unlike mouse naïve PSCs. However, the molecular mechanisms underlying the lower developmental competency of primed PSCs remain elusive.

**Results:** We examine the regulation of telomere maintenance, retrotransposon activity, and genomic stability of primed PSCs and compare them with naïve PSCs. Surprisingly, primed PSCs only minimally maintain telomeres and show fragile telomeres, associated with declined DNA recombination and repair activity, in contrast to naïve PSCs that robustly elongate telomeres. Also, we identify LINE1 family integrant *L1Md\_T* as naïve-specific retrotransposon and ERVK family integrant *IAPEz* to define primed PSCs, and their transcription is differentially regulated by heterochromatic histones and Dnmt3b. Notably, genomic instability of primed PSCs is increased, in association with aberrant retrotransposon activity.

**Conclusions:** Our data suggest that fragile telomere, retrotransposon-associated genomic instability, and declined DNA recombination repair, together with reduced function of cell cycle and mitochondria, increased apoptosis, and differentiation properties may link to compromised developmental potency of primed PSCs, noticeably distinguishable from naïve PSCs.

**Keywords:** Naïve and primed pluripotent state, Telomeres, 2C genes, Genome stability, Retrotransposons, Histone modifications

## Background

Murine embryonic stem cells (mESCs) from preimplantation embryos resemble an earlier stage in development known as naïve pluripotent state. Primed epiblast stem cells (mEpiSCs) are derived from murine post-implantation epiblast embryos commonly known to be in primed pluripotent state [1, 2]. Human ESCs derived from preimplantation embryos are transcriptionally similar to mEpiSCs [2–4]. Naïve mESCs can be sustained in the



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

presence of serum and leukemia inhibitory factor (LIF) condition [5], while the mEpiSCs can be maintained in the presence of FGF2 and Activin A condition [1, 2]. Naïve mouse pluripotent stem cells (mPSCs) exhibit germline competence as determined by chimera production test and can generate all-PSC mice by tetraploid embryo complementation (TEC) test, the most stringent functional assay of naïve pluripotency [6–9]. In contrast, primed mEpiSCs fail in germline competence and even rarely produce chimeras [1, 2], suggesting their reduced developmental and differentiation capacity. Under defined *in vitro* culture conditions, the primed mouse epiblast-like cells (mEpiLCs) can be established from naïve mESCs [10–12]. Naïve and primed pluripotent states exhibit different molecular signatures in the epigenome and transcriptome profile [3, 10, 11, 13–18].

Moreover, a “ground state” beyond naïve state has been achieved by culture under 2i/LIF condition (2i, GSK3, and MEK1/2 inhibitors) [19]. 2i reduces the DNA methyltransferases Dnmt3a/3b to promote hypomethylation and enhances Tet1/2 activity (consequently, increased 5-hydroxymethylcytosine (5hmC) levels) and “passive” loss of DNA methylation, resulting in a distinct transcriptional and epigenetic state that includes uniform expression of key pluripotency factors, such as Nanog and Prdm14 [20, 21]. Ground-state cells exhibit lower expression of lineage affiliated genes, reduced prevalence at promoters of the repressive histone modification H3K27me3, and fewer bivalent domains, which are thought to mark genes poised for either up- or downregulation [22]. Furthermore, female PSCs under the ground state or naïve state show both X chromosome activation, but inactivate one of the X chromosomes under primed state [18]. mESCs under 2i culture conditions also can produce all-ESC pups but prolonged inhibition with 2i could damage the developmental potential [23–25].

Notably, transposable elements, particularly endogenous retroviruses (ERVs) that have important roles in pluripotency, can wire the pluripotency network in humans and mice [26–33]. Naïve mouse ESCs periodically activate endogenous retrovirus (ERVs) and 2 cell (2C) genes [28]. ERVs together with other retrotransposons, including long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) occupy approximately 40% of mammalian genomes, and mostly are conserved and evolved [34–36]. Nevertheless, retrotransposition of active LINEs and ERVs could induce genomic instability, disrupt regulatory elements, cause mutations, or drive genome evolution [34, 37]. Therefore, the repression of retrotransposons is critical for the maintenance of genome stability in ESCs [38]. For instance, Kap1 and histone 3.3 repress ERVs in mESCs [38–40]. In addition, telomere elongation and maintenance are essential for unlimited self-renewal and pluripotency of PSCs [41, 42]. Short telomeres impair differentiation capacity of ESCs [43]. Yet, regulation of transposable elements and telomere length in primed PSCs remains largely unclear.

In this study, we sought to delineate telomere and retrotransposon features association with the genomic stability of primed PSCs in comparison with those of naïve PSCs in mice, given that the two states have been functionally defined by developmental pluripotency *in vivo*.

## Results

### Conversion of naïve mESCs to primed mEpiSCs results in repression of DNA recombination repair pathway

Naïve mESCs with distal Oct4 GFP fluorescence were maintained in serum/LIF medium on mitomycin C-inactivated MEFs served as feeder cells. The GFP reporter

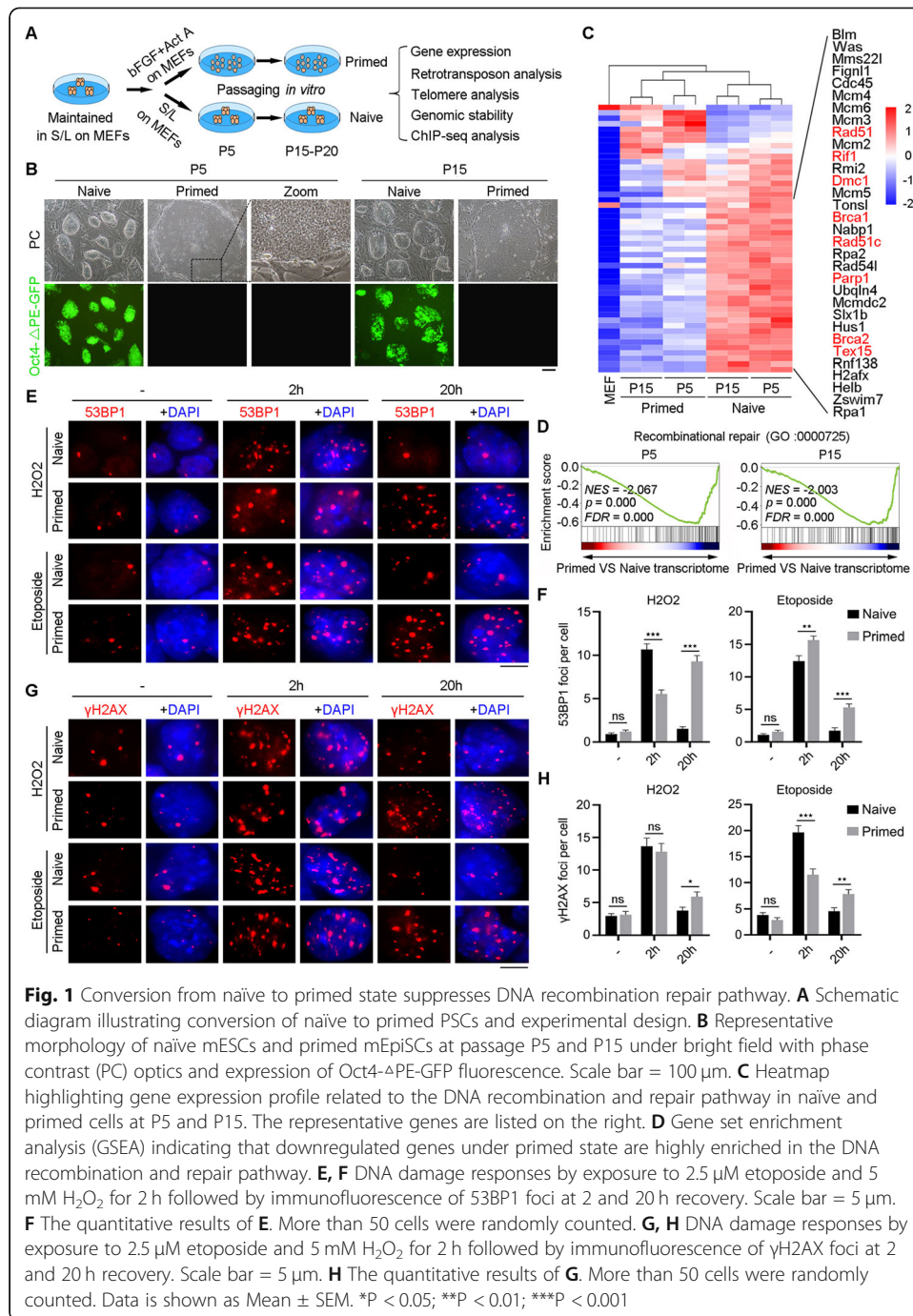
fluorescence driven by distal *Oct4* promoter and enhancer can indicate the naïve state of pluripotency [4, 44–46]. We converted naïve mESCs into the primed state by culture in F/A medium (Egf2 + ActivinA) based on the method described previously [1, 11, 18, 47]. Primed mEpiSCs also were cultured on feeder cells, like naïve mESCs [25]. Domed shaped naïve ES colonies began to flatten within 48 h following culture in F/A medium, as shown in a previous report [11].

To achieve stable state, we expanded the cells for five passages (P5, relatively early passages) and also for additional 10 or 15 passages (P15 or P20, relatively late passages) under naïve and primed conditions (Fig. 1a). Primed mEpiSCs grew as flat but larger colonies and were thus morphologically distinctive from naïve mESCs showing compact, dome-shaped but smaller colonies (Fig. 1b; and Additional file 1: Figure S1a). The GFP reporter fluorescence indicating the naïve state of pluripotency extinguished following conversion into primed state (Fig. 1b). Analysis of cell cycle at P5 showed that cells under primed state displayed extended G1 phase (Additional file 1: Figure S1b), in contrast to naïve mESCs with shorter G1 phase [48, 49]. The primed cells expressed two core pluripotent factors *Oct4* and *Nanog*, but at levels lower than those of naïve cells (Additional file 1: Figure S1c, e, f), as shown previously [11, 12, 18]. Additionally, expression of an important naïve marker *SSEA1* was reduced in primed cells (Additional file 1: Figure S1e, f), whereas *Dnmt3b*, a marker gene of primed state [2], was upregulated in primed cells (Additional file 1: Figure S1d, g). By RNA-seq analysis, naïve mESCs highly expressed known naïve marker genes such as *Rex1*, *Stella*, *Esrrb*, *Tbx3*, and *Nr0b1* [3], while marker genes for primed state such as *Otx2*, *Oct6*, *Fgf5*, *T*, *Cer1*, and *Pitx2* [2] were upregulated at either passage 5 or 15 (Additional file 1: Figure S1h). Our RNA-seq analysis also revealed differential global transcription profiles of the two pluripotent states (Additional file 1: Figure S1i; Additional file 2: Table S1; Additional file 3: Table S2,  $\text{Padj} < 0.01$ , fold change  $\geq 2$ ).

To validate the developmental potential of mouse naïve and primed cells, we performed chimera formation and tetraploid embryo complementation (TEC) assays. Primed mEpiSCs rarely can form chimeras (Additional file 1: Figure S2a, b, c), while naïve mESCs generated germline chimera mice and completely ESC pups by TEC assay (Additional file 1: Figure S2d, e, f) that were fertile (Additional file 1: Figure S2g). These results further support the notion that the developmental pluripotency of naïve PSCs is high, distinct from primed PSCs.

Naïve and primed PSCs exhibit different gene expression profiles and signaling pathways [3, 18]. By Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis, the most upregulated genes in naïve cells at P5 and P15 were enriched in signaling pathways regulating pluripotent stem cells (Additional file 1: Figure S3a, b), demonstrating that molecular pluripotency of naïve cells is indeed higher than that of primed cells. Signaling pathways that affect pluripotency such as MAPK, Wnt, and PI3K-Akt, and cell adhesion molecules (CAMs) and cytokine-cytokine receptor interaction also differ between the two states (Additional file 1: Figure S3c, d). Moreover, the transcription programs of naïve and primed cells resemble their in vivo counterparts ICM and post-implantation epiblasts respectively in terms of cell cycle and apoptosis pathway (Additional file 1: Figure S3e, f).

Remarkably, by gene ontology (GO) analysis, DNA recombinational repair pathway was significantly downregulated under primed state (Fig. 1c, d). Expression levels of



**Fig. 1** Conversion from naive to primed state suppresses DNA recombination repair pathway. **A** Schematic diagram illustrating conversion of naive to primed PSCs and experimental design. **B** Representative morphology of naive mESCs and primed mEpiSCs at passage P5 and P15 under bright field with phase contrast (PC) optics and expression of Oct4-ΔPE-GFP fluorescence. Scale bar = 100 μm. **C** Heatmap highlighting gene expression profile related to the DNA recombination and repair pathway in naive and primed cells at P5 and P15. The representative genes are listed on the right. **D** Gene set enrichment analysis (GSEA) indicating that downregulated genes under primed state are highly enriched in the DNA recombination and repair pathway. **E, F** DNA damage responses by exposure to 2.5 μM etoposide and 5 mM H<sub>2</sub>O<sub>2</sub> for 2 h followed by immunofluorescence of 53BP1 foci at 2 and 20 h recovery. Scale bar = 5 μm. **F** The quantitative results of **E**. More than 50 cells were randomly counted. **G, H** DNA damage responses by exposure to 2.5 μM etoposide and 5 mM H<sub>2</sub>O<sub>2</sub> for 2 h followed by immunofluorescence of γH2AX foci at 2 and 20 h recovery. Scale bar = 5 μm. **H** The quantitative results of **G**. More than 50 cells were randomly counted. Data is shown as Mean ± SEM. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001

*Rif1*, *Rad51*, *Dmc1*, *Brca1*, and *Brca2* were remarkably decreased in primed cells (Fig. 1c), and these genes are involved in double-strand break (DSB) repair process or telomere maintenance [50–52]. Mismatch repair and nucleotide-excision repair pathways were also repressed in primed, compared with naive mESCs, but non-homologous end joining (NHEJ) and base-excision repair pathways did not differ between the two states (Additional file 1: Figure S4a-h). These data imply the more robust DNA repair capacity likely suggestive of better genomic stability maintenance in naive mESCs compared with those of primed mEpiSCs.

To assess whether primed cells are more susceptible to DNA damage response, we performed immunofluorescence microscopy of  $\gamma$ H2AX and 53BP1, commonly used as DNA damage response markers [50, 53] in naïve and primed state PSCs by exposure to low dose of either etoposide or  $H_2O_2$ . We intuitively monitored 53BP1 and  $\gamma$ H2AX foci by immunofluorescence in etoposide or  $H_2O_2$ -treated cells. Increased 53BP1 and  $\gamma$ H2AX foci appeared at 2 h in both cell types, and the number of foci was significantly reduced likely indicative of DNA repair after 20 h recovery in naïve cells, and almost similar to that without exposure to the DNA damage at 0 h (Fig. 1e–h; Additional file 1: Figure S1j). However, more 53BP1 and  $\gamma$ H2AX foci persisted in primed cells than in naïve cells at 20 h (Fig. 1e–h; Additional file 1: Figure S1j). These data conclude that the DNA repair capacity is robust in naïve cells and is reduced in primed state.

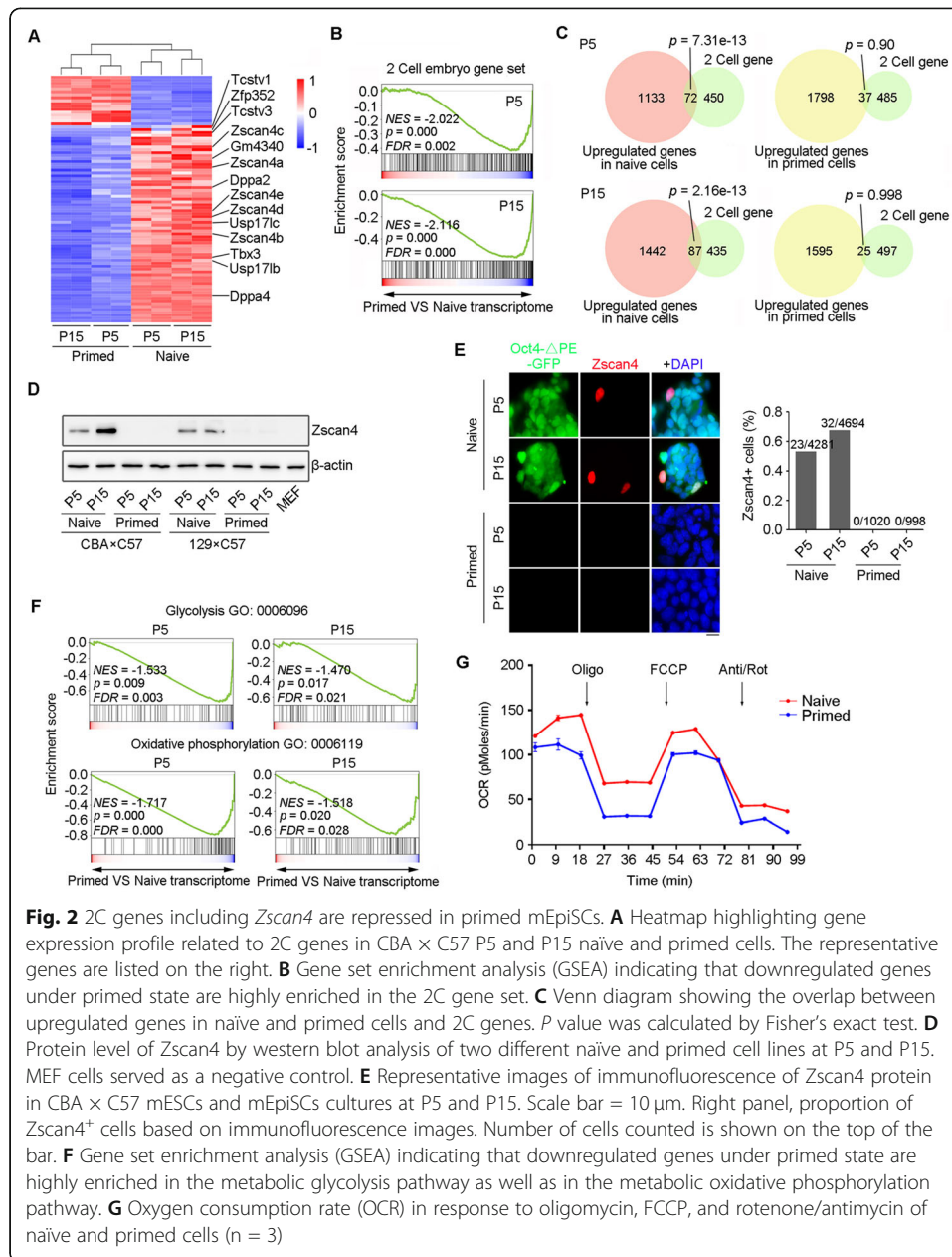
### Two-cell (2C) embryo genes are repressed in primed mEpiSCs

Furthermore, representative 2C genes, e.g., *Zscan4*, *Tcstv1/3*, *Usp17l*, *Dppa2*, and *Dppa4*, that were highly upregulated in naïve mESCs, were noticeably downregulated in primed mEpiSCs (Fig. 2a). We took advantage of previously reported lists of upregulated genes in 2C-like ESCs [28], and compared expression levels of 2C genes in naïve and primed cells. Conversion to primed state led to decreased enrichment of 2C embryo gene sets regardless of passages (Fig. 2b). 2C-like genes significantly overlapped with the upregulated genes in naïve cells compared with primed cells ( $p = 7.31e-13$  for P5 and  $p = 2.16e-13$  for P15), but showed no correlation with the upregulated genes in primed cells ( $p = 0.90$  for P5 and  $p = 0.998$  for P15) (Fig. 2c). These data suggested that 2C genes were repressed in primed mEpiSCs. Western blot and immunofluorescence microscopy confirmed sporadic expression of *Zscan4* protein in naïve but not in primed cells (Fig. 2d, e). These results were also repeated in another independent cell line at 129  $\times$  C57 genetic background (Additional file 1: Figure S5a, b). In addition, RNA-seq data and seahorse experiment further confirmed that the metabolic pathways of glycolysis and oxidative phosphorylation were decreased during the transition from naïve to primed state (Fig. 2f, g). Therefore, unlike naïve mESCs, primed mEpiSCs lack a 2C-like cell subpopulation.

### Primed PSCs do not elongate telomeres

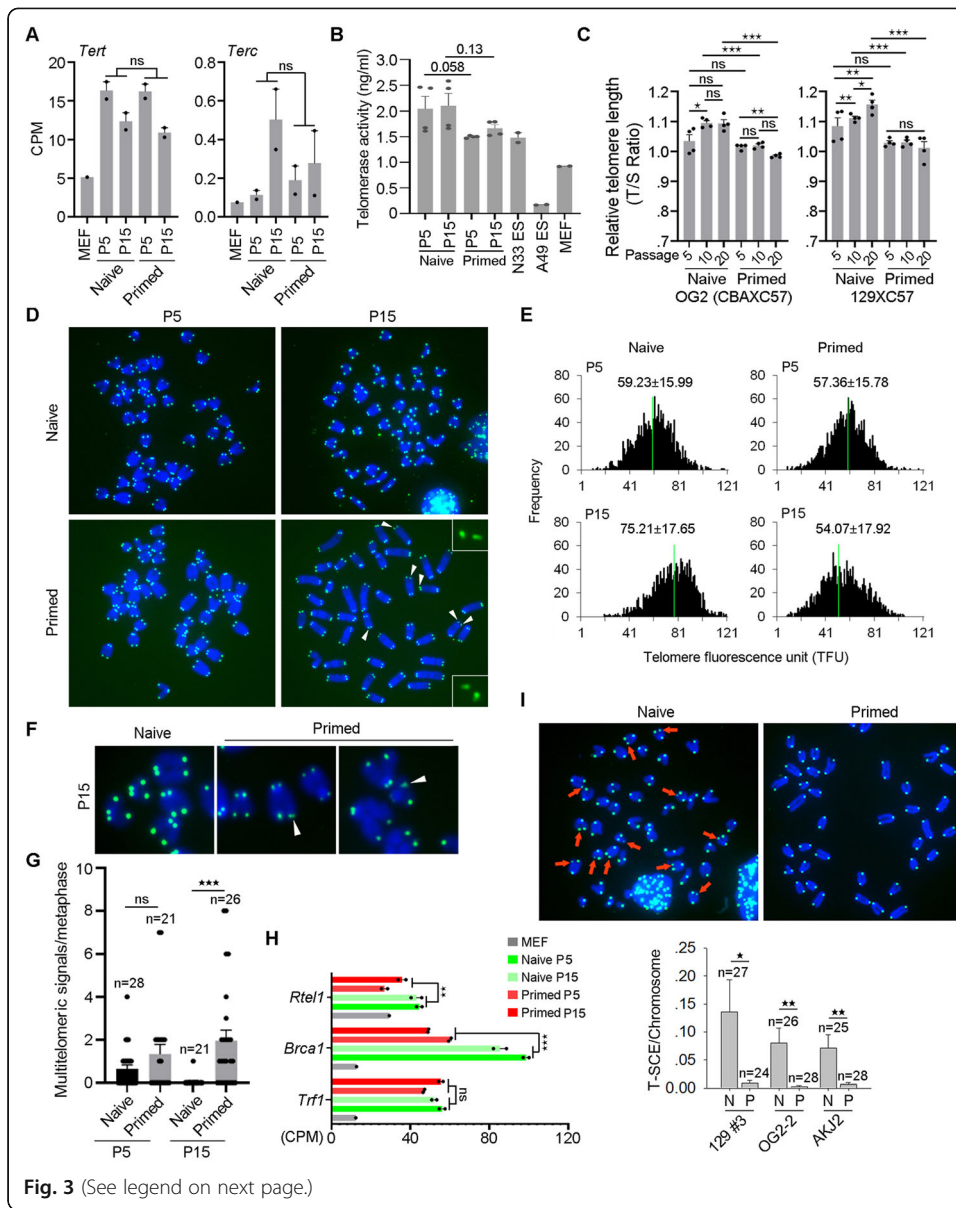
Primed PSCs do not express 2C genes and notably *Zscan4*, in contrast to naïve PSCs. *Zscan4* has been shown to be critical for telomere lengthening and maintaining the genomic stability of mESCs by telomere sister chromatid exchange (T-SCE)-based homologous recombination [52]. This prompted us to explore regulation of telomere maintenance of primed cells, compared with naïve cells. Telomerase genes *Tert* and *Terc* were expressed at similarly higher levels in naïve and primed cells, relative to MEF cells (Fig. 3a). In consistency, both pluripotent state cells expressed similarly high telomerase activity (Fig. 3b; Additional file 1: Figure S5c). Dynamics of relative telomere length revealed by qPCR, shown as telomere/single-copy gene (T/S) ratio demonstrated that telomeres lengthened in naïve mESCs but not in primed PSCs with increasing passages (Fig. 3c). By telomere Q-FISH, telomeres also lengthened significantly in naïve mESCs with passages, whereas primed mEpiSCs failed to elongate their telomeres but





exhibited slightly shortened telomeres (Fig. 3d, e). Similar results were obtained in another cell line from 129 × C57 genetic background (Additional file 1: Figure S5d, e, f).

Notably, following relatively longer-term culture, primed cell lines at a later passage exhibited fragile telomeres (Fig. 3f, g; Additional file 1: Figure S5g). TRF1, RTEL1, and BRCA1 were reported to be involved in the formation of fragile sites [54–56]. *Trf1* expression level did not change between the two states, but *Rtel1* and *Brca1* expression levels were decreased in the primed cells compared to naive cells (Fig. 3h). This, along with the declined DNA recombination repair capacity, might be linked to the fragile telomeres in primed mEpiSCs after longer passage. The fragile telomeres and declined recombination pathways are associated with minimal telomere maintenance in primed



**Fig. 3** (See legend on next page.)

(See figure on previous page.)

**Fig. 3** Distinct telomere maintenance in naïve and primed cells. **A** Expression levels (CPMs) of telomerase genes *Tert* and *Terc* in CBA × C57 naïve and primed cells at P5 and P15. Two biological repeats. **B** Quantification of telomerase activity of two different naïve and primed cell lines P5 and P15 by ELISA assay. WT N33 ESCs served as telomerase positive controls and telomerase-deficient G4 A49 ESCs served as negative controls. **C** Relative telomere length shown as T/S ratio by qPCR during passages (P5, P10, and P20) in naïve and primed cells in CBA × C57 or in 129 × C57 background. Four replicates are shown. \*P < 0.05; \*\*P < 0.01, \*\*\*P < 0.001. **D** Representative images displaying telomere FISH of naïve and primed cells at CBA × C57 background at P5 and P15. Blue, chromosomes stained with DAPI; green dots, telomeres. The white arrows indicate fragile telomeres. **E** Histogram shows distribution of relative telomere length displayed as TFU by Q-FISH analysis. Green line indicates medium telomere length. Mean ± s.d. of telomere length is shown above each panel. About 20 chromosome spreads were quantified for each group. Two independent experiments. **F** Enlarged telomere Q-FISH images, showing telomere fragility (white arrowhead) in primed cells but not in naïve cells. Telomeres were labeled with telomere PNA probes (green), and chromosomes labeled with DAPI (blue). **G** Frequency of telomere fragility per chromosome. n, number of spreads counted. Two independent experiments. **H** Expression levels (CPMs) of *Trf1*, *Brca1*, and *Rtel1* genes in MEF cells, naïve, and primed cells. Two biological repeats. **I** Representative micrographs showing telomere sister chromatid exchange (T-SCE, red arrows) in naïve and primed cells in CBA × C57 background by CO-FISH analysis. Bottom panel, frequency of T-SCE in the naïve and primed cell lines from three different genetic background. Two independent experiments. Data is shown as Mean ± SEM. \*P < 0.05; \*\*P < 0.01; \*\*\*P < 0.001

cells. These data indicated that vigorous telomere elongation takes place in naïve cells but not in primed cells.

Telomerase extends telomeres at a slow pace of about 50–100 nucleotides per cell cycle [57]. Although telomerase activity was high in both states, the difference in telomere length dynamics suggests that mechanism other than telomerase activity could be involved in telomere maintenance of these cells. Alternative lengthening of telomeres (ALT) is a telomerase-independent mechanism to elongate telomeres rapidly [58], in association with frequent telomere sister chromatic exchange (T-SCE), which can indicate the recombinational rate of telomeres [59–61]. We assessed T-SCE in naïve and primed cells by telomere chromosome orientation FISH (CO-FISH) analysis [59]. Frequency of T-SCE was extremely low in primed cells (Fig. 3i; Additional file 1: Figure S5h), associated with the declined expression level of *Dmc1* which is involved in telomere recombination process [52] (Fig. 1c).

To further explore whether the high expression of DNA recombination repair genes links to telomere maintenance, we knocked out *Brca1*, *Dmc1*, and *Rad51* in naïve cells respectively by CRISPR-Cas9 and measured telomeres (Additional file 1: Figure S6a, b). Cell morphology did not change much after knocking out each of these genes (Additional file 1: Figure S6c). Expression of pluripotent gene *Oct4* and 2C gene *Zscan4* also was not affected following knockout (KO) (Additional file 1: Figure S6d). Telomeres significantly shortened after *Dmc1* KO and slightly shortened after *Rad51* KO (Additional file 1: Figure S6e, f). Although telomere length slightly increased after *Brca1* KO, telomere fragility increased (Additional file 1: Figure S6e-h). Hence, higher expression levels of DNA recombination repair genes in naïve state facilitate telomere lengthening of mESCs, whereas low expression levels of DNA recombination repair genes in primed state may reduce telomere elongation.

DNA demethylation also could directly affect recombination level at telomere region, and increased DNA methylation level may inhibit telomere elongation [62, 63]. Expression levels of DNMTs varied in naïve and primed cells. *Dnmt3l* and *Dnmt3b*, as the naïve and primed marker genes [64, 65], were highly expressed in naïve and primed



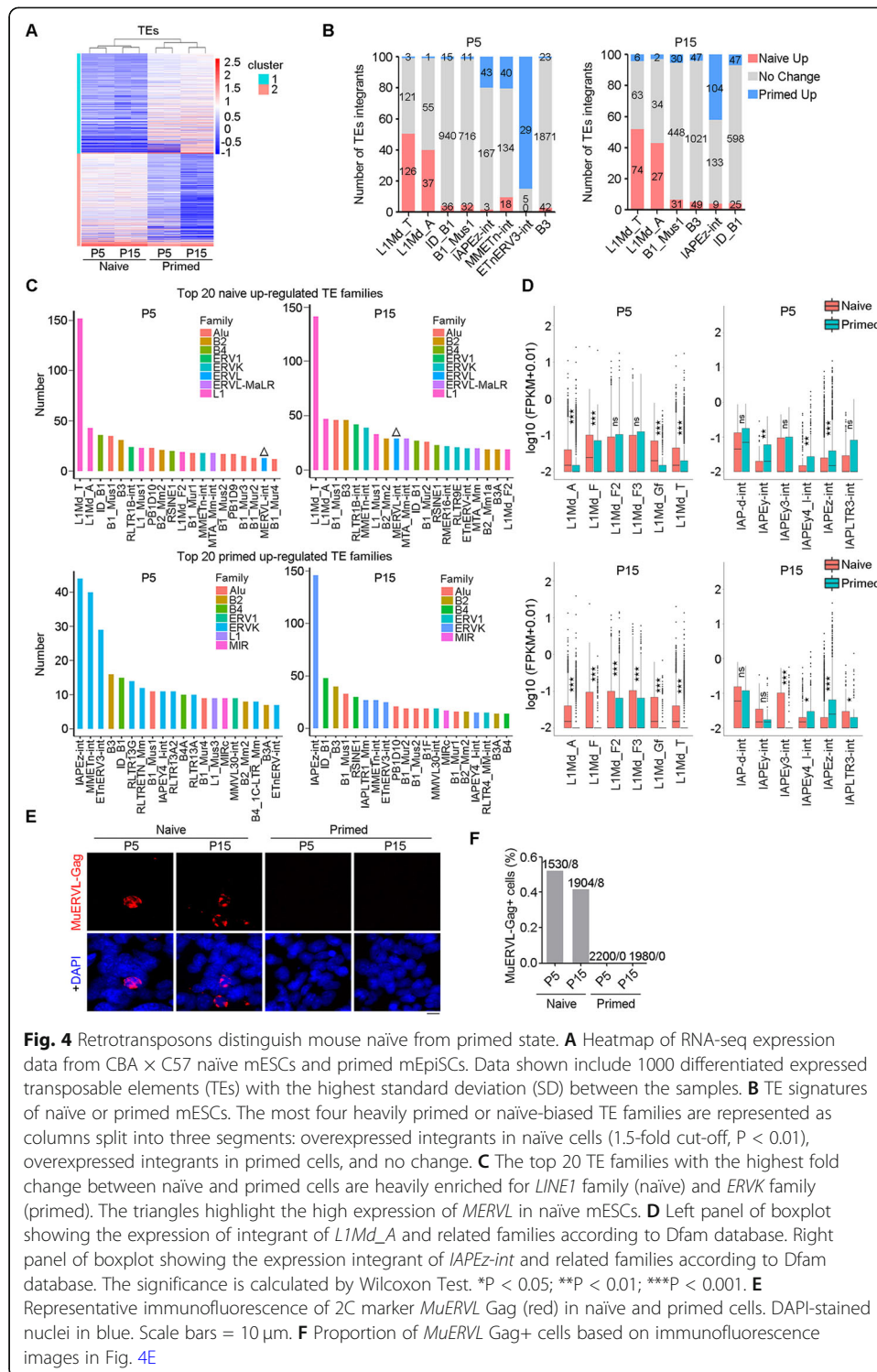
cells, respectively (Additional file 1: Figure S 7a). *Dnmt1*, *Uhrf1*, and *Dnmt3a* all were downregulated in primed cells (Additional file 1: Figure S7a). Expression levels of *Tet1* and *Tet2* were decreased, but *Tet3* increased in primed cells compared with naïve cells (Additional file 1: Figure S7a). Tet family proteins oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), an intermediate that can lead to DNA demethylation [66, 67]. By immunofluorescence microscopy and dot blot assay, 5mC level did not change, but 5hmC level was decreased in primed cells (Additional file 1: Figure S7b-d). Differences in DNA methylation/demethylation levels may also regulate telomeres of naïve and primed cells. Increased telomere recombination activity corroborates robust telomere lengthening in naïve PSCs.

### Retrotransposon transcription distinguishes primed from naïve PSCs

Furthermore, we compared the transcription of transposable elements (TEs) in naïve and primed cells using an improved RNA-sequencing analysis pipeline [68]. The top 1000 TEs with the largest standard deviation (SD) separated naïve from primed cells (Fig. 4a). We looked at which TEs were overexpressed in naïve and primed cells, respectively. Top 20 TE families that were highly expressed in naïve cells at P5 and P15 included multiple types of retrotransposons but no DNA repeats (Fig. 4c). Among them, *L1* family members, especially *L1Md\_Ts* were prominently upregulated in naïve cells compared with primed cells (Fig. 4b–d). Other *L1* family members, such as *L1Md\_A*, *L1\_Mus1*, and *L1Md\_F2* were also upregulated in naïve mESCs (Fig. 4b, c). Among all 129 differentially transcribed *L1Md\_Ts* at P5, 126 of them were upregulated in naïve mESCs, in contrast to only three of them upregulated in primed mEpiSCs. Similarly, 74 of 80 differentially expressed *L1Md\_Ts* were upregulated in naïve cells at P15, but only 6 were upregulated in primed cells (Fig. 4b; Additional file 4: Table S3; Additional file 5: Table S4). Other retrotransposons, such as *SINEs* including *Alu* family integrant *B1\_Mus1*, *B4* family integrant *ID\_B1*, and *B2* family integrant *B3*, also were elevated to various degrees in naïve cells (Fig. 4c). *MERVL*, which reportedly is highly expressed in naïve mESCs and can mark 2C-like state [28], was not the most significantly increased retrotransposons in comparison with those of primed cells (Fig. 4c). This might be because *MERVL* is only highly expressed in naïve 2C-like subpopulations, compromising its total expression levels in a large ESC population.

Moreover, all of the top 20 TE families upregulated in primed cells also were retrotransposons (Fig. 4c). Members of the *ERVK* family, including *MMETn-int* and *ETnERV3-int*, and particularly *IAPez-int*, were transcribed almost exclusively in primed cells (Fig. 4b–d). Among all 46 differentially transcribed *IAPez-ints*, 43 were upregulated in primed mEpiSCs at P5. This trend was even more obvious in cells at P15, in which 104 of 113 differential *IAPez-ints* were upregulated in primed cells, and only 9 upregulated in naïve cells (Fig. 4b; Additional file 4: Table S3; Additional file 5: Table S4). Other retrotransposons such as loci of the elements *ID\_B1*, *B3* and *B1\_Mus1* were upregulated in both naïve and primed cells (Fig. 4b, c).

Immunofluorescence microscopy confirmed that *MuERV1* was sporadically expressed in naïve PSCs but suppressed in primed PSCs (Fig. 4e, f). This result is in accordance with the repression of 2C genes in primed cells (shown above). These data demonstrated highly polarized retrotransposons in naïve and primed PSCs in that naïve



**Fig. 4** Retrotransposons distinguish mouse naïve from primed state. **A** Heatmap of RNA-seq expression data from CBA × C57 naïve mESCs and primed mEpiSCs. Data shown include 1000 differentiated expressed transposable elements (TEs) with the highest standard deviation (SD) between the samples. **B** TE signatures of naïve or primed mESCs. The most four heavily primed or naïve-biased TE families are represented as columns split into three segments: overexpressed integrants in naïve cells (1.5-fold cut-off,  $P < 0.01$ ), overexpressed integrants in primed cells, and no change. **C** The top 20 TE families with the highest fold change between naïve and primed cells are heavily enriched for *LINE1* family (naïve) and *ERVK* family (primed). The triangles highlight the high expression of *MERVL* in naïve mESCs. **D** Left panel of boxplot showing the expression of integrant of *L1Md\_A* and related families according to Dfam database. Right panel of boxplot showing the expression integrant of *IAPez-int* and related families according to Dfam database. The significance is calculated by Wilcoxon Test. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . **E** Representative immunofluorescence of 2C marker *MuERV-L* Gag (red) in naïve and primed cells. DAPI-stained nuclei in blue. Scale bars = 10  $\mu\text{m}$ . **F** Proportion of *MuERV-L*-Gag<sup>+</sup> cells based on immunofluorescence images in Fig. 4E

mESCs are marked by high transcripts of *L1Md\_T* whereas primed mEpiSCs are featured with high transcriptional level of *IAPez-int*.

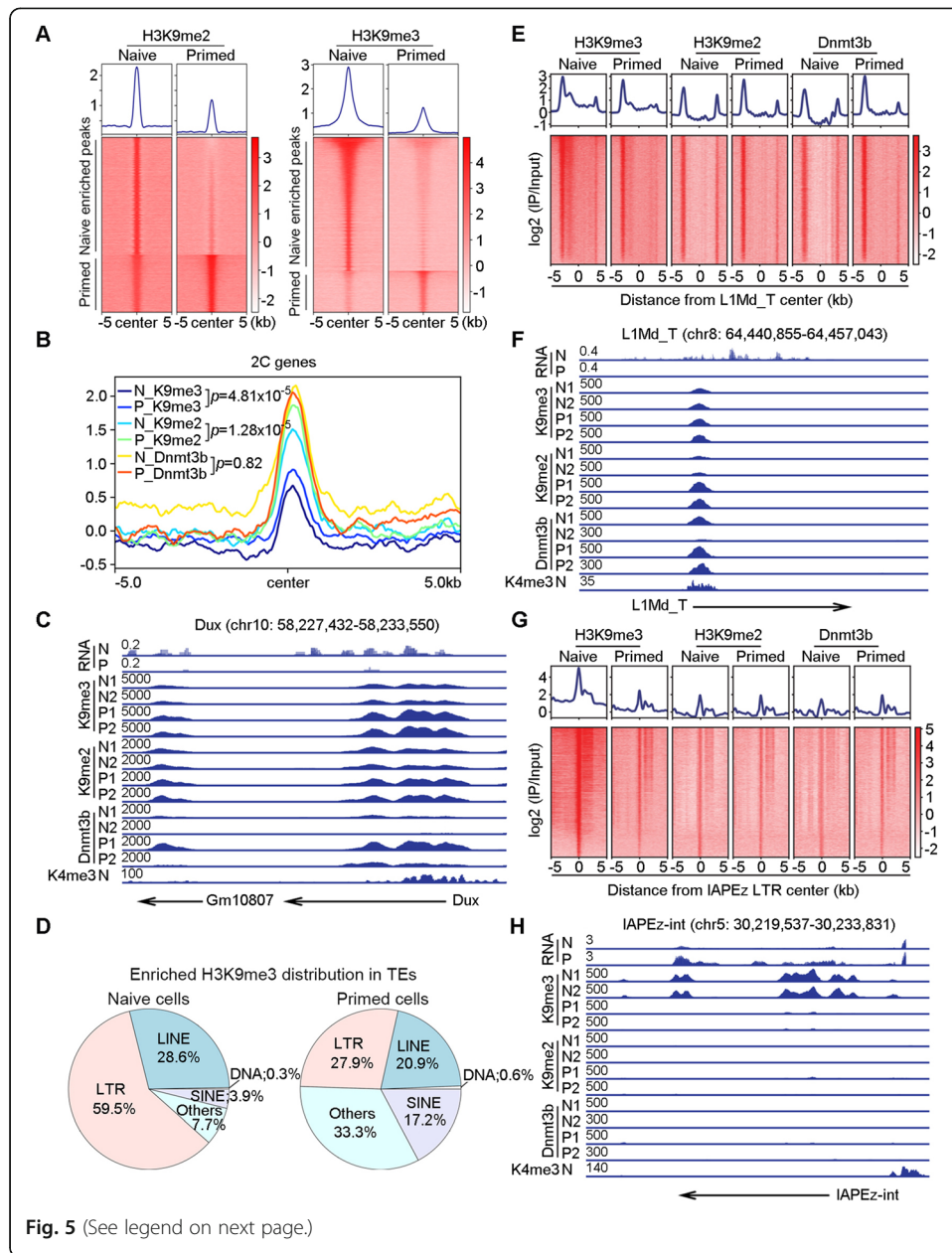
### Epigenetic regulation of transcription of naïve and primed PSCs

We explored the underlying mechanisms of how retrotransposons and 2C genes are repressed in primed cells. Initially, we examined the expression pattern of potential histone epigenetic modifications, including H3K9me3, H3K27me3, and H3K27ac, by immunofluorescence microscopy, which revealed no significant difference between naïve and primed cells (Additional file 1: Figure S7e). H3K9me3 exhibited typical foci with DAPI-stained heterochromatin, and H3K27ac and H3K27me3 showed similar nuclear distribution in naïve and primed cells. The expression pattern of PSCs differed greatly from that of differentiated MEF cells served as a control (Additional file 1: Figure S7e).

To determine the distribution of histone epigenetic modifications across the genome, we performed bulk chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) and ultra-low-input micrococcal nuclease-based native ChIP sequencing (ULI-NChIP-seq) assay of H3K9me3, H3K9me2, and Dnmt3b. At the genome-wide level, differential peaks between H3K9me2 and H3K9me3 were mainly enriched in naïve cells (Fig. 5a). Firstly, we looked into the enrichment at naïve marker genes such as *Tbx3* and *Klf4* by the ChIP-seq data. Distinctly, H3K9me3 enriched at the promoter region of *Tbx3*, while H3K9me2 and Dnmt3b enriched at the promoter region of *Klf4* in primed cells (Additional file 1: Figure S8a, b), likely repressing their expression, respectively.

Next, we assessed the abundance distribution of Dnmt3b and the two epigenetic modifications H3K9me2 and H3K9me3 across the 2C genes. H3K9me3 and H3K9me2 were more enriched in primed cells across the 2C genes compared to the naïve cells, but no significant difference in Dnmt3b enrichment was found (Fig. 5b). We analyzed the published *Setdb1*, *G9a*, and *Dnmt3b* (H3K9me3, H3K9me1/2, and DNA methyltransferases gene, respectively) KO RNA-seq data in naïve mESCs [70–72]. The naïve genes and 2C genes were upregulated after *Setdb1* and *G9a* KO, while only the naïve genes increased after *Dnmt3b* KO, the 2C genes were not affected (Additional file 1: Figure S8g-l), demonstrating that 2C genes were mainly regulated by H3K9 methylation rather than DNA methylation.

As 2C genes are expressed in only a small percentage (1–5%) of mESC populations, the average abundance signal of the histones could be masked when comparing the whole cell populations in naïve and primed PSCs. Nonetheless, some important genes still manifested differences in the abundance of histones between the two pluripotent states. For instance, *Dux* was discovered as a central transcription regulator of zygotic genome activation in 2C embryos and critical to the expression of *MERVL* and 2C genes during development and in mESCs [73–75]. In our RNA-seq and ChIP-seq data analysis, *Dux* was downregulated in primed mEpiSCs compared to naïve mESCs, and the downregulation was accompanied by elevated levels of H3K9me3, H3K9me2, and Dnmt3b at *Dux* promoter region marked by H3K4me3 (Fig. 5c). These data suggested that H3K9me3, H3K9me2, and Dnmt3b orchestrate in reducing expression of naïve pluripotency genes and suppressing specific 2C genes in primed cells.



**Fig. 5** (See legend on next page.)

(See figure on previous page.)

**Fig. 5** Epigenetic regulation of gene expression and TEs in naïve and primed cells. **A** H3K9me2 (left) and H3K9me3 (right) enrichment signal in differential enriched peaks. There are 28,347 and 32,326 enriched H3K9me2 and H3K9me3 peaks in naïve PSCs and 13,701 and 9,832 enriched H3K9me2 and H3K9me3 peaks in primed PSCs, respectively. The ChIP-seq signals are calculated as log<sub>2</sub> ratio of normalized reads relative to the input. **B** H3K9me3, H3K9me2, and Dnmt3b enrichment signal profile plot around the TSS of 2C genes in naïve and primed PSCs, respectively. The ChIP-seq signals are calculated as log<sub>2</sub> ratio of normalized reads relative to the input. The Wilcoxon signed-rank test was used to calculate the significance of difference. **C** Density plot of RNA-seq and ChIP-seq signal of *Dux* locus. Arrow indicates the direction of transcription. H3K4me3 was used to indicate the promoters of genes and analyzed based on published data from naïve mESCs cultured in serum/Lif [69]. RNA-seq signals are normalized by CPM, ChIP-seq signals are normalized by RPKM. **D** Differential enriched H3K9me3 peaks distribution in TEs. Peaks located in regions outside of TEs are labeled as Others. **E, G** Plot of H3K9me3, H3K9me2, and Dnmt3b binding profile and heatmap at all *L1Md\_T* and *IAPeZ-int* loci in naïve and primed PSCs. The ChIP-seq signals are calculated as log<sub>2</sub> ratio of normalized reads relative to the input. **F, H** Density plot of RNA-seq and ChIP-seq signal of representative *L1Md\_T* and *IAPeZ-int* loci in naïve and primed PSCs. Arrow indicates the direction of transcription. H3K4me3 were used to indicate the promoters of genes. RNA-seq signals are normalized by CPM, and ChIP-seq signals are normalized by RPKM

### Distinct epigenetic regulation of *L1Md\_T* and *IAPeZ* in naïve and primed PSCs

The highly expressed specific retrotransposons in naïve and primed PSCs prompted us to understand their molecular regulation by potential epigenetic modifications. Notably, H3K9me3 bound much more LTRs in naïve cells than in primed cells (Fig. 5d), suggesting that H3K9me3 might play a fundamental role in silencing ERVs in naïve pluripotent cells. Then we examined the top two expressed RTEs in naïve and primed cells, the L1 family *L1Md\_T* and *L1Md\_A*, ERVs *IAPeZ* and *ETnERV3*, respectively. As to L1 family, Dnmt3b bound more *L1Md\_T* and *L1Md\_A* in primed cells compared with those of naïve cells (Fig. 5e; Additional file 1: Figure S8c), indicating that DNA methylation might primarily regulate L1 RTEs in these two pluripotent cells. ChIP-seq data showing enrichment on specific sites of these two L1 families also confirm our conclusion. Especially, Dnmt3b bound more at *L1Md\_T* and *L1Md\_A* sites in primed cells than in naïve cells; meanwhile, H3K9me2 also seemed to play a repressive role in primed cells (Fig. 5f; Additional file 1: Figure S8d). As to the regulation of *IAPeZ* and *ETnERV3*, ChIP-seq data revealed significantly decreased enrichment of H3K9me3 near the TSS site and also across these two ERVs in primed cells (Fig. 5g, h; Additional file 1: Figure S8e, f); thus, this reduced heterochromatic distribution might de-repress *IAPeZ* and *ETnERV3*, resulting in their highly specific expression under the primed state. The H3K9me2 and Dnmt3b showed only minimal enrichment on these two ERVs sites (Fig. 5g, h; Additional file 1: Figure S8e, f). These data suggested that diverse epigenetic modifications regulate specific transcription of RTEs in naïve mESCs and primed mEpiSCs. Reduced Dnmt3b (DNA methylation) may promote specific L1 family transcription of *L1Md\_T* and *L1Md\_A* under naïve state, whereas *IAPeZ* and *ETnERV3* transcription is mainly regulated by H3K9me3 in primed PSCs.

Kap1 is a co-repressor in mESCs and tethers to DNA by sequence-specific Kruppel-associated box zinc finger proteins (KRAB-ZFPs) and induces local heterochromatin formation through the histone methyltransferase Setdb1, responsible for H3K9me3, and has been shown to control endogenous retroviruses in ESCs [38, 39]. Our RNA-seq data and western blot analysis showed that Kap1 was downregulated in primed cells compared to naïve cells (Additional file 1: Figure S9a). We hypothesized that decreased Kap1 could lead to upregulation of *IAPeZ-int* in primed cells. We analyzed the



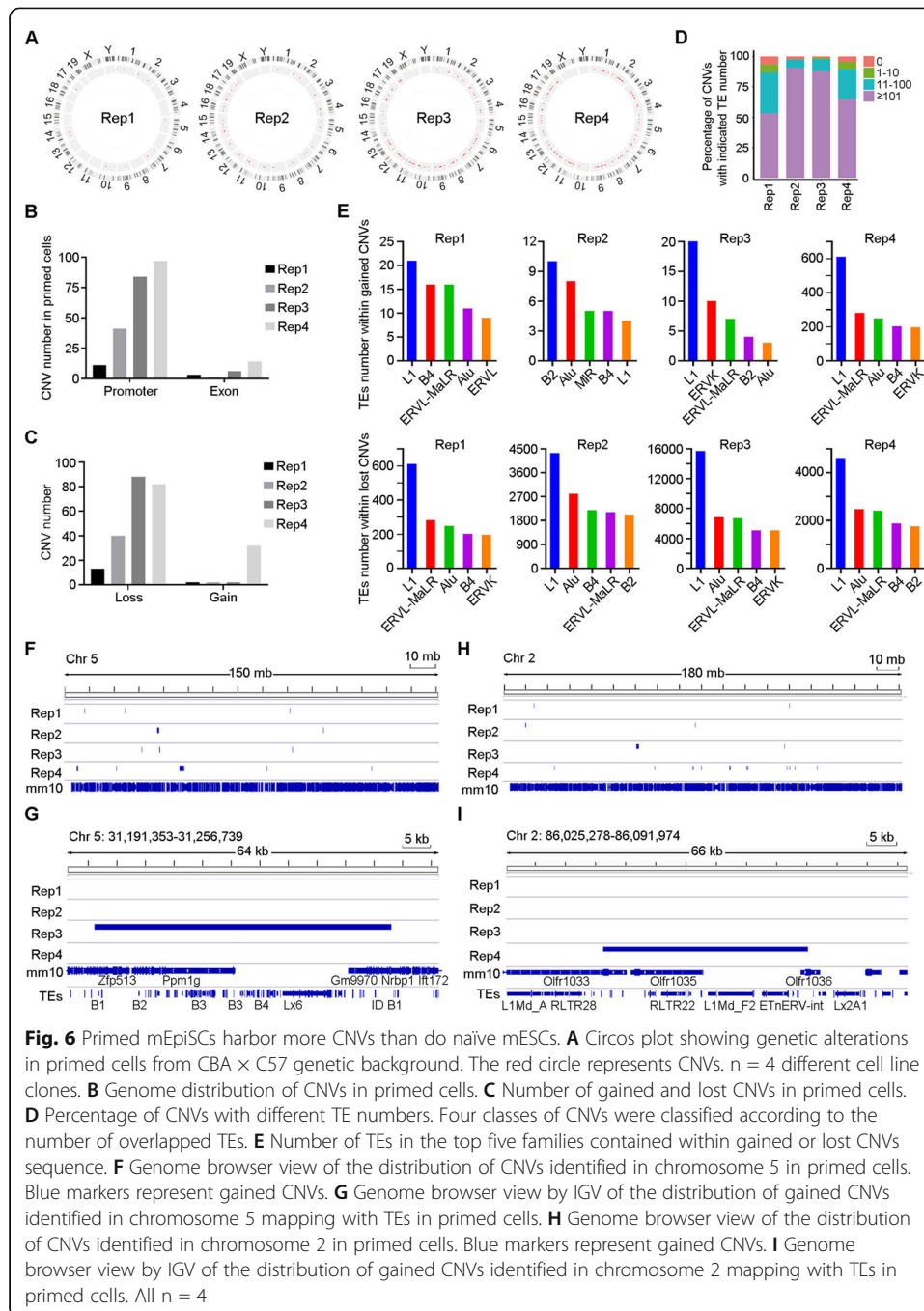
published sequencing data of *Kap1* knockdown (KD) ESCs [75, 76]. *IAPez-int* was the most significantly upregulated type of retrotransposons (Additional file 1: Figure S 9b, c). The enrichment of *Kap1* was very high at the *IAPez-int* loci in ESCs and the enrichment of H3K9me3 at the *IAPez-int* site was dramatically decreased after *Kap1* KD (Additional file 1: Figure S9d, e, g). Likewise, *Kap1*-mediated H3K9me3 also has an inhibitory effect on *LIMd\_T* (Additional file 1: Figure S 9d, f, g). These data together further support the notion that heterochromatic histone modifiers can regulate *IAPez-int* and *LIMd\_T* transcription in PSCs.

### Primed PSCs manifest increased genomic instability in association with retrotransposon insertion

Lastly, we were curious to know any consequences, if any, resulting from differential DNA repair capacity and expression of retrotransposons in naïve versus primed PSCs. We compared the genomic stability of seven different clones from two different genetic background cell lines under naïve and primed state by conducting the whole-exome sequencing (WES) analysis. We applied the naïve ESCs served as references and compared the primed with naïve ESCs, because the naïve ESCs efficiently produced germline chimeras as well as TEC pups, indicating that the naïve ESCs are genomic stable. By such comparison with naïve ESCs, any insertions and the TEs contained within copy number variations (CNVs) in primed ESCs presumably are considered as de novo and specific. Six out of seven primed cell line clones from two genetic backgrounds had more CNVs compared to the naïve clones, and these CNVs contained similar families of TEs like *LI*, *Alu*, *B2*, and *B4*. The variable number of CNVs and TEs within CNVs indicates the heterogeneity of different clones (Additional file 6: Table S5). Notably, primed cells in CBA × C57 background accumulated a higher number of CNVs compared to naïve cells (Fig. 6a). Among the four cell line clones at CBA × C57 background, CNVs of primed cells mainly were found in the promoter regions (Fig. 6b). Most of them showed lost CNVs (Fig. 6c). Remarkably, when all the lost and gained CNVs in primed cells were mapped to the genome, most of CNVs contained more than 100 TEs (Fig. 6d) of various families, especially *LI*, *Alu*, and *ERVK* (Fig. 6e). As examples, the CNVs occurred on chromosomes 2 and 5 and the TEs within CNVs were found variable in different clones (Fig. 6f–i). Similar observation was also found in another cell line at 129 × C57 background and again TEs contained in CNVs were variable (Additional file 6: Table S5). These results demonstrated increased genomic instability in the primed ES cells and suggested that more CNVs in primed ES cells could be introduced by insertion of aberrantly activated retrotransposons.

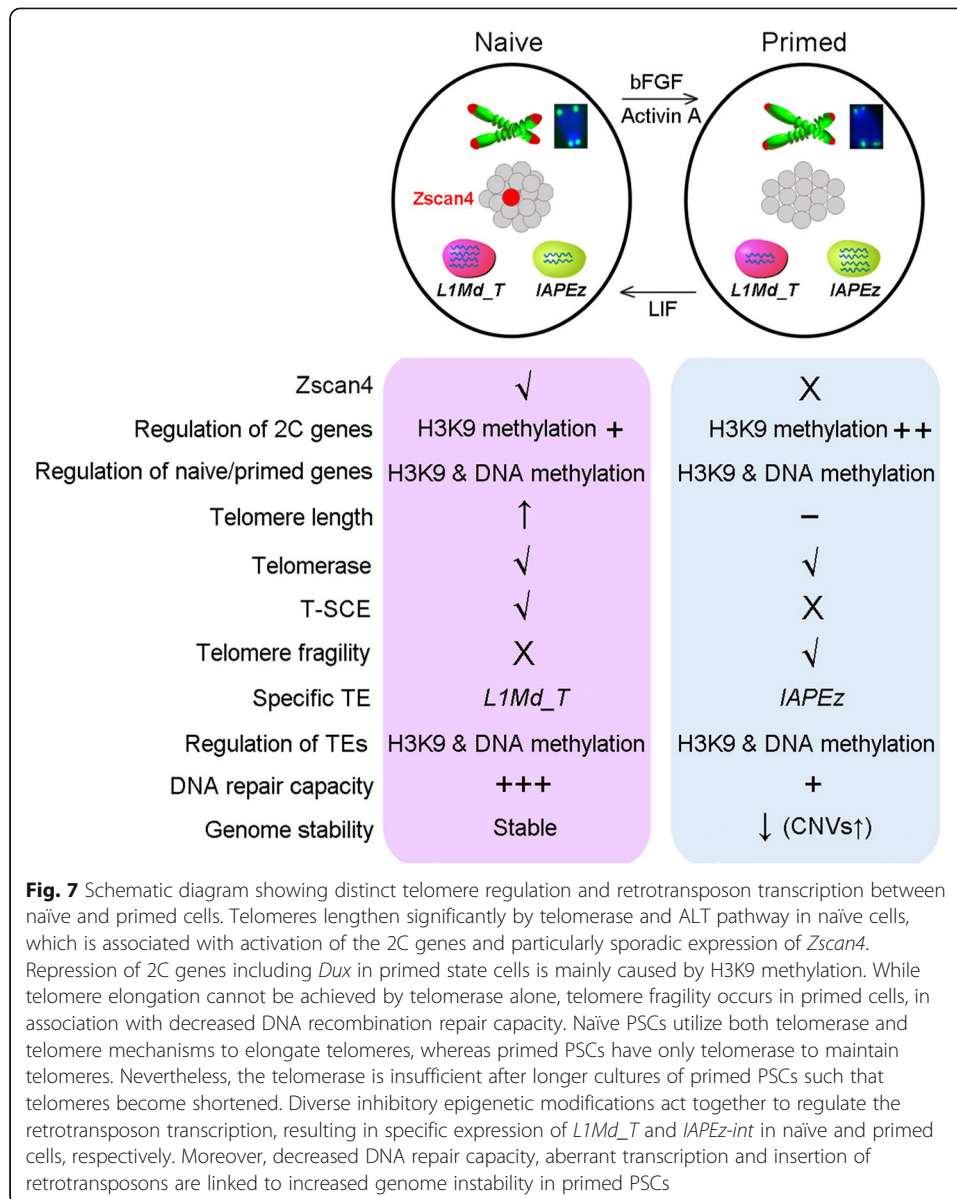
### Discussion

We find that primed PSCs exhibit fragile telomere with passages and retrotransposon regulation by various histone modifications at specific loci, and deficient DNA recombination repair, distinct from naïve PSCs. Consequently, primed PSCs harbor increased frequency with passages of CNVs and genomic instability where retrotransposon integration is found, in contrast to naïve PSCs (Fig. 7). Our naïve ESCs were maintained on



the presence of feeder cells under conventional serum/LIF culture conditions and exhibited developmental pluripotency as evidenced by efficient generation of germline chimeras as well as TEC pups that lived healthily to adulthood and reproduced, also indicating that the naïve ESCs were genomic stable.

Primed PSCs show reduced DNA recombination and repair capacity as well as low mismatch repair and nucleotide-excision repair capacity, different from naïve PSCs. DSB repair process is very rapid and homologous recombination-mediated repair (HRR) may predominate in mESCs [77]. *Rad51* and *Brca1*, involved in HRR [78], are downregulated under primed state. The high DNA recombination and repair capacity



are important for maintaining telomere function and genomic stability in naïve cells. Telomeres lengthen under naïve state whereas elongation of telomeres is suppressed in primed cells, and this is linked to the repression of 2C genes and particularly *Zscan4* that is critical for telomere lengthening and maintaining the genomic stability of mESCs by T-SCE-based homologous recombination [52, 79]. *Zscan4* repression for longer term can impair naïve pluripotency and developmental potential of mESCs in vivo [25] and induce expression of differentiation-related genes [80]. T-SCE also is suppressed and high telomerase activity is required for telomere elongation in conventional primed human ESCs [81]. Lack of telomere recombination-induced telomere elongation corroborates increased frequency of fragile telomeres found in primed PSCs at later passage. The fragile sites are unstable under replication stress, and the fragile telomeres, as a result of replication stress at telomeres, would be more likely to result in chromosome instability due to deficiency in DSB repair at subtelomeric regions [54,

82]. Impaired telomere elongation by the accumulation of repressive histones and declined capacity of DNA recombination repair in association with fragile telomeres may lead to genomic instability of primed cells. In consistency, telomere erosion in human pluripotent stem cells conventionally known as primed state of pluripotency induces DNA damage response and leads to ATR-mediated mitotic catastrophe [83]. Also, mESCs depleted for telomere maintenance protein *TRF2* elicit minimal or only mild DNA damage response, in contrast to non-pluripotent or differentiated cells showing that *TRF2*-mediated telomere protection is dispensable in pluripotent stem cells [84, 85]. Notably, *TRF2* depletion in mESCs activates a totipotent-like two-cell-stage transcriptional profile including high levels of *Zscan4* and the upregulation of *Zscan4* reduces DNA damage and elongates telomeres in the absence of *TRF2* [85], providing additional evidence in supporting importance of *Zscan4* in pluripotent stem cells.

Increasing evidence established that *Zscan4* is critical to telomere elongation and maintenance of naïve mouse ESCs. However, mechanisms underlying *Zscan4*-mediated telomere elongation remain elusive. Telomeres shortened after *Dmc1* KO in naïve ESCs and slightly shortened after *Rad51* KO (Additional file 1: Figure S6e, f), implying that DNA recombination is involved in telomere maintenance. We have attempted to exogenously overexpress *Dmc1* in the *Dmc1* KO naïve cell lines but failed in achieving the *Dmc1* rescue cells. As we did not overexpress *Dmc1* in WT naïve cells, we cannot conclude whether overexpression of *Dmc1* is toxic to the cells or whether this phenotype is an off-target effect. Also, it is likely that the telomere phenotype in primed state cells may be caused by comprehensive factors. Overexpression of a single factor alone may not be sufficient to rescue telomere phenotype in primed state cells. It will be necessary to systematically investigate telomere function and regulation in naïve and primed PSCs in future experiments.

Moreover, differential transcription of retrotransposons clearly distinguishes mouse primed from naïve states. *LIMd\_T* is highly transcribed in naïve ESCs and *IAPez-int* in primed cells. *LINE1* regulates global chromatin accessibility in the early mouse embryo and also acts as a repressor of *MERVL* to balance 2-cell-like activity in mESCs [86, 87]. These data also support the notion that *L1* contributes to the pluripotency network [86]. Indeed, we also observed higher expression levels of *L1* than those of *MERVL* in naïve PSCs. Expression of retrotransposons is known to be regulated by a variety of epigenetic modifications, including DNA methylation and histone methylation [88–90]. Retrotransposons in naïve and primed cells are differentially regulated by diverse epigenetic modifications. For example, downregulation of *LIMd\_T* in the primed state is mainly regulated by DNA methylation, and transcription of *IAPez-int* is mainly regulated by histone methylation, such as H3K9me3. In support, heterochromatin modifier SETDB1 prevents TET2-dependent activation of *IAP* retroelements (especially *IAPez-int*) in mESCs [91]. Diverse epigenetic modifications also regulate *Dux* expression, which in turn can activate downstream 2C genes such as *Zscan4* and *MERVL* that serve as enhancers of 2C genes including *Zscan4* itself [73–75, 80]. *Zscan4*, in turn, activates *MERVL* and cleavage embryo genes [80]. In naïve PSCs, the highly expressed *Dux* activates a series of downstream 2C genes, including *Zscan4*, which can activate *MERVL* and rapidly lengthens telomeres through T-SCE. A large amount of *LINE1* transcription is required to inhibit excessive 2C gene expression to maintain genome stability and pluripotency of naïve cells. In contrast, under primed state, diverse inhibitory

heterochromatin epigenetics act in combination at the *Dux* gene promoter region to inhibit *Dux* expression and further repress expression of downstream 2C genes. Therefore, the primed cells do not need *LINE1* to inhibit 2C genes. Recently, *Dux* expression is reported to facilitate nuclear transfer and somatic reprogramming by activating downstream 2C genes and plays an important role during 2C-like to pluripotent state transition process [92, 93].

However, aberrant *LI* transcripts that have retrotransposition activity may harm the host genome by random integration and consequently lead to genomic instability [94], even though fewer *LI* is activated in primed cells than in naïve cells. Indeed, exome-seq data reveals that primed cells accumulate more CNVs containing various TEs. Although a relatively large CNV contains one or many TEs, the CNV boundaries do not necessarily indicate the mechanism underlying the induction of CNV by TEs. RNA-seq data shows that primed ESCs express genes for lineage specification such as *Eomes* and *Gsc* as mesendoderm marker genes [95], suggestive of differentiation properties. Additionally, primed ESCs exhibit reduced DNA repair capacity, increased telomere fragility, declined mitochondria functions, and aberrant cell cycle. These factors together could compromise developmental potential of primed ESCs.

## Conclusions

Primed PSCs have been known to have lower pluripotency than do naïve PSCs. Our data reveals that primed PSCs in mice exhibit declined telomere maintenance and elevated retrotransposon-associated genomic instability, coincided with reduced DNA recombination repair capacity. Specifically, (1) Telomeres lengthen in naïve PSCs, which acquire robust recombination repair, but not in primed state, such that naïve PSCs maintain telomere integrity, whereas primed PSCs exhibit fragile telomeres; (2) Naïve and primed states show distinct retrotransposon activation and related epigenetic state; (3) Genomic instability is increased in primed PSCs, in association with the insertion of retrotransposons and increased CNVs in contrast to naïve PSCs; (4) Reduced function in cell cycle and mitochondria and differentiation properties are found in primed PSCs. Mitochondria dysfunction has been linked to telomere attrition. Fragile telomeres, decreased DNA recombination repair capacity, and aberrant retrotransposon activity together likely contribute to increased genomic instability found in primed mouse PSCs after prolonged culture in vitro. This also may explain why mouse naïve PSCs achieve developmental pluripotency by stringent functional assays, but primed PSCs do not. These findings could have implications in further derivation and characterization of human naïve and primed PSCs.

## Methods

### Animal care and use

Use of mice for this research was approved by the Nankai University Animal Care and Use Committee. All mice used in this study were taken care of and operated according to the relevant regulations. Mice were housed and cared in individually ventilated cages (IVCs) on a standard 12 h light: 12 h dark cycle in the sterile Animal Facility. Oct4-GFP (OG2) mice (CBA × C57, JAX stock #004654) that carry Oct4 distal promoter-driven GFP were purchased from Model Animal Research Center of Nanjing University. 129 ×



C57 mice, albino ICR mice, and albino Kunming (KM) mice were purchased from Beijing Vital River Laboratory Animal Technology Co., Ltd. The mice were humanely euthanized for specific experiments.

#### Derivation of naïve mESCs

Conventional naïve mESC lines were established and characterized based on the method described and reviewed previously [96]. Blastocysts were isolated from the uterine horns of pregnant females at embryo (E) 3.5 using a dissection microscope in HKSOM and plated onto mitomycin C-treated MEF cells served as feeders in KSR/DMEM (K/DL) medium and cultured for 7 days to form outgrowths. Emerging ICM outgrowths were directly picked into serum/LIF (S/L) medium on feeders to establish stable naïve mESC lines. mESCs were maintained by dissociating cells with 0.25% TE every 2–3 days and re-plating them onto feeder cells. K/DL medium contains knockout DMEM (Invitrogen) supplemented with 20% Knockout serum replacement (KSR, Invitrogen), 1 mM L-glutamine, 100 µM NEAA, 1% 2A, 0.1 mM β-mercaptoethanol (β-me, Invitrogen), 1 µM PD0325901 (Miltenyi), and 1000 IU/ml mouse LIF (mLIF, Millipore). S/L medium (ESC culture medium) contains knockout DMEM supplemented with 20% FBS (ES quality, Hyclone), 1 mM L-glutamine, 100 µM NEAA, 1% 2A, 0.1 mM β-me, and 1000 IU/ml mLIF.

#### ESC cultures

Three mouse ES cell lines were used in this study. OG2 ESC line was derived from C57BL6 × CBA mice that carry *Oct4* distal promoter-driven GFP, 129 × C57 ESC line from B6 × 129F1 mice, and AKJ2 ESC line from actin-GFP mice in C57BL/6J origin. Naïve mESCs were routinely cultured based on the method described previously [23]. Briefly, mESCs were cultured under 5% CO<sub>2</sub> at 37 °C on mitomycin C-treated MEF feeder in S/L medium consisting of knockout DMEM supplemented with 20% fetal bovine serum (FBS, ES quality, Hyclone), 1000 U/ml leukemia inhibitory factor (LIF) (ESGRO, Chemicon), 0.1 mM non-essential amino acids, 0.1 mM β-mercaptoethanol, 1 mM L-glutamine, and penicillin (100 U/ml) and streptomycin (100 µg/ml). Naïve mESCs were maintained by dissociating cells with 0.25% TE every 2–3 days and re-plating them onto feeder cells.

Conversion of mEpiSCs from naïve mESCs was achieved based on the method described [11]. In brief, mESCs were seeded at a density of 10<sup>5</sup> cells per 6-well plate in S/L medium on feeder cells. On the next day, the medium was replaced with mEpiSCs bFGF/Activin A (F/A) medium. After culture for 3–4 days, the surviving cells grew and formed large compact colonies. These primed EpiSCs were passaged every 3–4 days on feeder cells by collagenase IV (1 mg/ml; Invitrogen) dissociated as cell clumps or mechanical dissociation by syringe needles. The F/A medium contains a 1:1 mixture of DMEM/F12 (Invitrogen) supplemented with N2 (Invitrogen) and Neurobasal (Invitrogen) supplemented with B27 (Invitrogen), 1 mM L-glutamine, 0.1 mM non-essential amino acids, penicillin (100 U/ml) and streptomycin (100 µg/ml), 0.1 mM β-mercaptoethanol, 1% KSR, 12 ng/ml bFGF (Invitrogen), and 20 ng/ml Activin A (Peprotech). The density of feeder cells was crucial to maintaining these mEpiSCs passaging in vitro in an undifferentiated state.

### Production of chimeras and genotyping

Approximately 10–15 naïve mESCs or primed mEpiSCs were injected into 4–8-cell embryos of ICR mice as hosts using a Piezo injector as described [96]. Injected embryos were cultured overnight in KSOM<sub>AA</sub> medium. Blastocysts were transferred into uterine horns of E2.5 surrogate mice. Pregnant females delivered pups naturally at about E19.5. Pups were identified initially by coat color. Contribution of mESCs or mEpiSCs to various tissues in chimeras was confirmed by standard DNA microsatellite genotyping analysis using *D12Mit136* primers: 5'-TTA ATT TTG AGT GGG TTT GGC-3' and 5'-TTG CTA CAT GTA CAC TGA TCT CCA-3'.

### Tetraploid embryo complementation (TEC)

To generate mice by tetraploid embryo complementation, two-cell embryos were collected from the oviducts of ICR females and electrofused to produce one-cell tetraploid embryos that were then cultured in KSOM media. Naïve mESCs were injected into the tetraploid blastocyst cavity. The blastocysts were placed in KSOM with amino acids until embryo transfer. Approximately thirty injected blastocysts were transferred to each uterine horn of 2.5-day-postcoitum pseudopregnant ICR females. Pregnant recipients with tetraploid embryos were subjected to cesarean section on day 18.5 of gestation. TEC mice further were mated with ICR mice to test their fertile capacity. The contribution of naïve mESCs to TEC mice was confirmed by standard DNA microsatellite genotyping analysis using *D12Mit136* and *D12Mit99* primers (5'-CTT ACA GAA AAT GAA AAC CAA AAC A-3' and 5'-CCT CTG CTT TAG AGG CAA ACG-3').

### DNA damage response

To induce DNA damage, naïve and primed PSCs were exposed to 2.5  $\mu$ M etoposide or 5 mM H<sub>2</sub>O<sub>2</sub> for 2 h. Two hours and 20 h following the exposure, cells were collected for immunofluorescence microscopy of 53BP1 or  $\gamma$ H2AX foci, commonly used as DNA damage response markers [53]. More than 50 cells were counted for quantification of each group.

### Immunofluorescence microscopy

Cells were washed twice in PBS, fixed in freshly prepared 3.7% paraformaldehyde for 30 min on ice, washed once in PBS, and permeabilized in 0.1% Triton X-100 in blocking solution (3% goat serum plus 0.1% BSA in PBS) for 30 min at room temperature, then washed once in PBS, and left in blocking solution for 2 h. Cells were incubated overnight at 4 °C with primary antibodies anti-Oct4, Nanog, SSEA1, Dnmt3b, 53BP1,  $\gamma$ H2AX, or Zscan4. Cells were washed three times with blocking solution and incubated for 1 h with secondary antibodies at room temperature. Goat Anti-Mouse IgG (H + L) FITC, Goat Anti-Rabbit IgG (H + L) Alexa Fluor® 594, and Donkey Anti-Goat IgG (H + L) Alexa Fluor® 594 diluted 1:300 with blocking solution were used. Samples were washed and counterstained with 0.5  $\mu$ g/ml DAPI (Roche) in Vectashield (Vector) mounting medium. Fluorescence was detected and imaged using Axio-Imager Z2 Fluorescence Microscope (Carl Zeiss).

### Western blot

Cells were washed twice in PBS, collected, and lysed in cell lysis buffer on ice for 30 min and then sonicated for 1 min at 60 of amplitude with 2-s intervals. After centrifugation at 10,000g, 4 °C for 10 min, supernatant was transferred into new tubes. The protein concentration was measured by bicinchoninic acid, and then protein samples were boiled in SDS Sample Buffer at 95 °C for 10 min. Same amount of protein of each cell extract was resolved by 10% Acr-Bis SDS-PAGE and transferred to polyvinylidene difluoride membranes (PVDF, Millipore). Nonspecific binding was blocked by incubation in 5% skim milk in TBST at room temperature for 2 h. Blots were then probed with primary antibodies overnight by incubation at 4 °C with Oct4, Nanog, Dnmt3b, or Zscan4, and  $\beta$ -actin served as a loading control. Immunoreactivity bands were then probed for 2 h at room temperature with the appropriate horseradish peroxidase (HRP)-conjugated secondary antibodies, goat anti-Rabbit IgG-HRP, or goat anti-Mouse IgG (H + L)/HRP. Protein bands were detected by chemiluminescent HRP substrate (WBKLS0500, Millipore). Information of antibodies is listed in Additional file 7: Table S6.

### Dot blot

Total 5mC and 5hmC levels were measured by dot blot analysis based on the method described [97]. Briefly, genomic DNA was denatured in 0.4 N NaOH and 10 mM EDTA at 95 °C for 10 min, and 2-fold serial dilutions were spotted on positively charged nylon membranes. The membranes were immunoblotted with 5mC and 5hmC antibody, followed by HRP-conjugated goat anti-mouse or anti-rat antibodies. DNA loading was verified by staining with methylene blue.

### Generation of *Dmc1*, *Rad51*, or *Brca1* KO naïve mESCs by CRISPR/Cas9

Briefly, CRISPR/Cas9 expression vector pSpCas9(BB)-2A-Puro (PX459) was digested with BbsI endonuclease (Fermentas). The sgRNA targeting *Dmc1*, *Rad51*, or *Brca1* was designed using <http://crispor.tefor.net/crispor.py>. sgRNA-F and sgRNA-R were annealed into a double strand as 95 °C for 30 s, 72 °C for 2 min, 56 °C for 2 min, 37 °C for 2 min, 25 °C for 2 min, and 4 °C for storage. Annealed double strand was diluted to 100-fold and constructed to a linearized vector. Positive clones were picked for sequencing, and the sequencing primers were Human U6 Promoter-F (ACT ATC ATA TGC TTA CCG TAA C). After sequencing, the correct bacterial solution was chosen for expansion culture and the ultrapure plasmid extracted for cell transfection. After transfection, cells were screened by puromycin and then selected and expanded from a single clone. The knockout cells were characterized by PCR and sequencing. The knockout cells were cultured for additional 8 passages prior to measurement of telomeres. The sequences of sgRNA and PCR detection primers are listed in Additional file 7: Table S6.

### Telomerase activity by TRAP assay

Telomerase activity was determined by the Stretch PCR method according to manufacturer's instruction using TeloChaser Telomerase assay kit (T0001, MD Biotechnology). About  $2.5 \times 10^4$  cells from each sample were lysed. Lysis buffer served as negative



integrated using the TFL-TELO program (kindly provided by Peter Lansdorp, Terry Fox Laboratory). More than 15 metaphase spreads were counted for each group.

#### **Telomere chromosome orientation-fluorescence in situ hybridization (CO-FISH)**

CO-FISH assay was performed as described [59], with slight modification. mESCs or mEpiSCs were incubated with BrdU (10  $\mu$ M) for 12 h. Nocodazole at 0.3  $\mu$ g/ml was added for 3 h prior to cell harvest, and metaphase spreads were prepared by a routine method. Chromosome slides were treated with RNase A, fixed with 4% formaldehyde, then stained with Hoechst 33258 (0.5 mg/ml), incubated in 2  $\times$  SSC (Invitrogen) for 15 min, and exposed to 365 nm UV light (Stratalinker 1800UV irradiator) for 40 min. BrdU-substituted DNA was digested with Exonuclease III (Takara). Slides were then dehydrated through ethanol series and air-dried. PNA-FISH was performed with FITC-OO-(CCCTAA)<sub>3</sub> (Panagene, F1009). Slides were hybridized, washed, dehydrated, mounted, and counterstained with 1.25  $\mu$ g/ml DAPI in VectaShield antifade medium. Digital images were captured using CCD camera on Zeiss Imager Z2 microscope. To analyze telomere sister chromatid exchange (T-SCE), one signal at each end of the chromosome was counted as no T-SCE, while two signals at both chromatids on one chromosome end were identified as one T-SCE. At least 15 metaphase spreads were counted for the frequency of T-SCE. Pairwise comparisons for statistical significance were made by t-tests.

#### **Metabolic flux analysis**

Agilent Seahorse XFe24 Analyzers was used to measure oxygen consumption rate of naïve and primed state PSCs. Cells were plated in XF24 Cell Culture Microplates pre-coated with 1% Matrigel at a density of 60,000 per well. The next day, cells were treated with Seahorse XF Cell Mito Stress Test Kit (10  $\mu$ g/ml oligomycin, 1  $\mu$ M FCCP, and 1  $\mu$ M rotenone and antimycin) and measured following the manufacturer's instructions.

#### **Library preparation and RNA-sequencing**

mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEB Next First-Strand Synthesis Reaction Buffer (5x). First-strand cDNA was synthesized using random hexamer primer and M-MLV Reverse Transcriptase (RNase H<sup>-</sup>). Second-strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEB Next Adaptors with hairpin loop structure were ligated to prepare for hybridization. To select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3  $\mu$ l USER Enzyme (NEB, USA) was used with size-selected and cDNA adaptor-ligated at 37 °C for 15 min followed by 5 min at 95 °C followed by PCR. PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index Primer. At last, PCR products were purified using AMPure XP system and library quality assessed on the Agilent Bioanalyzer 2100 system. Cluster of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit (Illumina) according to the



manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq platform.

#### **Bioinformatics analysis for differentially expressed genes**

The adapter sequences of RNA-seq data and low-quality reads with Phred score < 5 were deleted with Cutadapt before further processing. Hisat2 was used to map RNA-seq data to mm10 with parameter -k 20. Genes were annotated according to the Ensembl database. Transposable element annotations were from UCSC Genome Browser (RepeatMasker). Reads were counted using featureCounts. TE transcripts with parameter “-mode multi” was used to measure differentially expressed ERVs. Genes with expression fold change > 2 and adjusted *P* value < 0.01 according to DESeq2 were used for GO and KEGG analysis by clusterProfiler. Clustering and analysis of 2-cell embryo, naïve, and primed pluripotent-specific genes were done according to published RNA-seq data [2, 28]. Z-score of selected genes was used for heatmap. DEGs related to cell cycle and apoptotic signaling pathway (downloaded from <https://www.gsea-msigdb.org>) were plotted in heatmaps.

#### **Bioinformatics analysis of transcription of transposable elements (TEs)**

To precisely estimate the expression of single TE locus, we used the recently published SQuIRE method with “total” mode [68]. SQuIRE function “squire Fetch” was used to download TE annotation from RepeatMasker, then “squire Clean” was used to filter RepeatMasker file for repeats of interest, collapses overlapping repeats. Function “squire Map” was used to align reads to mm10, then “squire Count” was used to quantify reads aligning to TEs and genes. For the TE subfamily level, we add up all count of the TE loci. Only quantification of TE was used by DESeq2 to calculate the differentially expressed TE locus. TE locus whose foldchange > 1.5 and *P* value < 0.01 were considered significant.

#### **Bulk chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) data analysis**

Bulk ChIP experiment was performed based on the method described [80]. Briefly, naïve and primed PSCs were collected after removing off feeders and fixed with 1% paraformaldehyde, lysed, and sonicated to achieve the majority of DNA fragments with 100–1000 bp. DNA fragments were enriched by immunoprecipitation with 5 µg antibody and dynabeads M280 (Life Technologies). The immunoprecipitated material was eluted from the beads by heating for 30 min at 68 °C. To reverse the crosslinks, samples were incubated with Proteinase K at 42 °C for 2 h followed by 67 °C for 6 h. The samples were then extracted with phenol:chloroform:isoamyl alcohol (25:24:1, pH > 7.8) followed by chloroform, ethanol precipitated in the presence of glycogen, and re-suspended in TE buffer. The adapter sequences and low-quality base were removed by Cutadapt. Bowtie2 was used to map reads to mm10. ChIP-seq signal enrichment was analyzed and obtained by bamCompare from DeepTools [101, 102]. ChIP signal heatmap and line plot were also generated by computeMatrix and plotHeatmap from DeepTools. For gene locus visualization, duplicated reads were marked by MarkDuplicates from picard and removed by samtools. Differential enriched peaks were called by sicer

and peaks with  $\log_{10}$  of  $FDR > 10$  were kept. To analyze the enrichment signal on TE subfamily by ChIP-seq, we constructed saf file which considers TE subfamily as meta-feature and used featureCounts to assign mapped reads to the corresponding TE subfamily.

#### **Ultra-low-input micrococcal nuclease-based native ChIP (ULI-NChIP)**

For ULI-NChIP-seq, about  $5 \times 10^4$  naïve or primed cells were used per reaction. The ULI-NChIP procedure was performed as previously described [103, 104]. Sequencing libraries were generated using AT seq library following manufacturer's recommendations and index codes were added. The library quality was assessed on the Qubit 2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. At last, the library was sequenced on an Illumina HiSeq x platform and 150 bp paired-end reads were generated.

#### **Whole-exome sequencing**

Paired-end DNA library was prepared according to manufacturer's instructions (Agilent). Genomic DNAs (gDNA) from cell samples were sheared into 180–280 bp fragments by Covaris S220 sonicator. Ends of gDNA fragments were repaired and 3' ends were adenylated. Both ends of gDNA fragments were ligated at the 3' ends with paired-end adaptors (Illumina) with a single 'T' base overhang and purified using AMPure SPRI beads from Agencourt. The adaptor-modified gDNA fragments were enriched by six cycles of PCR using SureSelect Primer and SureSelect ILM Indexing PreCapture PCR Reverse Primer. The concentration and size distribution of the libraries were determined on an Agilent Bioanalyzer DNA 1000 chip. Whole-exome capture was carried out using SureSelect Mouse All Exon V1 Agilent 5190-4642. An amount of 0.5  $\mu$ g prepared gDNA library was hybridized with capture library for 5 min at 95 °C followed by 24 h at 65 °C. The captured DNA–RNA hybrids were recovered using Dynabeads MyOne Streptavidin T1. Capture products were eluted from the beads and desalted using Qiagen MinElute PCR purification columns. The purified capture products were then amplified using the SureSelect ILM Indexing Post Capture Forward PCR Primer and PCR Primer Index (Agilent) for 12 cycles. After DNA quality assessment, captured DNA library was sequenced on Illumina Hiseq 2000 sequencing platform (Illumina) according to the manufacturer's instructions for paired-end 150 bp reads (Novogene). Libraries were loaded onto paired-end flow cells at concentrations of 14–15 pM to generate cluster densities of 800,000–900,000 per  $\text{mm}^2$  using Illumina cBot and HiSeq paired-end cluster kit.

#### **Whole-exome sequencing data analysis**

The analysis was performed according to the protocols specified for mouse cells [105]. Briefly, the adaptors of raw reads were trimmed by cutadapt, then bwa was used to map reads to mm10 genome. Picard was used to process the aligned reads and GATK to recalibrate the base quality scores. Copy number variant detection was used by software package CNVkit [106], which used both the targeted reads and the nonspecifically captured off-target reads to infer copy number evenly across the genome. We used segmentation algorithm hmm and set threshold of 500 bp and higher to define CNVs. To

obtain the de novo CNVs from primed cells, we pooled all the naïve samples served as a reference and filtered out the CNVs shared among four (CBA × C57) or three (129 × C57) primed cell line clones. TE annotation file RepeatMasker (downloaded from UCSC) was used to obtain the number and type of TEs contained within CNVs.

### Statistical analysis

Data were analyzed by Student's t test if not indicated. Overlap of upregulated genes in naïve and primed cells and 2C genes shown by Venn diagram was analyzed by Fisher's exact test. Wald test was used for analysis of differentially expressed genes and TE locus. Wilcoxon test was used for analysis of differentially expressed TE locus derived from FPKM value and differentially enriched ChIP-seq signal on 2C genes. Significant differences were defined as \* $P < 0.05$ , \*\* $P < 0.01$ , or \*\*\* $P < 0.001$ .

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02417-9>.

**Additional file 1: Figure S1.** Pluripotency and developmental potential of naïve mESCs and primed mEpiSCs. **Figure S2.** Developmental potential of naïve mESCs and primed mEpiSCs *in vivo*. **Figure S3.** Comparison of naïve and primed pluripotent state *in vitro* and *in vivo* ICM and post-implantation epiblast counterpart. **Figure S4.** Comparison of naïve and primed pluripotent state by GO analysis related to DNA repair pathways. **Figure S5.** Differential telomere length dynamics in naïve and primed state cells at 129 × C57BL/6 genetic background. **Figure S6.** DNA recombination repair genes link to telomere maintenance in naïve mESCs. **Figure S7.** Methylation and histone modifications in naïve and primed ESCs. **Figure S8.** Histone modifications and Dnmt3b enrichment at naïve and primed pluripotent genes and specific TEs. **Figure S9.** Kap1-mediated H3K9me3 regulates *L1Md\_T* and *IAPez-int* transcription in naïve and primed PSCs. **Figure S10.** Uncropped scans of Western blot with molecular weight markers.

**Additional file 2: Table S1.** Differential gene transcriptome at passage 5.

**Additional file 3: Table S2.** Differential gene transcriptome at passage 15.

**Additional file 4: Table S3.** Differential TE transcriptome at passage 5.

**Additional file 5: Table S4.** Differential TE transcriptome at passage 15.

**Additional file 6: Table S5.** Summary of CNVs containing TEs in naïve and primed PSCs in two different genetic background cell lines.

**Additional file 7: Table S6.** Key resource table.

**Additional file 8.** Review history.

### Acknowledgements

We thank Yujie Zhao and Meng Zhao from Beijing Novogene Bioinformatics Technology Co., Ltd. for their Exome-seq services and Panpan Shi from our lab for assistance with CRISPR/Cas9 experiments.

### Review history

The review history is available as Additional file 8.

### Peer review information

Anahita Bishop was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Authors' contributions

H.F. performed most of the experiments and wrote the manuscript; W.Z. performed bioinformatics analyses of the RNA-seq, ChIP-seq data, and exome-seq data; N.L. performed the ChIP experiments; J.Y. performed the DNA damage and IF experiments; X.Y. performed the TEC and 4–8 cell injection experiments; C.T. provided the materials, X.L. discussed the data and revised manuscript, L.L. designed and advised the experiments, and revised the manuscript. All author(s) read and approved the final manuscript.

### Funding

This study was supported by funding from China National Key R&D Program (2018YFA0107002, 2018YFC1003004).

### Availability of data and materials

RNA-seq data and ChIP-seq data are available from Gene Expression Omnibus under accession number GEO: GSE140665 [107] and GSE154487 [108]. The accession number for Exome-seq data is NCBI Sequence Read Archive: PRJNA725383 [109]. All public data used in this study have been listed and well referenced in Additional file 7: Table S6, including *in vivo* embryos RNA-seq data: GSE98150 [104], RNA-seq data from *Setdb1* KO mESC: PRJNA544540 [70],

*G9a* KO mESC: GSE49669 [71], *Dnmt3b* KO mESC: GSE72855 [72] and *Kap1* KD mESC: GSE95720 [76], and H3K9me3 and *Kap1* ChIP-seq data: GSE94323 [75]. All cell lines used in this study have been authenticated and are available upon request.

## Declarations

### Ethics approval and consent to participate

Use of mice for this research was approved by the Nankai University Animal Care and Use Committee. All mice used in this study were taken care of and operated according to the relevant regulations. The mice were humanely euthanized for specific experiments.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>State Key Laboratory of Medicinal Chemical Biology, Nankai University, Tianjin, China. <sup>2</sup>Department of Cell Biology and Genetics, College of Life Sciences, Nankai University, Tianjin, China. <sup>3</sup>The Key Laboratory of Bioactive Materials Ministry of Education, Nankai University, Tianjin, China. <sup>4</sup>College of Pharmacy, Nankai University, 94 Weijin Road, Tianjin 300071, China.

Received: 28 October 2020 Accepted: 24 June 2021

Published online: 09 July 2021

## References

- Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, et al. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*. 2007;448(7150):191–5. <https://doi.org/10.1038/nature05950>.
- Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*. 2007;448(7150):196–9. <https://doi.org/10.1038/nature05972>.
- Nichols J, Smith A. Naive and primed pluripotent states. *Cell Stem Cell*. 2009;4(6):487–92. <https://doi.org/10.1016/j.stem.2009.05.015>.
- De Los AA, Ferrari F, Xi R, Fujiwara Y, Benvenisty N, Deng H, et al. Hallmarks of pluripotency. *Nature*. 2015;525:469–78.
- Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. *Nature*. 1981;292(5819):154–6. <https://doi.org/10.1038/292154a0>.
- Nagy A, Rossant J, Nagy R, Abramow-Newerly W, Roder JC. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci U S A*. 1993;90(18):8424–8. <https://doi.org/10.1073/pnas.90.18.8424>.
- Eggen K, Rode A, Jentsch I, Samuel C, Hennek T, Tintrup H, et al. Male and female mice derived from the same embryonic stem cell clone by tetraploid embryo complementation. *Nat Biotechnol*. 2002;20(5):455–9. <https://doi.org/10.1038/nbt0502-455>.
- Zhao XY, Li W, Lv Z, Liu L, Tong M, Hai T, et al. iPS cells produce viable mice through tetraploid complementation. *Nature*. 2009;461(7260):86–90. <https://doi.org/10.1038/nature08267>.
- Jaenisch R, Young R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell*. 2008;132(4):567–82. <https://doi.org/10.1016/j.cell.2008.01.015>.
- Guo G, Yang J, Nichols J, Hall JS, Eyres I, Mansfield W, et al. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development*. 2009;136(7):1063–9. <https://doi.org/10.1242/dev.030957>.
- Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, et al. Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell*. 2014;14(6):838–53. <https://doi.org/10.1016/j.stem.2014.04.003>.
- Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell*. 2011;146(4):519–32. <https://doi.org/10.1016/j.cell.2011.06.052>.
- Ghimire S, Van der Jeught M, Neupane J, Roost MS, Anckaert J, Popovic M, et al. Comparative analysis of naive, primed and ground state pluripotency in mouse embryonic stem cells originating from the same genetic background. *Sci Rep*. 2018;8(1):5884. <https://doi.org/10.1038/s41598-018-24051-5>.
- Tosolini M, Brochard V, Adenot P, Chebrou M, Grillo G, Navia V, et al. Contrasting epigenetic states of heterochromatin in the different types of mouse pluripotent stem cells. *Sci Rep*. 2018;8(1):5776. <https://doi.org/10.1038/s41598-018-23822-4>.
- Hackett JA, Surani MA. Regulatory principles of pluripotency: from the ground state up. *Cell Stem Cell*. 2014;15(4):416–30. <https://doi.org/10.1016/j.stem.2014.09.015>.
- Loh KM, Lim B, Ang LT. Ex uno plures: molecular designs for embryonic pluripotency. *Physiol Rev*. 2015;95(1):245–95. <https://doi.org/10.1152/physrev.00001.2014>.
- Wu J, Izpisua Belmonte JC. Dynamic pluripotent stem cell states and their applications. *Cell Stem Cell*. 2015;17(5):509–25. <https://doi.org/10.1016/j.stem.2015.10.009>.
- Weinberger L, Ayyash M, Novershtern N, Hanna JH. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat Rev Mol Cell Biol*. 2016;17(3):155–69. <https://doi.org/10.1038/nrm.2015.28>.
- Ying QL, Wray J, Nichols J, Battle-Morera L, Doble B, Woodgett J, et al. The ground state of embryonic stem cell self-renewal. *Nature*. 2008;453(7194):519–23. <https://doi.org/10.1038/nature06968>.
- Ficz G, Hore TA, Santos F, Lee HJ, Dean W, Arand J, et al. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell*. 2013;13(3):351–9. <https://doi.org/10.1016/j.stem.2013.06.004>.
- Leitch HG, McEwen KR, Turp A, Encheva V, Carroll T, Grabole N, et al. Naive pluripotency is associated with global DNA hypomethylation. *Nat Struct Mol Biol*. 2013;20(3):311–6. <https://doi.org/10.1038/nsmb.2510>.

22. Marks H, Kalkan T, Menafra R, Denissov S, Jones K, Hofemeister H, et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*. 2012;149(3):590–604. <https://doi.org/10.1016/j.cell.2012.03.026>.
23. Choi J, Huebner AJ, Clement K, Walsh RM, Savol A, Lin K, et al. Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. *Nature*. 2017;548(7666):219–23. <https://doi.org/10.1038/nature23274>.
24. Yagi M, Kishigami S, Tanaka A, Semi K, Mizutani E, Wakayama S, et al. Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation. *Nature*. 2017;548(7666):224–7. <https://doi.org/10.1038/nature23286>.
25. Guo R, Ye X, Yang J, Zhou Z, Tian C, Wang H, et al. Feeders facilitate telomere maintenance and chromosomal stability of embryonic stem cells. *Nat Commun*. 2018;9(1):2620. <https://doi.org/10.1038/s41467-018-05038-2>.
26. Fort A, Hashimoto K, Yamada D, Salimullah M, Keya CA, Saxena A, et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet*. 2014;46(6):558–66. <https://doi.org/10.1038/ng.2965>.
27. Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol*. 2014;21(4):423–5. <https://doi.org/10.1038/nsmb.2799>.
28. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*. 2012;487(7405):57–63. <https://doi.org/10.1038/nature11244>.
29. Wang J, Xie G, Singh M, Ghanbarian AT, Rasko T, Svetnik A, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature*. 2014;516(7531):405–9. <https://doi.org/10.1038/nature13804>.
30. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet*. 2010;42(7):631–4. <https://doi.org/10.1038/ng.600>.
31. Goke J, Lu X, Chan YS, Ng HH, Ly LH, Sachs F, et al. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell*. 2015;16(2):135–41. <https://doi.org/10.1016/j.stem.2015.01.005>.
32. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature*. 2015;522(7555):221–5. <https://doi.org/10.1038/nature14308>.
33. Gao L, Wu K, Liu Z, Yao X, Yuan S, Tao W, et al. Chromatin accessibility landscape in human early embryos and its association with evolution. *Cell*. 2018;173(1):248–59 e215. <https://doi.org/10.1016/j.cell.2018.02.028>.
34. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet*. 2009;10(10):691–703. <https://doi.org/10.1038/nrg2640>.
35. Leung DC, Lorincz MC. Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem Sci*. 2012;37(4):127–33. <https://doi.org/10.1016/j.tibs.2011.11.006>.
36. Thompson PJ, Macfarlan TS, Lorincz MC. Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol Cell*. 2016;62(5):766–76. <https://doi.org/10.1016/j.molcel.2016.03.029>.
37. Burns KH, Boeke JD. Human transposon tectonics. *Cell*. 2012;149(4):740–52. <https://doi.org/10.1016/j.cell.2012.04.019>.
38. Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*. 2010;463(7278):237–40. <https://doi.org/10.1038/nature08674>.
39. Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y, et al. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res*. 2013;23(3):452–61. <https://doi.org/10.1101/gr.147678.112>.
40. Elsasser SJ, Noh KM, Diaz N, Allis CD, Banaszynski LA. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature*. 2015;522:240–4.
41. Huang J, Wang F, Okuka M, Liu N, Ji G, Ye X, et al. Association of telomere length with authentic pluripotency of ES/iPS cells. *Cell Res*. 2011;21(5):779–92. <https://doi.org/10.1038/cr.2011.16>.
42. Marion RM, Strati K, Li H, Tejera A, Schoeftner S, Ortega S, et al. Telomeres acquire embryonic stem cell characteristics in induced pluripotent stem cells. *Cell Stem Cell*. 2009;4(2):141–54. <https://doi.org/10.1016/j.stem.2008.12.010>.
43. Pucci F, Gardano L, Harrington L. Short telomeres in ESCs lead to unstable differentiation. *Cell Stem Cell*. 2013;12(4):479–86. <https://doi.org/10.1016/j.stem.2013.01.018>.
44. Bao S, Tang F, Li X, Hayashi K, Gillich A, Lao K, et al. Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature*. 2009;461(7268):1292–5. <https://doi.org/10.1038/nature08534>.
45. Yeom FJ, Fuhrmann G, Ovitt CE, Brehm A, Ohbo K, Gross M, et al. Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonic cells. *Development*. 1996;122(3):881–94. <https://doi.org/10.1242/dev.122.3.881>.
46. Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X, et al. Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell*. 2010;6(5):468–78. <https://doi.org/10.1016/j.stem.2010.03.015>.
47. Hanna J, Markoulaki S, Mitalipova M, Cheng AW, Cassady JP, Staerk J, et al. Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell*. 2009;4(6):513–24. <https://doi.org/10.1016/j.stem.2009.04.015>.
48. Liu L, Michowski W, Kolodziejczyk A, Sicsinski P. The cell cycle in stem cell proliferation, pluripotency and differentiation. *Nat Cell Biol*. 2019;21(9):1060–7. <https://doi.org/10.1038/s41556-019-0384-4>.
49. Gonzales KA, Liang H, Lim YS, Chan YS, Yeo JC, Tan CP, et al. Deterministic restriction on pluripotent state dissolution by cell-cycle pathways. *Cell*. 2015;162(3):564–79. <https://doi.org/10.1016/j.cell.2015.07.001>.
50. Lombard DB, Chua KF, Mostoslavsky R, Franco S, Gostissa M, Alt FW. DNA repair, genome stability, and aging. *Cell*. 2005;120(4):497–512. <https://doi.org/10.1016/j.cell.2005.01.028>.
51. Dan J, Liu Y, Liu N, Chiourea M, Okuka M, Wu T, et al. Rif1 maintains telomere length homeostasis of ESCs by mediating heterochromatin silencing. *Dev Cell*. 2014;29(1):7–19. <https://doi.org/10.1016/j.devcel.2014.03.004>.
52. Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, Lee SL, et al. Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature*. 2010;464(7290):858–63. <https://doi.org/10.1038/nature08882>.
53. Zimmermann M, de Lange T. 53BP1: pro choice in DNA repair. *Trends Cell Biol*. 2014;24(2):108–17. <https://doi.org/10.1016/j.tcb.2013.09.003>.
54. Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL, et al. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell*. 2009;138(1):90–103. <https://doi.org/10.1016/j.cell.2009.06.021>.
55. Vannier JB, Sandhu S, Petalcorin MI, Wu X, Nabi Z, Ding H, et al. RTEL1 is a replisome-associated helicase that promotes telomere and genome-wide replication. *Science*. 2013;342(6155):239–42. <https://doi.org/10.1126/science.1241779>.



56. Arlt MF, Xu B, Durkin SG, Casper AM, Kastan MB, Glover TW. BRCA1 is required for common-fragile-site stability via its G2/M checkpoint function. *Mol Cell Biol*. 2004;24(15):6701–9. <https://doi.org/10.1128/MCB.24.15.6701-6709.2004>.
57. Zhao Y, Sfeir AJ, Zou Y, Buseman CM, Chow TT, Shay JW, et al. Telomere extension occurs at most chromosome ends and is uncoupled from fill-in in human cancer cells. *Cell*. 2009;138(3):463–75. <https://doi.org/10.1016/j.cell.2009.05.026>.
58. Cesare AJ, Reddel RR. Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet*. 2010;11(5):319–30. <https://doi.org/10.1038/nrg2763>.
59. Bailey SM, Brenneman MA, Goodwin EH. Frequent recombination in telomeric DNA may extend the proliferative life of telomerase-negative cells. *Nucleic Acids Res*. 2004;32(12):3743–51. <https://doi.org/10.1093/nar/gkh691>.
60. McKenna MJ, Robinson E, Goodwin EH, Cornforth MN, Bailey SM. Telomeres and NextGen CO-FISH: directional genomic hybridization (Telo-dGH). *Methods Mol Biol*. 2017;1587:103–12. [https://doi.org/10.1007/978-1-4939-6892-3\\_10](https://doi.org/10.1007/978-1-4939-6892-3_10).
61. Ourliac-Garnier I, Londono-Vallejo A. Telomere strand-specific length analysis by fluorescent in situ hybridization (Q-CO-FISH). *Methods Mol Biol*. 2017;1587:41–54. [https://doi.org/10.1007/978-1-4939-6892-3\\_4](https://doi.org/10.1007/978-1-4939-6892-3_4).
62. Yang J, Guo R, Wang H, Ye X, Zhou Z, Dan J, et al. Tet enzymes regulate telomere maintenance and chromosomal stability of mouse ESCs. *Cell Rep*. 2016;15(8):1809–21. <https://doi.org/10.1016/j.celrep.2016.04.058>.
63. Gonzalo S, Jaco I, Fraga MF, Chen T, Li E, Esteller M, et al. DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nat Cell Biol*. 2006;8(4):416–24. <https://doi.org/10.1038/ncb1386>.
64. Theunissen TW, Powell BE, Wang H, Mitalipova M, Faddah DA, Reddy J, et al. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell*. 2014;15(4):524–6. <https://doi.org/10.1016/j.stem.2014.09.003>.
65. Takashima Y, Guo G, Loos R, Nichols J, Ficiz G, Krueger F, et al. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell*. 2015;162(2):452–3. <https://doi.org/10.1016/j.cell.2015.06.052>.
66. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*. 2013;502(7472):472–9. <https://doi.org/10.1038/nature12750>.
67. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 2009;324(5929):930–5. <https://doi.org/10.1126/science.1170116>.
68. Yang WR, Ardeljan D, Pacyna CN, Payer LM, Burns KH. SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Res*. 2019;47(5):e27. <https://doi.org/10.1093/nar/gky1301>.
69. Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, et al. Cooperative binding of transcription factors orchestrates reprogramming. *Cell*. 2017;168(3):442–59 e420. <https://doi.org/10.1016/j.cell.2016.12.016>.
70. Wu K, Liu H, Wang Y, He J, Xu S, Chen Y, et al. SETDB1-mediated cell fate transition between 2C-like and pluripotent states. *Cell Rep*. 2020;30(1):25–36 e26. <https://doi.org/10.1016/j.celrep.2019.12.010>.
71. Mozzetta C, Pontis J, Fritsch L, Robin P, Portoso M, Proux C, et al. The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Mol Cell*. 2014;53(2):277–89. <https://doi.org/10.1016/j.molcel.2013.12.005>.
72. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, et al. Intragenic DNA methylation prevents spurious transcription initiation. *Nature*. 2017;543(7643):72–7. <https://doi.org/10.1038/nature21373>.
73. Whiddon JL, Langford AT, Wong CJ, Zhong JW, Tapscott SJ. Conservation and innovation in the DUX4-family gene network. *Nat Genet*. 2017;49(6):935–40. <https://doi.org/10.1038/ng.3846>.
74. Hendrickson PG, Dorais JA, Grow EJ, Whiddon JL, Lim JW, Wike CL, et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet*. 2017;49(6):925–34. <https://doi.org/10.1038/ng.3844>.
75. De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet*. 2017;49(6):941–5. <https://doi.org/10.1038/ng.3858>.
76. Coluccio A, Ecco G, Duc J, Offner S, Turelli P, Trono D. Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naive embryonic stem cells. *Epigenet Chromatin*. 2018;11:7.
77. Tichy ED, Stambrook PJ. DNA repair in murine embryonic stem cells and differentiated cells. *Exp Cell Res*. 2008;314(9):1929–36. <https://doi.org/10.1016/j.yexcr.2008.02.007>.
78. Zhao W, Steinfeld JB, Liang F, Chen X, Maranon DG, Jian Ma C, et al. BRCA1-BARD1 promotes RAD51-mediated homologous DNA pairing. *Nature*. 2017;550(7676):360–5. <https://doi.org/10.1038/nature24060>.
79. Liu L. Linking telomere regulation to stem cell pluripotency. *Trends Genet*. 2017;33(1):16–33. <https://doi.org/10.1016/j.tig.2016.10.007>.
80. Zhang W, Chen F, Chen R, Xie D, Yang J, Zhao X, et al. Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. *Nucleic Acids Res*. 2019;47(16):8485–501. <https://doi.org/10.1093/nar/gkz594>.
81. Zeng S, Liu L, Sun Y, Xie P, Hu L, Yuan D, et al. Telomerase-mediated telomere elongation from human blastocysts to embryonic stem cells. *J Cell Sci*. 2014;127(Pt 4):752–62. <https://doi.org/10.1242/jcs.131433>.
82. Murnane JP. Telomere dysfunction and chromosome instability. *Mutat Res*. 2012;730(1-2):28–36. <https://doi.org/10.1016/j.mrfmmm.2011.04.008>.
83. Vessoni AT, Zhang T, Quinet A, Jeong HC, Munroe M, Wood M, et al. Telomere erosion in human pluripotent stem cells leads to ATR-mediated mitotic catastrophe. *J Cell Biol*. 2021;220(6):e202011014. <https://doi.org/10.1083/jcb.202011014>.
84. Ruis P, Van Ly D, Borel V, Kafer GR, McCarthy A, Howell S, et al. TRF2-independent chromosome end protection during pluripotency. *Nature*. 2021;589(7840):103–9. <https://doi.org/10.1038/s41586-020-2960-y>.
85. Markiewicz-Potoczny M, Lobanova A, Loeb AM, Kirak O, Olbrich T, Ruiz S, et al. TRF2-mediated telomere protection is dispensable in pluripotent stem cells. *Nature*. 2021;589(7840):110–5. <https://doi.org/10.1038/s41586-020-2959-4>.
86. Jachowicz JW, Bing X, Pontabry J, Boskovic A, Rando OJ, Torres-Padilla ME. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat Genet*. 2017;49(10):1502–10. <https://doi.org/10.1038/ng.3945>.
87. Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, et al. A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell*. 2018;174(2):391–405 e319. <https://doi.org/10.1016/j.cell.2018.05.043>.
88. Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, et al. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature*. 2010;464(7290):927–31. <https://doi.org/10.1038/nature08858>.

89. Sharif J, Endo TA, Nakayama M, Karimi MM, Shimada M, Katsuyama K, et al. Activation of endogenous retroviruses in Dnmt1(-/-) ESCs involves disruption of SETDB1-mediated repression by NP95 binding to hemimethylated DNA. *Cell Stem Cell*. 2016;19(1):81–94. <https://doi.org/10.1016/j.stem.2016.03.013>.
90. Yang BX, El Farran CA, Guo HC, Yu T, Fang HT, Wang HF, et al. Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell*. 2015;163(1):230–45. <https://doi.org/10.1016/j.cell.2015.08.037>.
91. Deniz O, de la Rica L, Cheng KCL, Spensberger D, Branco MR. SETDB1 prevents TET2-dependent activation of IAP retroelements in naive embryonic stem cells. *Genome Biol*. 2018;19(1):6. <https://doi.org/10.1186/s13059-017-1376-y>.
92. Yang L, Liu X, Song L, Di A, Su G, Bai C, et al. Transient Dux expression facilitates nuclear transfer and induced pluripotent stem cell reprogramming. *EMBO Rep*. 2020;21:e50054.
93. Fu X, Djekidel MN, Zhang Y. A transcriptional roadmap for 2C-like-to-pluripotent state transition. *Sci Adv*. 2020;6:eay5181.
94. Kazazian HH Jr, Goodier JL. LINE drive: retrotransposition and genome instability. *Cell*. 2002;110(3):277–80. [https://doi.org/10.1016/S0092-8674\(02\)00868-1](https://doi.org/10.1016/S0092-8674(02)00868-1).
95. Alexanian M, Maric D, Jenkinson SP, Mina M, Friedman CE, Ting CC, et al. A transcribed enhancer dictates mesendoderm specification in pluripotency. *Nat Commun*. 2017;8(1):1806. <https://doi.org/10.1038/s41467-017-01804-w>.
96. Huang J, Deng K, Wu H, Liu Z, Chen Z, Cao S, et al. Efficient production of mice from embryonic stem cells injected into four- or eight-cell embryos by piezo micromanipulation. *Stem Cells*. 2008;26(7):1883–90. <https://doi.org/10.1634/stemcells.2008-0164>.
97. Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*. 2011;473(7347):398–402. <https://doi.org/10.1038/nature10008>.
98. Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res*. 2002;30(10):e47. <https://doi.org/10.1093/nar/30.10.e47>.
99. Liu L, Bailey SM, Okuka M, Munoz P, Li C, Zhou L, et al. Telomere lengthening early in development. *Nat Cell Biol*. 2007;9(12):1436–41. <https://doi.org/10.1038/ncb1664>.
100. Poon SS, Martens UM, Ward RK, Lansford PM. Telomere length measurements using digital fluorescence microscopy. *Cytometry*. 1999;36(4):267–78. [https://doi.org/10.1002/\(SICI\)1097-0320\(19990801\)36:4<267::AID-CYTO1>3.0.CO;2-O](https://doi.org/10.1002/(SICI)1097-0320(19990801)36:4<267::AID-CYTO1>3.0.CO;2-O).
101. Zhou M, Palanca AMS, Law JA. Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. *Nat Genet*. 2018;50(6):865–73. <https://doi.org/10.1038/s41588-018-0115-y>.
102. Gehre M, Bunina D, Sidoli S, Lubke MJ, Diaz N, Trovato M, et al. Lysine 4 of histone H3.3 is required for embryonic stem cell differentiation, histone enrichment at regulatory regions and transcription accuracy. *Nat Genet*. 2020;52(3):273–82. <https://doi.org/10.1038/s41588-020-0586-5>.
103. Brind'Amour J, Liu S, Hudson M, Chen C, Karimi MM, Lorincz MC. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nat Commun*. 2015;6(1):6033. <https://doi.org/10.1038/ncomms7033>.
104. Wang C, Liu X, Gao Y, Yang L, Li C, Liu W, et al. Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat Cell Biol*. 2018;20(5):620–31. <https://doi.org/10.1038/s41556-018-0093-4>.
105. Lange S, Engleitner T, Mueller S, Maresch R, Zwiebel M, Gonzalez-Silva L, et al. Analysis pipelines for cancer genome sequencing in mice. *Nat Protoc*. 2020;15(2):266–315. <https://doi.org/10.1038/s41596-019-0234-7>.
106. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *Plos Comput Biol*. 2016;12(4):e1004873. <https://doi.org/10.1371/journal.pcbi.1004873>.
107. Fu H, Zhang W, Li N, Yang J, Ye X, Tian C, Lu, and Liu L. Elevated retrotransposon activity and genomic instability in primed pluripotent stem cells. RNA-seq data GSE140665. 2021. <https://www.ncbi.nlm.nih.gov/search/all/?term>. Accessed 22 Aug 2020.
108. Fu H, Zhang W, Li N, Yang J, Ye X, Tian C, Lu, and Liu L. Elevated retrotransposon activity and genomic instability in primed pluripotent stem cells. ChIP-seq data GSE154487. 2021. <https://www.ncbi.nlm.nih.gov/search/all/?term>. Accessed 22 Aug 2020.
109. Fu H, Zhang W, Li N, Yang J, Ye X, Tian C, Lu, and Liu L. Elevated retrotransposon activity and genomic instability in primed pluripotent stem cells. Exome-seq data PRJNA725383. 2021. <https://www.ncbi.nlm.nih.gov/search/all/?term>. Accessed 8 Apr 2021.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

