

## Review Article

# Transposable Element Dynamics and Regulation during Zygotic Genome Activation in Mammalian Embryos and Embryonic Stem Cell Model Systems

Yixuan Low <sup>1</sup>, Dennis Eng Kiat Tan <sup>1</sup>, Zhenhua Hu <sup>1</sup>, Shawn Ying Xuan Tan <sup>1</sup>,  
and Wee-Wei Tee <sup>1,2</sup>

<sup>1</sup>*Chromatin Dynamics and Disease Epigenetics Group, Institute of Molecular and Cell Biology, Agency for Science, Technology, and Research (A\*STAR), Singapore 138673*

<sup>2</sup>*Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117593*

Correspondence should be addressed to Wee-Wei Tee; [wwtee@imcb.a-star.edu.sg](mailto:wwtee@imcb.a-star.edu.sg)

Received 21 May 2021; Revised 31 August 2021; Accepted 8 September 2021; Published 15 October 2021

Academic Editor: Xinyi Lu

Copyright © 2021 Yixuan Low et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Transposable elements (TEs) are mobile genetic sequences capable of duplicating and reintegrating at new regions within the genome. A growing body of evidence has demonstrated that these elements play important roles in host genome evolution, despite being traditionally viewed as parasitic elements. To prevent ectopic activation of TE transposition and transcription, they are epigenetically silenced in most somatic tissues. Intriguingly, a specific class of TEs—retrotransposons—is transiently expressed at discrete phases during mammalian development and has been linked to the establishment of totipotency during zygotic genome activation (ZGA). While mechanisms controlling TE regulation in somatic tissues have been extensively studied, the significance underlying the unique transcriptional reactivation of retrotransposons during ZGA is only beginning to be uncovered. In this review, we summarize the expression dynamics of key retrotransposons during ZGA, focusing on findings from *in vivo* totipotent embryos and *in vitro* totipotent-like embryonic stem cells (ESCs). We then dissect the functions of retrotransposons and discuss how their transcriptional activities are finetuned during early stages of mammalian development.

## 1. Introduction

Annotations of eukaryotic genomes have revealed that repetitive elements interspersed between protein-coding genes are prevalent [1–3], constituting up to two-thirds of the human genome [4]. TEs are DNA sequences that can reintegrate to other genomic regions within the same cell of origin. Based on their mechanism of transposition, TEs can be divided into two main classes: retrotransposons (class I) and DNA transposons (class II) (Figure 1) [1, 5]. DNA transposons are the smallest class of mobile genetic elements, making up approximately 3% of the human genome, and they copy themselves via a “cut and paste” mechanism [1, 6]. On the other hand, retrotransposons represent the largest class of TEs, approximately 37% of the human genome, and they

transpose through an RNA intermediate in a “copy and paste” mechanism [3, 6, 7]. Retrotransposons can be further subdivided into three subgroups, namely, the long terminal repeat (LTR) containing endogenous retroviruses, long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs) [6]. Notably, DNA transposons and most retrotransposons are no longer functional in mice and humans, owing to the accumulation of genetic mutations across evolutionary time [8].

Originally, TEs were thought to be genetic parasites [9–11]. Specifically, the transposition activity of TEs contributes to DNA rearrangements, deletions, and insertions, thereby threatening the host genome with deleterious disruptions to gene regulatory networks. Unsurprisingly, TEs and their spurious activities have been linked to various

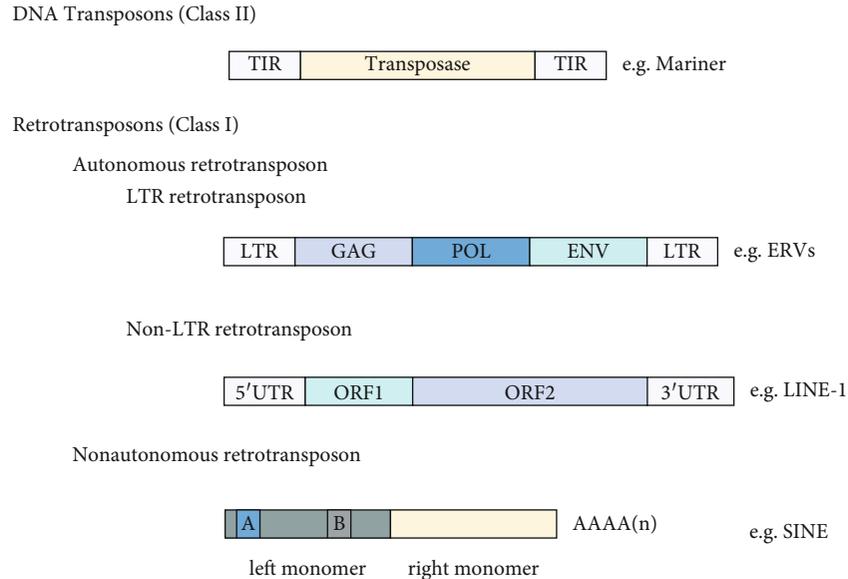


FIGURE 1: Schematic representations of key TE classes present in mammalian genomes. TEs can be broadly classified into DNA transposons (class II) and retrotransposons (class I). For a specific breakdown of each order, please refer to the review by [6]. Abbreviations: GAG: capsid protein; ENV: envelope protein; POL: coding region that encodes one or more of the following: proteinase, reverse transcriptase, RNaseH, polymerase, or integrase; TIR: terminal inverted repeat; LTR: long-terminal repeat; UTR: untranslated regions; EN: endonuclease; RT: reverse transcriptase; A and B: RNA polymerase III promoter box.

mutations and diseases [12–16]. However, this traditional view of TEs as parasitic elements is oversimplified, as seminal work by Barbara McClintock on TE regulation of neighboring genes in maize suggested that TEs could harbor rich sources of regulatory elements suitable for host co-option [17]. Work in later years indicated that selective forces may be acting to domesticate certain TEs for regulatory purposes, catalyzing the evolution of eukaryotic gene regulatory networks [18, 19]. Examples of these include RAG enzymes that are involved in the generation of antibody repertoire [20] and syncytin in placental development [21]. In normal development, a specific TE, LINE1, is expressed in neuronal progenitor cells and contributes to neuronal diversity [22]. Other studies further supplemented evidence supporting the notion that a significant fraction of TEs is implicated in transcriptional and epigenetic programs involved in development [21–26] and various phenotypes [27–30].

Notably, TEs have diverged so rapidly that even within mammals, their abundance and activities are highly variable [31]. In light of this, it is remarkable to observe a surge of transcriptional activation of TEs, more specifically retrotransposons, during preimplantation development across various mammalian embryos, albeit with differences in timing and class of retrotransposons [31]. In humans [32] and mice [33, 34], the increase in the transcription of species-specific retrotransposons is evident as early as in the zygote and is maintained up till the blastocyst stage. It remains unclear how disparate TE compositions across the mammalian genomes become involved in a highly conserved process. Yet, this conservation suggests that retrotransposons likely play a crucial role in mediating some aspects of preimplantation development. Although most retrotransposons are nonfunctional fossil remnants in the genome, some families,

such as LINE1, remain transposition-competent. However, the observed LINE1 retrotransposition activity is disproportionately low, given its high transcript abundance in mouse embryos [35–37]. Thus, the temporal upregulation of TE transcription during early mouse development may exert additional regulatory functions beyond the mere expansion of retrotransposons.

## 2. TE Expression Coincides with ZGA and Totipotency

Maternal-to-zygotic transition, also known as zygotic genome activation (ZGA), is the first major developmental transition after the fertilization of the gametes. During ZGA, maternally inherited transcripts are depleted, and the quiescent zygotic genome becomes transcriptionally active. ZGA occurs in two phases: (1) minor ZGA, characterized by the synthesis of a small set of transcripts from the paternal pronucleus, and (2) major ZGA, during which reprogramming of the gene regulatory networks and expression of stage-specific transcripts peaks [38–40]. This change in gene expression control is coordinated with changes in the cell cycle, chromatin state, and cellular contents. During ZGA, the parental genomes are epigenetically reset: heterochromatin is lost, DNA methylation depletes transiently [41], and histone mobility increases [42]. For an indepth discussion on epigenetic resetting during preimplantation development, please refer to these excellent reviews [43, 44]. Briefly, this reprogramming process gives rise to a more open chromatin architecture and coincides with the establishment of totipotency [42, 45, 46]. Crucially, two epigenetic machineries involved in the silencing of retrotransposons are remodeled during these early stages: DNA

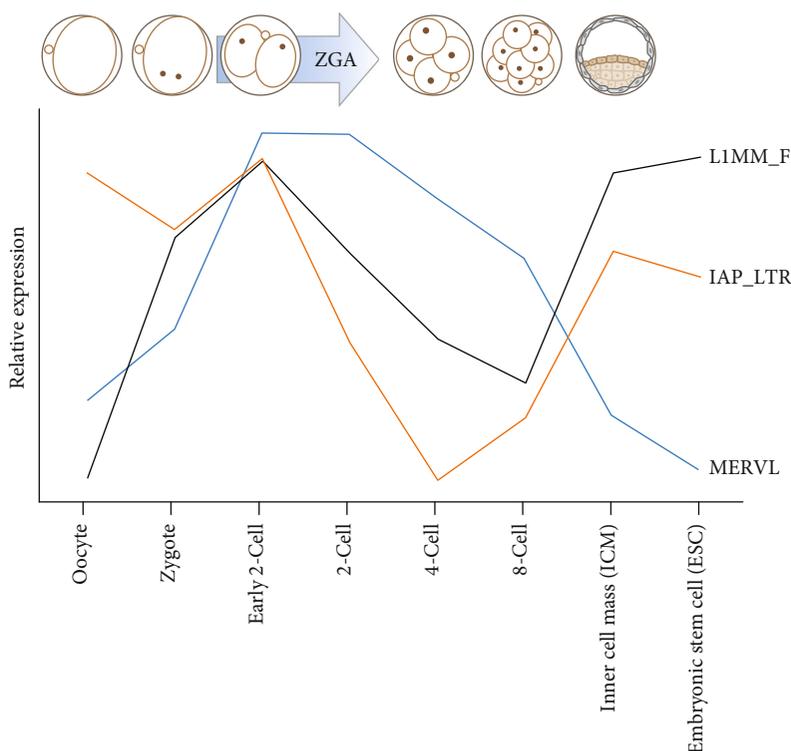


FIGURE 2: Graphical representation of the dynamics of LINE1, MuERV1, and IAP transcript levels during mouse preimplantation development and in mESCs. The relative expression levels of the TEs plotted are based on the analysis of published RNAseq datasets [45].

methylation and heterochromatic marks, such as H3K9me3. Deficiencies in these silencing mechanisms can lead to transcriptional activation of distinct TEs in preimplantation embryos, as well as in embryonic stem cells (ESCs). For instance, a subclass of retrotransposons, endogenous retroviruses (ERVs), is one of the earliest transcribed sequences in the mouse 2-cell embryos, during which maternal-to-zygotic transition ensues and cellular totipotency is established [47]. The expression of distinct ERV subclasses during various stages of preimplantation was also observed in human embryos [32, 48].

Early analyses of gene expression patterns in mouse preimplantation embryos revealed that several retrotransposons have distinct expression levels across different developmental stages (Figure 2). The expression of the earliest characterized retrotransposons include LINE1 and ERVs, namely, class II intracisternal A particle (IAP) and class III murine ERVL with leucine tRNA primer (MuERV1) and MaLRs (internally deleted nonautonomous counterparts of class III ERVL) [33, 49–52]. LINE-1 RNA is detected in the 1-cell zygotic stage embryos, peaks at the 2-cell stage, and remains active throughout embryonic development [33]. IAPs are expressed in oocytes that are rapidly degraded after fertilization and are suppressed during ZGA [53]. Intriguingly, IAPs are reexpressed as development progresses, and their expression peaks at the blastocyst stage [53]. Finally, MuERV1 and MaLR expression are restricted to the zygote and 2-cell stage embryos [34, 51, 52, 54].

Efforts to decipher the implications of the retrotransposon expression in the 2-cell context have been facilitated by

the discovery of an *in vitro* cellular model of totipotency, stemming from the work of Macfarlan et al. [54], who identified the existence of a rare subpopulation of mESCs that bear strong molecular and epigenetic resemblance to 2-cell stage embryos, termed 2C-like mESCs. (for a comprehensive review of 2C-like ESC features, refer to [55]). Briefly, both 2-cell embryos and 2C-like mESCs robustly express 2C-specific genes such as DUX [56] and ZSCAN4 [57–59], display high histone mobility [42, 60], and possess relaxed chromatin architectures [45, 56]. 2C-like mESCs also exhibit an increased propensity to contribute to the extra-embryonic lineage, reflective of their expanded cellular potency. Since then, numerous groups have employed the 2C-like system to elucidate molecular factors governing ZGA and uncover molecular features of 2-cell blastomeres [26, 42, 56, 61–83]. Notably, similar to the 2-cell embryos, 2C-like mESCs also exhibit high transcriptional output at LTR elements, specifically at MuERV1 elements [54]. Apart from 2C-like mESCs, other ESC models of expanded potency have also been reported; although, they exhibit distinct transcriptomic profiles [84–86]. It is also important to note that the epigenetic and transcriptomic profiles of 2C-like mESCs and 2-cell embryos exhibit differences [87], highlighting nuances between *in vivo* and *in vitro* experimental models. Notwithstanding, 2C-like mESCs represent a relevant and tractable model to study ZGA and totipotency outside of the early embryos. Below, we summarize the functional relevance of retrotransposons during development, using studies from both 2C-like mESCs and preimplantation embryos.

### 3. Functional Relevance of Retrotransposons during Preimplantation Development

**3.1. Retrotransposons Promote TE-Gene Chimeric Transcripts during Preimplantation Development.** Throughout evolution, there have been multiple precedents for the co-option of *cis*-regulatory elements of retrotransposons and their role in shaping cell type-specific gene networks. This includes TE contribution to the morphological diversification of the mammalian placenta [88] and regulation of tissue-specific gene expression [89, 90]. LTR elements being utilized as alternative promoters, enhancers, and exons of early embryonic genes have been extensively reported [32, 34, 47, 48, 56]. Indeed, characterizations of mouse oocytes, preimplantation embryos, and 2C-like mESCs revealed that LTRs of MuERV1 elements are co-opted as alternative promoters to drive the expression of a high volume of preimplantation-specific genes [34, 91]. These LTR-driven transcripts are termed chimeric transcripts. Similarly, specific primate-specific ERVs are robustly upregulated during each preimplantation stage [32, 48]. Thus, the contribution of retrotransposons, via its *cis*-regulatory elements or transcripts, has emerged as potentially playing a role in the transcriptional regulation of the totipotency program and ZGA.

Notably, it was reported that 90% of mapped chimeric 2C-like transcripts contain Open Reading Frames (ORFs) preceded by LTR elements fused to exons [54]. Moreover, a single-cell transcriptomic analysis of differentially upregulated genes in 2C-like mESCs revealed that a significant proportion contains an MT2\_mm (solo LTR of canonical MuERV1) element in close proximity [67], whereas this occurrence is absent in downregulated genes. Therefore, the widespread co-option of LTR-driven transcription could have evolved to coordinate the temporal rewiring of gene networks during ZGA by placing a large proportion of 2C-related genes under the regulation of LTR elements.

**3.2. LINE1 and MuERV1 Transcripts Regulate Gene Expression and Developmental Potency.** As previously mentioned, retrotransposons are also rich sources of noncoding regulatory elements. During mouse ZGA, LINE1 RNA performs a function similar to long noncoding RNAs, by recruiting nucleolin/KAP1 to repress 2C-specific transcription factor (TF) DUX and activate rRNA synthesis [26]. The inactivation of the LINE1 expression leads to developmental arrest in 2-cell embryos and promotes transition to the 2C-like state in mESCs [26]. These observations suggest that the *trans*-acting functions of LINE1 are crucial for the modulation of cellular identity during early development.

Overall, there is mounting evidence to suggest that retrotransposons play important roles during development *in vivo* and cellular plasticity *in vitro*. This is achieved via multiple mechanisms acting in both *cis* and *trans*. Indeed, class III ERVs, which occupy significant proportions of the oocyte and embryonic transcriptomes in both mouse and human, are required for developmental progression [92]. siRNA targeting of 80.5% of the MuERV1 elements in mice contributes to the failure of the chimeric transcript expres-

sion and a decrease in the GAG protein content [92]. Moreover, even a modest reduction of MuERV1 transcripts was sufficient to cause developmental delays, thereby implicating MuERV1 in the regulation of early developmental programs [92]. However, it remains unclear whether the resultant phenotype is a direct consequence of the loss of MuERV1 transcripts or the depletion of chimeric transcripts. The identification of the exact contributor to developmental abnormalities in MuERV1-depleted cells would be an important next step towards understanding the role of retrotransposons, in particular ERVs, in ZGA and totipotency.

**3.3. Retrotransposons May Organize Chromatin Architecture in Preimplantation Embryos.** The significance of the activation of retrotransposons during ZGA has not been entirely elucidated. In this regard, TE activation could simply be a consequence of increased chromatin accessibility during ZGA, or it could contribute to the unique chromatin features of early embryos. LINE1 and MuERV1 have been implicated in modulating chromatin accessibility and organization during early development. The expression of LINE1 in 2C-embryos promotes increased chromatin accessibility, and its subsequent depletion following ZGA is a prerequisite for developmental progression [62, 93]. In a functional study, premature silencing of LINE1 elements led to a decrease in chromatin accessibility, while prolonged activation prevents chromatin compaction and delays developmental progression. Notably, the transcription of LINE1 appears to predominantly impact chromatin structure without overt changes in the global gene expression [62]. These observations imply that LINE1 functions at the chromatin level regulate chromatin accessibility, via its transcriptional activation, and may contribute to the shaping of the early embryonic chromatin architecture *in vivo*. It is also interesting to note that while LINE1-overexpressing embryos showed developmental arrest, experimental induction of LINE1 chromatin decondensation with an acidic peptide resulted in a milder developmental phenotype [62], hinting at a potential function of the LINE1 transcript itself, perhaps a feedback loop to reinforce chromatin relaxation, that is yet to be elucidated.

Beyond a transcriptional function, MuERV1 elements are also involved in shaping the 3D genome during development, as evidenced by Hi-C analysis of MT2\_mm and canonical MuERV1, which revealed the establishment of local and global domain boundaries in both 2C-like mESCs and 2C embryo datasets [94], preprint). These domain boundaries are correlated with the transcriptional upregulation of genes downstream of these retrotransposons and increased chromatin accessibility. The potential role of MuERV1 in shaping chromatin structure is undoubtedly interesting, but this observation is preliminary and pending further review [94], preprint). While these studies report the involvement of MuERV1 and LINE1 in 3D genome organization, their significance in development remains unclear. In line with findings of LINE1-mediated chromatin accessibility, Kruse et al. [94], preprint) also demonstrated that MuERV1-driven domain organization is not related to its gene regulatory activity. Rather, this organization is likely

driven by DUX binding and precedes the activation of gene expression from MuERV1 elements within the domains. The purpose of these domain boundaries could be twofold: First, during the onset of ZGA, this could concentrate 2C regulatory factors, such as DUX, to promote transcription efficiency. Second, following ZGA, MuERV1-driven transcripts within these domains could be easily packaged into heterochromatic structures to facilitate developmental progression.

#### 4. Regulation of Transposable Element Expression during Early Development and ZGA

In the following section, we will outline the TFs and epigenetic mechanisms involved in the regulation of retrotransposons in the context of 2-cell embryos and 2C-like mESCs (Table 1).

##### 4.1. Epigenetic-Based Regulation

**4.1.1. Histone Modifications.** Constitutively repressive H3K9 histone methylation is required for the maintenance of the TE repression following preimplantation development [95] and in somatic cells [96, 97]. In mammals, there are numerous H3K9 histone lysine methyltransferases (KMTs), including Suv39h1, Suv39h2, G9a/GLP, and SETDB1. The depletion of some of these KMTs in mESCs has been shown to promote the activation of specific TEs and transition into the 2C-like state [98, 99]. Of the retrotransposons studied, IAPs are most robustly repressed via the SETDB1-TRIM28/KAP1 silencing complex [100]. Further compaction of IAP into heterochromatin is promoted by the H3K9me3-dependent recruitment of heterochromatin protein 1 (HP1) transcriptional repressor [98, 101, 102]. However, SETDB1-deficient mESCs do not exhibit strong upregulation of MuERV1, indicating that SETDB1-mediated H3K9me3 is likely not responsible for the silencing of MuERV1 elements in mESCs. Instead, MuERV1 elements are enriched for G9a-dependent H3K9me2 [47], and catalytically active G9a is required for silencing MuERV1 LTR-driven transcripts in mESCs [98]. Moreover, G9a depletion in mESCs led to the upregulation of LTR-driven transcripts and a subset of 2C genes [54, 98]. In the same vein, the genomic depletion of SETDB1 in oocytes correlates with the ectopic reactivation of several TEs including IAP and LINE1, but not MuERV1 [103]. Intriguingly, H3K9me3 is less enriched on MuERV1 elements in 2-cell embryos, suggesting that SETDB1 may have an indirect role in repressing MuERV1 elements [104]. Post-ZGA, MuERV1- and LTR-containing retrotransposons are then marked with H3K9me3 from the 4-cell stage onwards, and the H3 histone chaperone, CAF1, is crucial for this H3K9me3-mediated LTR silencing [104].

Krüppel-associated box zinc finger proteins (KRAB-ZFPs) are crucial in mediating TE silencing [101, 105] via its KRAB domain, which contains specific DNA-binding regions and interacts with epigenetic modifiers. This ability to bind at specific genomic sites enables KRAB-ZFPs to

TABLE 1: Factors with roles in the regulation of the retrotransposon expression during (1) mouse ZGA and (2) in 2C-like mESCs. Not all factors demonstrated to be involved in retrotransposon regulation in ZGA are involved in the context of 2C-like mESCs, and vice versa.

	ZGA	2C-like mESCs
<i>Transcription factors</i>		
DUX	[63]	[56, 63, 64, 164]
p53	-	[64]
DPPA2/DPPA4	-	[77, 78]
NELFA	-	[65]
GATA2	[80]	[80]
ZSCAN4	-	[57, 59, 77, 186]
<i>Posttranscriptional regulators</i>		
miR-344	-	[71]
miR-34a	-	[80]
<i>Posttranslational modifiers</i>		
SUMO2	-	[161]
PIAS4	[79]	[79]
<i>Chromatin-associated regulators</i>		
KMTs (Suv39h1, Suv39h2, G9a/GLP, SETDB1)	[104]	[98, 99]
KDM1A/LSD1	[117]	[47]
ZMYM2	[71]	[71]
FACT complex	-	[116]
USP7	-	[67, 116]
DNMT1-UHRF1	[139]	-
TET2-PSPC1	-	[83]
CAF1	-	[66]
GBAF	-	[187]
STELLA	[92]	-
SMCHD1	[179]	[70]
LINE1	[26, 163]	[26, 72, 163]
ZFPs (ZFP809, RYBP, REX1)	[115]	[74, 114, 195]
HP1	-	[98]
TRIM28/KAP1	-	[98, 101]
RIF1	-	[160]

direct sequence-specific epigenetic silencing. Epigenetic modifiers such as KMTs (SETDB1), TRIM28/KAP1 scaffold, DNMTs, HP1, and nucleosome remodelers, KDM1A/LSD1, and histone deacetylation (NuRD) complex can then be specifically targeted to TE sites [106–111]. Notably, the role of ZFP-TRIM28/KAP1 interaction in regulating development and pluripotency has been well established [101, 112, 113]. In mESCs, SETDB1-TRIM28/KAP1 is recruited by ZFP809 for retrotransposon repression [114, 115]. The expression of ZFP809 is particularly interesting because ZFP809 and its MT2<sub>mm</sub> initiated chimeric transcripts are both robustly expressed in 2C-like mESCs [45, 114, 116]. This suggests that the MT2<sub>mm</sub> expression in 2C-like mESCs could potentiate its own repression by promoting the expression

The physiological function of mCH remains poorly described in both mouse and human. Based on current profiling efforts, mCH is found in human ESCs (hESCs) [144], induced pluripotent stem cells (iPSCs) [145], brain neurons [146, 147], oocytes [148, 149], PGCs [150, 151], and select organs [152]. Interestingly, there is an unusually high abundance of mCH in the oocyte, enriched at neighboring genomic regions with high-density mCG levels, but its functional significance is presently unclear [148, 153]. In the context of development, mCH is more enriched in oocytes than in sperm cells [154], and both mCG and mCH are lost from the zygote upon fertilization [153]. Interestingly, mCH is highly enriched on repetitive elements, suggesting a potential repressive role that is similar to mCG [155, 156]. Indeed, mCH appears to contribute to retrotransposon silencing during spermatogenesis [157]. DNMT3L, which is catalytically inactive, plays a crucial adapter role in establishing 5mC in germ cells [158]. In a DNMT3L loss-of-function study, a mutation to its ADD domain (responsible for interaction with histone H3) led to the ectopic reactivation of IAPs and LINEs in male germ cells [157]. When compared to controls, the most significant form of 5mC that was lost was in the context of mCH where in total, and only 15% (mCHH) and 11% (mCHG) were retained in the mutant respectively, in contrast to mCG (84%) [157]. The authors further verified that mCH is indeed abundant on both LINEs and IAPs, which is lost in the ADD mutant, suggesting that mCH likely plays a role in suppressing the expression of IAPs and LINEs, or perhaps retrotransposons broadly. In support of this, hESCs also harbor elevated mCH levels at SINE repeats [159]. However, it is important to note that although mCH loss is extensive on both LINEs and IAPs, mCG levels are also partially depleted in the DNMT3L-ADD mutant. Thus, the exact contribution by mCH to retrotransposon repression requires further investigation [157]. Future targeted epigenetic engineering experiments are critical to clarify the function of mCH in TE regulation and its function in development. Additionally, given that mCH is also enriched in cell types lacking DNMT3L, further studies are required to understand how noncanonical methylation is established in those settings.

Box 1: Possible Role of mCH in Development and Regulation of TEs.

of repressive factors, such as ZFP809, thereby creating an autofeedback loop. Another ZFP is RYBP, which is found to be crucial in the repression of retrotransposons and 2C-specific genes in mESCs, implicating RYBP in the exit from the ZGA program [74].

One of the earliest identified roadblocks to 2C-like reprogramming is the histone H3K4/K9 demethylase, KDM1A/LSD1. KDM1A/LSD1 mutant mESCs harbor significantly higher MuERVL transcript levels [47], indicating that repressive heterochromatin restrains MuERVL transcription [54, 117]. Interestingly, KDM1A/LSD1 is a maternally inherited factor, and its depletion leads to lethality in embryos prior to gastrulation [118]. Furthermore, a requirement of KDM1A/LSD1-mediated chromatin compaction for the exit from ZGA was demonstrated in KDM1A/LSD1 knockout (KO) zygotes that showed developmental arrest at the 2-cell stage [117]. Notably, KDM1A/LSD1 KO zygotes displayed robust upregulation of LINE1 transcripts, but not upregulation of MuERVL [117]. This is in direct contrast to KDM1A/LSD1 mutant mESCs, suggesting that multiple epigenetic regulators likely act in concert to activate MuERVL during ZGA *in vivo*.

Recently, another histone chaperone, FACT, which mediates H2A/H2B exchange, has also been implicated in the pluripotency-to-2C transition. Specifically, FACT recruits the H2B deubiquitinase USP7 to repress MuERVL- and LTR-driven chimeric transcript expression in mESCs [116]. Loss of either FACT or USP7 in mESCs led to robust upregulation of MuERVL and chimeric transcripts driven by MuERVL, concomitant with the expression of 2C-specific genes. Notably, this finding agrees with the siRNA screen performed by Rodriguez-Terrones et al., which identified ubiquitination pathway proteins, including USP7, as major 2C-like reprogramming roadblocks [73].

**4.1.2. DNA Modifications.** DNA methylation (5mC) is the most abundant epigenetic modification that plays a major

role in the silencing of retrotransposons [119, 120]. 5mC exists in two contexts—canonical in CpG dinucleotides (mCG) and noncanonical CH (mCH, where  $H = A, C, \text{ or } T$ ). 5mC is established by a group of highly conserved DNA methyltransferases (DNMTs), namely, DNMT1, which preferentially methylates hemimethylated CpG dinucleotides to maintain the 5mC landscape, and DNMT3A/B, which perform *de novo* methylation at unmethylated CpG [121]. 5mC level changes dynamically throughout development. Here, we summarize the 5mC landscape and its changes, including oxidized 5mC, during development, as well as the current knowledge on mCH in development in Box 1.

**4.1.3. Dynamic Changes in Levels of 5mC and Its Oxidized Derivatives during Development.** During preimplantation development, the developing zygote undergoes two waves of DNA demethylation. Shortly after fertilization, both maternal and paternal genomes are globally demethylated in the zygotes [122]. 5mC levels reach a relatively low level in the preimplantation embryo, which is followed by increased methylation after the onset of implantation [123]. The second wave of demethylation then occurs in primordial germ cells (PGCs) of the postimplantation embryo. In the first wave, the rapid loss of 5mC is mediated by ten-eleven translocation 3 (TET3), an  $\alpha$ -ketoglutarate dependent methylcytosine dioxygenase, which iteratively oxidizes 5mC to generate 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) [124, 125]. In contrast, 5mC levels of the maternal genome remain largely unchanged, with a less pronounced accumulation of 5hmC [124–126]. Interestingly, DNA demethylation during the first wave is still an area of intense debate, in which both active and passive models have been proposed. In the latter model, DNA demethylation is thought to be achieved through successive rounds of replication-dependent dilution and lack of 5mC maintenance, as the maternally supplied DNMT1 is excluded from

the nucleus during replication [124, 127, 128]. In support of this, the deliberate inhibition of DNA replication blocks DNA demethylation independently of TET3 activity [126]. However, a recent study suggested that DNA demethylation can still occur in the absence of replication in the developing zygote through active DNA demethylation [129].

**4.1.4. TE Reactivation Coincides with 5mC Loss and Gain of 5hmC.** The loss and gain of 5mC and 5hmC correlated with increased chromatin accessibility and expression of retrotransposons such as LINE1 and MuERVL, but not IAPs, during development. The function of 5mC in the silencing retrotransposon expression has been well characterized and is exemplified by the following DNMT1 loss-of-function study: in mouse midgastrulation embryos lacking DNMT1, the normally silenced IAPs become aberrantly reactivated from the loss of 5mC, leading to developmental delays and embryonic lethality [130]. In embryos lacking DNMT3A/B, MuERVL and 2C transcripts were strongly induced, and this activation is correlated with the global loss of 5mC [131]. Interestingly, IAPs were not reactivated in the DNMT3A/B KO embryos. Mechanistically, DNMT3A/B KO embryos were found to possess elevated expression of 2C-specific TFs DUX, DPPA2, and DPPA4, which are presumably activated due to the loss of the 5mC suppression, thereby potentially explaining the transcription of 2C genes [131]. However, there was no observable increase in the MuERVL and 2C-specific TF expression in DNMT1 KO embryos, despite their significant loss of 5mC [131]. These observations suggest that retrotransposons may be differentially regulated by maintenance and de novo establishment of 5mC and additionally points to a possible noncatalytic function of DNMT3A/B unique to embryos that has yet to be established.

Intriguingly, unlike in embryos, the loss of 5mC in DNMT-KO mESCs does not result in the reactivation of retrotransposons. Specifically, DNMT1 ablation [81, 132] and triple KO of DNMT1, DNMT3A, and DNMT3B [81, 133] in mESCs did not significantly upregulate MuERVL transcription, despite significant global DNA hypomethylation. To address this conundrum, Walter et al. examined the dynamics of retrotransposon reactivation in mESCs in response to global 5mC loss, by readapting the cells in serum-free 2i (PD0325901, CHIR99021) media with vitamin C supplementation. The combined chemical treatment resulted in extensive 5mC depletion that led to the reactivation of select retrotransposons (LINE1, IAP, and MuERVL) [81, 134, 135]. However, following this initial reactivation, the retrotransposons were eventually silenced by acquiring repressive histone modifications. It was found that whereas H3K9me2 levels were globally diminished, and H3K9me3 marks remained unchanged. Importantly, retrotransposons rapidly gained H3K27me3 in response to the loss of 5mC-mediated silencing [134], thereby illuminating an “epigenetic switch” in adaptation of 5mC loss to repressive histone pathways (H3K9me3 and H3K27me3) to maintain retrotransposon dynamics.

This adaptation and alteration in epigenetic repressive mechanisms are also observed during preimplantation

development. For instance, H3K9me3 is implicated in repressing select retrotransposons in the absence of active 5mC deposition. Specifically, IAPs and some ERVs are marked with H3K9me3, and these retrotransposons resist 5mC loss during ZGA. Mechanistically, it is suggested that 5mC levels at these retrotransposons are maintained by UHRF1, which recognizes H3K9me3 [136–139]. UHRF1, a cofactor of DNMT1, is essential for maintaining 5mC levels on IAPs in preimplantation embryos [139]. This adaptation is further exemplified by the silencing of MuERVL- and LTR-containing retrotransposons post-ZGA, during which MuERVL and LTRs rapidly acquire H3K9me3 and H3K27me3 following ZGA from the 4-cell to the late blastocyst stage [104]. This increase in H3K9/27me3 correlates with MuERVL- and LTR-containing retrotransposon repression during this period in the absence of 5mC, as zygotic DNMTs are only expressed in the late blastocyst. Collectively, these observations in both early embryos and mESCs highlight the importance of H3K9me3- and H3K27me3-mediated chromatin pathways in retrotransposon silencing in the absence of 5mC.

However, the role of TETs and 5hmC in regulating the retrotransposon expression is less straightforward. In TET1/3 KO mouse embryos, 5mC levels at LTR (IAPs) and non-LTR (LINE, SINE) retrotransposons are higher than controls, which correlate with lower expression of these TEs [140]. Genetic ablation of GADD45, a key interactor of TET enzymes and a component of the DNA demethylation machinery, hindered 2C entry in mESCs [76]. GADD45 is an adapter that directs and tethers TETs to genomic loci for DNA demethylation [141]. The loss of GADD45 function negatively impacts DNA demethylation, as the recruitment of TETs and the required accessory cofactors is affected. GADD45 (GADD45a, GADD45b, GADD45g) triple knockout (TKO) mESCs exhibit higher levels of 5mC when compared to controls and consequently impaired expression of prototypic 2C genes. This result corroborates a previous observation of global 5mC loss during 2C state cycling in mESCs [81]. GADD45 double knockout (DKO) embryos are sublethal, showing impaired upregulation of ZGA-associated genes and reduced implantation success [76]. Nonetheless, some LINE1 elements are upregulated in the DKO embryos, suggesting a repressive role of TET enzymes on LINE1. Indeed, TET1/2 has been reported to repress LINE1 in mESCs [142]. Interestingly, even though TET1/2 and 5hmC are enriched at young LINE1 elements, this DNA demethylation did not result in the reactivation of LINE1, as TET1 was found to recruit the SIN3A corepressive complex to maintain LINE1 silencing in the absence of 5mC [142].

In addition to LINE1, TET2 has also been found to repress MuERVL in mESCs [83]. Mechanistically, the RNA-binding protein paraspeckle component 1 (PSPC1) recruits TET2 to posttranscriptionally destabilize MuERVL and MuERVL-driven RNAs through 5hmC modifications [83]. PSPC1-TET2 can also recruit histone deacetylase 1 and 2 (HDAC1/2) to repress MuERVL transcription. Loss of PSPC1 not only drives the expression of MuERVL but also a subset of 2C-like genes. Notably, this effect of

PSPC1-TET2 regulation is specific to TE classes. For example, PSPC1-TET2 interaction transcriptionally activates the class II ERVK (IAP and MusD) expression, but not class II MuERVL in mESCs. Unlike TET3, the TET2 expression is low in 2-cell embryos and only increases during the blastocyst stage [143]. As such, it would be interesting to dissect the different roles of each TET member and further investigate whether PSPC1-mediated TET2 hydroxymethylation could be involved in modulating ZGA exit *in vivo*. In summary, TETs and 5hmC may potentially exert dual roles—first to relieve 5mC repression during ZGA and second to repress retrotransposon expression as development progresses to later stages.

**4.1.5. Histone Chaperones.** 2C-like mESCs and 2-cell embryos are known to display higher chromatin mobility [42]. A crucial roadblock to 2C-like reprogramming in mESCs is CAF1, a replication-coupled H3/H4 histone chaperone [66]. In 2-cell embryos, it was reported that the p60 subunit of CAF1 is transiently depleted from the replicating chromatin in the early S phase, indicating a decoupling of chromatin assembly with replication during ZGA. This correlates with ATAC-seq observations of large stretches of highly accessible chromatin regions, including MuERVL, in the early 2-cell embryos [45]. This delayed chromatin assembly may transiently render the chromatin more accessible, thereby promoting TE and other 2C gene expressions. Additionally, CAF1 is also responsible for mediating the deposition of repressive H3K9me3 on LTRs, thereby protecting preimplantation embryos from endogenous retrotransposon expression post-ZGA [104]. Interestingly, the replication-associated factor, RIF1, has also been identified to negatively regulate the MuERVL expression in both mESCs and hESCs, and it inhibits mESC transition to the 2C-like state [160]. RIF1 recruits histone modifiers and promotes the establishment of repressive histone marks and DNA methylation, possibly via its interaction with KMTs [160] and CAF1 [161]. Furthermore, the transcriptomic profiles of CAF1 KD and RIF1 KD 2C-like mESCs are highly similar, suggesting that both factors could function in the same axis [160].

**4.1.6. Noncoding RNAs.** TEs are rich sources of *trans*-acting factors. Notably, TE-derived sequences are highly overrepresented in vertebrate noncoding RNAs, including lncRNAs, siRNAs, piRNAs, and microRNAs [162]. For example, LINE1 RNA acts as a scaffold to recruit the RNA-binding proteins nucleolin and KAP1 and together regulate the exit of the 2C-like state in mESCs [26]. The importance of LINE1 RNA functioning as a scaffold in regulating the 2C-like state and 2-cell embryos is further highlighted by two recent studies demonstrating that N<sup>6</sup>-adenosine methyltransferase (METTL3) and YT521-B homology domain C1 (YTHDC1) m<sup>6</sup>A mRNA reader modulate 2C-like transitions in mESCs [72]. Mechanistically, YTHDC1 binds to m<sup>6</sup>A-modified LINE1 transcripts and facilitates the recruitment of nucleolin and KAP1 to the LINE1 scaffold [72, 163]. Loss of either YTHDC1 or METTL3 results in a depletion of H3K9me3 on the gene bodies of retrotransposons and robust activation of

the 2C-like program, including MuERVL [72, 163]. Moreover, YTHDC1 KO embryos displayed developmental defects, reinforcing the importance of YTHDC1-LINE1-nucleolin-KAP1 in finetuning the transcriptional activity of 2C genes and retrotransposons during early development. This finding also hints at another instance of retrotransposons functioning to regulate their own expression via a feedback loop, wherein LINE1 RNA-bound YTHDC1 is specifically recruited to TE gene bodies. Altogether, a diverse cast of epigenetic regulators contributes to the enforcement of specific and timely TE activity.

## 4.2. Transcriptional Regulation of Retrotransposons

**4.2.1. DUX Pioneer Factor Directly Activates MuERVL Transcription.** In mouse, one of the key TFs involved in the activation of MuERVL is DUX [56, 63, 164], a double homeodomain TF conserved amongst mammals [165]. DUX (DUX4 in human) was first identified to be aberrantly expressed in facioscapulohumeral muscular dystrophy (FSHD) in humans, a disorder that is characterized by an unusually high transcriptional output of ERVs [166]. In mESCs, the ectopic expression of DUX results in the transcriptional activation of MuERVL LTRs and a subset of the ZGA transcriptome corresponding to the 2-cell stage embryo [56, 63, 164]. *In vivo*, DUX was also observed to be upregulated in the early 2-cell embryos, positioning it as a pioneer factor for ZGA [56, 63]. DUX regulates the expression of MuERVL via its interaction with DUX recognition motifs present on LTRs. In fact, a significant proportion of genes expressed during ZGA is in close proximity to LTRs bearing DUX binding motifs. In parallel, the DUX4 overexpression in hESCs triggers the expression of HERVL and a subset of ZGA genes that are expressed in 4-cell human embryos [56]. Throughout evolution, there have been precedents of convergent co-option of TEs as regulatory regions for specific gene networks that define specific cellular states [167]. Therefore, the coevolution of DUX and MuERVL may serve as a means to coordinate ZGA, a complex process governed by multiple genes, with TEs serving as alternative promoters that can only be activated during ZGA by the 2-cell stage specific DUX TF.

A notable target of DUX is the microRNA, miR-344. DUX binds to miR-344's promoter to activate its expression, which then posttranscriptionally represses ZMYM2, a recruiter and stabilizer of KDM1A/LSD1 [71]. The overexpression of miR-344 alone is sufficient to induce 2C-gene and MuERVL expression, indicating that miR-344 is a robust activator of the 2C-like state downstream of DUX. Importantly, transient siRNA knockdown of ZMYM2 in zygotes leads to developmental arrest at the 2-cell stage and more robust MuERVL expression, implicating the DUX-miR-344-ZMYM2-KDM1A/LSD1 axis in regulating the TE expression during ZGA.

**4.2.2. Multiple Maternally Inherited Factors Promote MuERVL and LINE1 Transcription.** Although DUX is a key inducer of the 2C gene expression program, the DUX expression only begins during minor ZGA [165]. This

suggests that upstream maternal factors may be involved in the activation of ZGA, either in a DUX-dependent or DUX-independent manner. Indeed, high expression levels of maternally inherited TFs, DPPA2, and DPPA4 are observed in 2-cell embryos and 2C-like mESCs [77–79, 168]. The overexpression of DPPA2 and DPPA4 induces 2C-like transitions in mESCs, as well as the expression of MuERVL and LINE1 transcripts through the transcriptional activation of DUX [77]. Importantly, depletion of DPPA2 and DPPA4 significantly reduced the efficiency of 2C-like induction in mESCs [77, 78].

In addition to promoting DUX activation, DPPA2 and DPPA4 are implicated in shaping the epigenetic landscape of LINE1 elements that harbor DPPA2-binding sites at their 5' ends [169]. Mechanistically, DPPA2 and DPPA4 prevent de novo DNA methylation at LINE1 elements so that they remain competent for reactivation during lineage specification [169]. As previously discussed, LINE1 promotes chromatin relaxation during ZGA; in this case, DPPA2 and DPPA4 upregulation during ZGA may also contribute to LINE1-mediated chromatin accessibility.

What regulates DPPA2 and DPPA4? DPPA2 is regulated posttranslationally by the sumo ligase PIAS4, which sumoylates and inactivates DPPA2. Accordingly, PIAS4 is downregulated in 2-cell embryos, during which DPPA2 is active and MuERVL is robustly transcribed [79]. Notably, in a proteomic screen, SUMO2/3 was also found to be involved in DPPA2 and DPPA4 inactivation and impediment of reprogramming to the 2C-like state [170]. The ectopic overexpression of PIAS4 in zygotes impaired the activation of the ZGA program and 2-cell specific genes, including MuERVL, suggesting that PIAS4 inhibition of DPPA2 indirectly regulates the MuERVL expression. This finding also indicates that the SUMO pathway may be implicated in the modulation of 2C-like transition and ZGA. In support of this, sumoylation of PRC1.6 components contributes to the repression of DUX, a potent activator of the MuERVL expression [171]. Furthermore, the SUMO pathway has also emerged as a repressor of MuERVL in mESCs through a genome-wide siRNA screen for proviral repressors [161].

Apart from DPPA2/4, the mammalian-specific factor STELLA (encoded by *Dppa3*) is also maternally inherited and is required for proper preimplantation development of mouse embryos [92, 172, 173]. Early studies have implicated STELLA in the protection of the maternal pronucleus from TET3-mediated active demethylation in oocytes and maintenance of DNA methylation at a subset of imprinted genes and retrotransposons [174, 175]. However, the dispensability of TET3 and maintenance of imprinted genes for preimplantation development indicate that STELLA could play additional roles in preimplantation development [176, 177]. Indeed, Huang et al. demonstrated that STELLA maternal/zygotic knockout (M/Z KO) 2-cell embryos showed impairment in ZGA and a failure to upregulate 2-cell, LTR-driven, and MuERVL transcripts [92]. The aberrant MuERVL expression is directly attributed to the loss of STELLA, given that MuERVL can be activated in arrested 2-cell embryos with functional STELLA [117]. Interestingly, the overexpression or loss of *Dppa3* did not significantly

alter the MuERVL expression levels in mESCs, highlighting context-specific differences.

In type 2 FSHD (FSHD2), the DUX4 overexpression is most often due to loss-of-function of the structural maintenance of chromosomes hinge domain 1 (SMCHD1) gene [178]. In mice, SMCHD1 is also responsible for the DUX repression following ZGA. SMCHD1 mRNA transcript and protein are maternally inherited, and transient depletion of SMCHD1 mRNA in zygotes leads to the protracted DUX expression and developmental defects [179]. SMCHD1 KO mESCs also display upregulation of DUX and MuERVL [70]. Mechanistically, SMCHD1 sequesters TET proteins from DUX promoter, leading to hypermethylation and silencing of DUX. Recently, our lab also identified a maternal factor, negative elongation factor A (NELFA), which partners with DNA topoisomerase 2A (TOP2A) to promote 2C genes and TE expression in mESCs. We determined that the NELFA overexpression in mESCs is sufficient to activate the 2C program, including robust DUX and MuERVL activation [65]. Consistent with our observations, a role of TOP2A-mediated DNA double-strand breaks was implicated in ZGA in the *C.elegans* germline [180].

Several evidence also point to a role of the DNA damage response in ZGA/2C induction [59, 61, 181–183]. Notably, in a recent study by Grow et al. [64], it was found that p53, a key transcriptional effector of DNA damage, mediates the expression of DUX and 2C-specific transcripts, including MuERVL in mESCs [64]. Crucially, p53 is maternally inherited, and its activation coincides with the accumulation of endogenous DNA damage in the early embryos. Similar to the observed requirement of p53 to activate DUX and 2C transcript expression in mESCs, p53 maternal/zygotic KO embryos showed lower, but not complete elimination of DUX and 2C gene expression levels. Importantly, DUX4 in human iPSCs is also activated by p53, suggesting that p53 could play similar regulatory roles during human ZGA.

**4.2.3. GATA2-miR-34a Axis Regulates the MuERVL Expression.** GATA2 is another TF implicated in regulating the 2-cell gene network in mESCs and is under the regulation of the microRNA, miR-34a [80]. Loss of miR-34a resulted in the upregulation of MuERVL elements. Investigation of the 18 most highly upregulated MuERVL loci revealed the presence of GATA2 TF binding sites [80]. Interestingly, not only is GATA2 upregulated in 2-cell embryos but its expression pattern is also correlated to the MuERVL expression during preimplantation development, thereby implicating GATA2 as a transcriptional activator of MuERVL in 2-cell embryos. Indeed, GATA2 can bind to MuERVL LTRs when overexpressed in miR-34a KO mESCs, and loss of GATA2 led to the impaired MuERVL expression in these cells. However, the overexpression of GATA2 alone in wildtype mESCs is not sufficient to induce the MuERVL expression, suggesting that other miR-34a targets in addition to GATA2 may be required to cooperatively bind and activate the MuERVL expression. Another notable observation is that, similar to what was observed in DUX KO mice, miR-34a KO embryos can undergo successful

preimplantation development [80], reinforcing the notion that MuERVL regulation during ZGA is modulated by complex and partially redundant regulatory networks.

**4.2.4. ZSCAN4 Is Both a TF and Scaffold for Chromatin Modifiers at MuERVL Sites.** ZSCAN4 is another 2C-specific TF that is robustly expressed in both late 2-cell embryos and 2C-like cells. It consists of several paralogs (ZSCAN4a-f and three pseudogenes ZSCAN4-ps1-3) [58]. ZSCAN4 functions to maintain telomere length and ensure genomic integrity in mESCs [59, 184, 185]. This gene cluster has also been implicated in promoting 2C-like transition and activating preimplantation genes [57, 59, 81, 186]. Moreover, ZSCAN4-depleted embryos display developmental delays during ZGA [58]. In agreement with these observations, ZSCAN4c, in particular, directly binds to and activates the enhancer of MT2\_mm and increases the 2C-specific gene expression in mESCs [187]. Mechanistically, ZSCAN4c recruits BAF-containing chromatin remodeling complex, GBAF, via its SCAN domain, to MT2\_mm sites [187]. It has been proposed that ZSCAN4c-GBAF complex could be responsible for the deposition of activating histone marks (H3K27ac, H3K4me1, and H3K14ac) on MT2\_mm, highlighting an epigenetic function of ZSCAN4c in regulating the TE expression. Given that the expression level of ZSCAN4c is significantly higher in 2-cell embryos compared to mESCs, the authors speculated that ZSCAN4c could similarly activate MT2\_mm during ZGA. Additionally, it was observed that DPPA2, DPPA4, and DUX were highly upregulated upon the ZSCAN4c overexpression in mESCs. However, unlike the DUX overexpression, the overexpression of ZSCAN4c in DPPA2/DPPA4 double KO mESCs could not promote 2C-like transitions [77]. The upregulation of DUX is interesting given that DUX is known to bind and activate the ZSCAN4 cluster in 2C-like mESCs [56]. This implies that DUX may not be the sole driver of the 2-cell program, and that 2-cell specific TFs may also function to reinforce each other's expression, thereby contributing to a positive feedback loop to activate the 2-cell gene expression program.

**4.2.5. CCCTC-Binding Factor (CTCF) Restrains 2C-Like Reprogramming.** The chromatin architecture protein, CCCTC-binding factor (CTCF), was recently discovered to be a barrier to 2C-like reprogramming [73]. CTCF is a zinc finger binding protein with roles in chromatin compaction and the insulation of topologically associated domains (TADs) [188]. Loss of CTCF promotes entry into the 2C-like state, in a ZSCAN4-dependent manner [73]. Interestingly, this study found that the upregulation of ZSCAN4 precedes that of DUX and MuERVL, placing ZSCAN4 expression upstream of DUX. Consistent with a role of CTCF in restraining 2C gene induction, the CTCF expression is lower in 2-cell embryos than in the ICM, the former characterized by a more relaxed chromatin state that is associated with weak TAD boundaries [189]. Taken together, CTCF may be a major repressor of the 2C-like state in mESCs, as well as during ZGA *in vivo*.

## 5. Conclusion

The contribution of TEs to gene regulation and chromatin dynamics is evident in the functional conservation of specific TE subclasses, even though there is limited conservation of TE sequences and activity of subclass type across species, and extensive TE polymorphisms are prevalent within species [190–192]. In particular, the convergence of retrotransposon regulation illustrates the importance of its precise expression during development. It is now clear that the reactivation of specific retrotransposons during early embryogenesis is not merely a consequence of genome-wide reprogramming, but exerts key biological functions. In fact, MuERVL activation alone, without DUX induction, is sufficient to induce the expression of a subset of 2C genes, reinforcing that the critical regulatory role MuERVL plays during early development [83]. Nonetheless, how different retrotransposons participate in sculpting the totipotency program and their mechanisms of action remains open questions.

Recent studies have highlighted key differences between retrotransposon expression and regulation in 2-cell embryos versus 2C-like mESC counterparts. For example, DUX is a key driver of mESCs to 2C-like transitions [54], but is not essential for ZGA *in vivo* [193], and its expression only activates a subset of the 2-cell program [56, 63, 164]. Moreover, DUX only occupies a quarter of the accessible chromatin in 2C-like cells, suggesting that this system may not fully recapitulate the complexity of ZGA [56]. Indeed, maternal and zygotic DUX KO mice are able to develop into adulthood, and a subset of presumably DUX-activated 2C genes in 2C-like mESCs can still be activated in DUX KO mice [78, 194]. Therefore, these findings suggest that multiple TFs that remain to be determined are likely involved in the regulation of the 2-cell transcriptional program.

These emerging studies on the dispensability of factors hypothesized to play key roles in the activation of retrotransposons and 2C genes during embryonic development are indicative of overlaps between the functional roles of retrotransposons and the pathways regulating their transcriptional activities. In this regard, it will be important to assess whether findings from 2C-like systems can be recapitulated in early embryos. It is likely that the transcriptional output of retrotransposons is dependent on the contribution of both nuclear factors and chromatin dynamics. Taken together, these observations paint a highly complex landscape of retrotransposon regulation in totipotency, delineating how specific classes of retrotransposons function and their dynamic nature of regulation will be integral in illuminating the diverse roles of retrotransposons during early development and cell fate determination.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Yixuan Low and Dennis Eng Kiat Tan contributed equally to this work.

## Acknowledgments

WT is supported by the National Research Foundation (NRF), Singapore, under the NRF fellowship (NRF-NRFF2016-06); Biomedical Research Council, Agency for Science, Technology, and Research (1531C00144); and Singapore National Medical Research Council, NMRC (OFIRG19nov-0015).

## References

- [1] E. S. Lander, L. M. Linton, B. Birren et al., “International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome,” *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.
- [2] R. N. Platt, M. W. Vandeweghe, and D. A. Ray, “Mammalian transposable elements and their impacts on genome evolution,” *Chromosom Resolutions*, vol. 26, no. 1–2, pp. 25–43, 2018.
- [3] G. Bourque, “Transposable elements in gene regulation and in the evolution of vertebrate genomes,” *Current Opinion in Genetics & Development*, vol. 19, no. 6, pp. 607–612, 2009.
- [4] A. P. J. de Koning, W. Gu, T. A. Castoe, M. A. Batzer, and D. D. Pollock, “Repetitive elements may comprise over two-thirds of the human genome,” *PLOS Genetics*, vol. 7, no. 12, article e1002384, 2011.
- [5] D. J. Finnegan, “Eukaryotic transposable elements and genome evolution,” *Trends in Genetics*, vol. 5, no. 4, pp. 103–107, 1989.
- [6] T. Wicker, F. Sabot, A. Hua-van et al., “A unified classification system for eukaryotic transposable elements,” *Nature Reviews Genetics*, vol. 8, no. 12, pp. 973–982, 2007.
- [7] T. H. Eickbush and V. K. Jamburuthugoda, “The diversity of retrotransposons and the properties of their reverse transcriptases,” *Virus Research*, vol. 134, no. 1–2, pp. 221–234, 2008.
- [8] A. Hayward, “Origin of the retroviruses: when, where, and how?,” *Current Opinion in Virology*, vol. 25, pp. 23–27, 2017.
- [9] P. J. Thompson, T. S. Macfarlan, and M. C. Lorincz, “Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire,” *Molecular Cell*, vol. 62, no. 5, pp. 766–776, 2016.
- [10] W. F. Doolittle and C. Sapienza, “Selfish genes, the phenotype paradigm and genome evolution,” *Nature*, vol. 284, no. 5757, pp. 601–603, 1980.
- [11] L. E. Orgel and F. H. C. Crick, “Selfish DNA: the ultimate parasite,” *Nature*, vol. 284, no. 5757, pp. 604–607, 1980.
- [12] P. A. Callinan and M. A. Batzer, “Retrotransposable Elements and Human Disease,” in *Genome and Disease*, pp. 104–115, 2006.
- [13] K. Wimmer, T. Callens, A. Wernstedt, and L. Messiaen, “The NF1 gene contains hotspots for L1 endonuclease-dependent De Novo insertion,” *PLoS Genetics*, vol. 7, no. 11, article e1002371, 2011.
- [14] M. Taşkesen, G. B. Collin, A. V. Evsikov et al., “Novel Alu retrotransposon insertion leading to Alström syndrome,” *Human Genetics*, vol. 131, no. 3, pp. 407–413, 2012.
- [15] P. M. Green, R. D. Bagnall, N. H. Waseem, and F. Giannelli, “Haemophilia a mutations in the UK: results of screening one-third of the population,” *British Journal of Haematology*, vol. 143, no. 1, pp. 115–128, 2008.
- [16] G. N. Gallus, E. Cardaioli, A. Rufa et al., “Alu-element insertion in an OPA1 intron sequence associated with autosomal dominant optic atrophy,” *Molecular Vision*, vol. 16, pp. 178–183, 2010.
- [17] B. McClintock, “The origin and behavior of mutable loci in maize,” *Proceedings of the National Academy of Sciences*, vol. 36, no. 6, pp. 344–355, 1950.
- [18] E. H. Davidson and R. Britten, “Regulation of gene expression: possible role of repetitive sequences,” *Science*, vol. 204, no. 4397, pp. 1052–1059, 1979.
- [19] R. J. Britten and E. H. Davidson, “Gene regulation for higher cells: a theory,” *Science (80-)*, vol. 165, no. 3891, pp. 349–357, 1969.
- [20] S. D. Fugmann, “The origins of the *Rag* genes –From transposition to V(D)J recombination,” *Seminars in Immunology*, vol. 22, no. 1, pp. 10–16, 2010.
- [21] A. Dupressoir, C. Vernochet, O. Bawa et al., “Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 29, pp. 12127–12132, 2009.
- [22] A. R. Muotri, V. T. Chu, M. C. N. Marchetto, W. Deng, J. V. Moran, and F. H. Gage, “Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition,” *Nature*, vol. 435, no. 7044, pp. 903–910, 2005.
- [23] D. Schmidt, P. C. Schwalie, M. D. Wilson et al., “Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages,” *Cell*, vol. 148, no. 1–2, pp. 335–348, 2012.
- [24] M. Xie, C. Hong, B. Zhang et al., “DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape,” *Nature Genetics*, vol. 45, no. 7, pp. 836–841, 2013.
- [25] P. É. Jacques, J. Jeyakani, and G. Bourque, “The majority of primate-specific regulatory sequences are derived from transposable elements,” *PLOS Genetics*, vol. 9, no. 5, article e1003504, 2013.
- [26] M. Percharde, C. J. Lin, Y. Yin et al., “A LINE1-nucleolin partnership regulates early development and ESC identity,” *Cell*, vol. 174, no. 2, pp. 391–405.e19, 2018.
- [27] J. Durruthy-Durruthy, V. Sebastiano, M. Wossidlo et al., “The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming,” *Nature Genetics*, vol. 48, no. 1, pp. 44–52, 2016.
- [28] E. B. Chuong, N. C. Elde, and C. Feschotte, “Regulatory evolution of innate immunity through co-option of endogenous retroviruses,” *Science (80-)*, vol. 351, no. 6277, pp. 1083–1087, 2016.
- [29] A. E. V. Hof, P. Campagne, D. J. Rigden et al., “The industrial melanism mutation in British peppered moths is a transposable element,” *Nature*, vol. 534, no. 7605, pp. 102–105, 2016.
- [30] M. Ong-Abdullah, J. M. Ordway, N. Jiang et al., “Loss of *Karma* transposon methylation underlies the mantled somaclonal variant of oil palm,” *Nature*, vol. 525, no. 7570, pp. 533–537, 2015.
- [31] D. Rodriguez-Terrones and M. E. Torres-Padilla, “Nimble and ready to mingle: transposon outbursts of early development,” *Trends in Genetics*, vol. 34, no. 10, pp. 806–820, 2018.

- [32] J. Göke, X. Lu, Y. S. Chan et al., “Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells,” *Cell Stem Cell*, vol. 16, no. 2, pp. 135–141, 2015.
- [33] A. Fadloun, S. le Gras, B. Jost et al., “Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA,” *Nature Structural & Molecular Biology*, vol. 20, no. 3, pp. 332–338, 2013.
- [34] A. E. Peaston, A. V. Evsikov, J. H. Graber et al., “Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos,” *Developmental Cell*, vol. 7, no. 4, pp. 597–606, 2004.
- [35] S. R. Richardson, P. Gerdes, D. J. Gerhardt et al., “Heritable L1 retrotransposition in the mouse primordial germline and early embryo,” *Genome Research*, vol. 27, no. 8, pp. 1395–1405, 2017.
- [36] H. Kano, I. Godoy, C. Courtney et al., “L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism,” *Genes & Development*, vol. 23, no. 11, pp. 1303–1312, 2009.
- [37] S. J. Newkirk, S. Lee, F. C. Grandi et al., “Intact piRNA pathway prevents L1 mobilization in male meiosis,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 28, pp. E5635–E5644, 2017.
- [38] P. Braude, V. Bolton, and S. Moore, “Human gene expression first occurs between the four- and eight-cell stages of preimplantation development,” *Nature*, vol. 332, no. 6163, pp. 459–461, 1988.
- [39] K. E. Latham, J. I. Garrels, C. Chang, and D. Solter, “Quantitative analysis of protein synthesis in mouse embryos. I. Extensive reprogramming at the one- and two-cell stages,” *Development*, vol. 112, no. 4, pp. 921–932, 1991.
- [40] F. Aoki, D. M. Worrada, and R. M. Schultz, “Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo,” *Developmental Biology*, vol. 181, no. 2, pp. 296–307, 1997.
- [41] Z. D. Smith, M. M. Chan, T. S. Mikkelsen et al., “A unique regulatory phase of DNA methylation in the early mammalian embryo,” *Nature*, vol. 484, no. 7394, pp. 339–344, 2012.
- [42] A. Bošković, A. Eid, J. Pontabry et al., “Higher chromatin mobility supports totipotency and precedes pluripotency in vivo,” *Genes & Development*, vol. 28, no. 10, pp. 1042–1047, 2014.
- [43] W. Xia and W. Xie, “Rebooting the epigenomes during mammalian early embryogenesis,” *Stem Cell Reports*, vol. 15, no. 6, pp. 1158–1175, 2020.
- [44] M. A. Eckersley-Maslin, C. Alda-Catalinas, and W. Reik, “Dynamics of the epigenetic landscape during the maternal-to-zygotic transition,” *Nature Reviews Molecular Cell Biology*, vol. 19, no. 7, pp. 436–450, 2018.
- [45] J. Wu, B. Huang, H. Chen et al., “The landscape of accessible chromatin in mammalian preimplantation embryos,” *Nature*, vol. 534, no. 7609, pp. 652–657, 2016.
- [46] L. Gao, K. Wu, Z. Liu et al., “Chromatin accessibility landscape in human early embryos and its association with evolution,” *Cell*, vol. 173, no. 1, pp. 248–259.e15, 2018.
- [47] T. S. Macfarlan, W. D. Gifford, S. Agarwal et al., “Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A,” *Genes & Development*, vol. 25, no. 6, pp. 594–607, 2011.
- [48] E. J. Grow, R. A. Flynn, S. L. Chavez et al., “Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells,” *Nature*, vol. 522, no. 7555, pp. 221–225, 2015.
- [49] A. Y. Chong, K. K. Kojima, J. Jurka et al., “Evolution and gene capture in ancient endogenous retroviruses - insights from the crocodylian genomes,” *Retrovirology*, vol. 11, no. 1, pp. 1–15, 2014.
- [50] W. E. Johnson, “Endogenous Retroviruses in the Genomics Era,” *Annual Review of Virology*, vol. 2, no. 1, pp. 135–159, 2015.
- [51] D. Kigami, N. Minami, H. Takayama, and H. Imai, “MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos,” *Biology of Reproduction*, vol. 68, no. 2, pp. 651–654, 2003.
- [52] P. Svoboda, P. Stein, M. Anger, E. Bernstein, G. J. Hannon, and R. M. Schultz, “RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos,” *Developmental Biology*, vol. 269, no. 1, pp. 276–285, 2004.
- [53] A. Inoue, S. Matoba, and Y. Zhang, “Transcriptional activation of transposable elements in mouse zygotes is independent of Tet3-mediated 5-methylcytosine oxidation,” *Cell Research*, vol. 22, no. 12, pp. 1640–1649, 2012.
- [54] T. S. Macfarlan, W. D. Gifford, S. Driscoll et al., “Embryonic stem cell potency fluctuates with endogenous retrovirus activity,” *Nature*, vol. 487, no. 7405, pp. 57–63, 2012.
- [55] M. Genet and M. E. Torres-Padilla, “The molecular and cellular features of 2-cell-like cells: a reference guide,” *Development*, vol. 147, no. 16, 2020.
- [56] P. G. Hendrickson, J. A. Doráis, E. J. Grow et al., “Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERV1/HERV1 retrotransposons,” *Nature Genetics*, vol. 49, no. 6, pp. 925–934, 2017.
- [57] T. Amano, T. Hirata, G. Falco et al., “Zscan4 restores the developmental potency of embryonic stem cells,” *Nature Communications*, vol. 4, no. 1, article 1966, 2013.
- [58] G. Falco, S. L. Lee, I. Stanghellini, U. C. Bassey, T. Hamatani, and M. S. H. Ko, “Zscan4: A novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells,” *Developmental Biology*, vol. 307, no. 2, pp. 539–550, 2007.
- [59] R. Srinivasan, N. Nady, N. Arora et al., “Zscan 4 binds nucleosomal microsatellite DNA and protects mouse two-cell embryos from DNA damage,” *Science Advances*, vol. 6, no. 12, article eaaz9115, 2020.
- [60] M. Ooga, H. Fulka, S. Hashimoto, M. G. Suzuki, and F. Aoki, “Analysis of chromatin structure in mouse preimplantation embryos by fluorescent recovery after photobleaching,” *Epigenetics*, vol. 11, no. 1, pp. 85–94, 2016.
- [61] S. Atashpaz, S. Samadi Shams, J. M. Gonzalez et al., “ATR expands embryonic stem cell fate potential in response to replication stress,” *Elife*, vol. 9, pp. 1–30, 2020.
- [62] J. W. Jachowicz, X. Bing, J. Pontabry, A. Bošković, O. J. Rando, and M. E. Torres-Padilla, “LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo,” *Nature Genetics*, vol. 49, no. 10, pp. 1502–1510, 2017.
- [63] A. De Iaco, E. Planet, A. Coluccio, S. Verp, J. Duc, and D. Trono, “DUX-family transcription factors regulate zygotic genome activation in placental mammals,” *Nature Genetics*, vol. 49, no. 6, pp. 941–945, 2017.
- [64] E. J. Grow, B. D. Weaver, C. M. Smith et al., “p53 convergently activates *\_Dux\_ / \_DUX4\_* in embryonic stem cells

- and in facioscapulohumeral muscular dystrophy cell models,” *Nature Genetics*, vol. 53, no. 8, pp. 1207–1220, 2021.
- [65] Z. Hu, D. E. K. Tan, G. Chia et al., “Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells,” *Nature Cell Biology*, vol. 22, no. 2, pp. 175–186, 2020.
- [66] T. Ishiuchi, R. Enriquez-Gasca, E. Mizutani et al., “Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly,” *Nature Structural & Molecular Biology*, vol. 22, no. 9, pp. 662–671, 2015.
- [67] D. Rodriguez-Terrones, X. Gaume, T. Ishiuchi et al., “A molecular roadmap for the emergence of early-embryonic-like cells in culture,” *Nature Genetics*, vol. 50, no. 1, pp. 106–119, 2018.
- [68] D. Rodriguez-Terrones, G. Hartleben, X. Gaume et al., “A distinct metabolic state arises during the emergence of 2-cell-like cells,” *EMBO reports*, vol. 21, no. 1, pp. 1–14, 2020.
- [69] S. S. C. Hung, R. C. B. Wong, A. A. Sharov, Y. Nakatake, H. Yu, and M. S. H. Ko, “Repression of global protein synthesis by Eif1a-Like genes that are expressed specifically in the two-cell embryos and the transient Zscan4-Positive state of embryonic stem cells,” *DNA Research*, vol. 20, no. 4, pp. 391–402, 2013.
- [70] Z. Huang, J. Yu, W. Cui, B. K. Johnson, K. Kim, and G. P. Pfeifer, “The chromosomal protein SMCHD1 regulates DNA methylation and the 2C-like state of embryonic stem cells by antagonizing TET proteins,” *Science Advances*, vol. 7, no. 4, pp. 1–16, 2021.
- [71] F. Yang, X. Huang, R. Zang et al., “DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces a Totipotent 2C-like State,” *Cell Stem Cell*, vol. 26, no. 2, pp. 234–250.e7, 2020.
- [72] J. Liu, M. Gao, J. He et al., “The RNA m<sup>6</sup>A reader YTHDC1 silences retrotransposons and guards ES cell identity,” *Nature*, vol. 591, no. 7849, pp. 322–326, 2021.
- [73] T. Olbrich, M. Vega-Sendino, D. Tillo et al., “CTCF is a barrier for 2C-like reprogramming,” *Nature Communications*, vol. 12, no. 1, 2021.
- [74] K. Hisada, C. Sánchez, T. A. Endo et al., “RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells,” *Molecular and Cellular Biology*, vol. 32, no. 6, pp. 1139–1149, 2012.
- [75] J. Dan, M. Li, J. Yang et al., “Roles for *Tbx3* in regulation of two-cell state and telomere elongation in mouse ES cells,” *Scientific Reports*, vol. 3, no. 1, pp. 1–9, 2013.
- [76] K. M. Schüle, M. Leichsenring, T. Andreani et al., “GADD45 promotes locus-specific DNA demethylation and 2C cycling in embryonic stem cells,” *Genes & Development*, vol. 33, no. 13–14, pp. 782–798, 2019.
- [77] M. Eckersley-Maslin, C. Alda-Catalinas, M. Blotenburg, E. Kreibich, C. Krueger, and W. Reik, “Dppa 2 and Dppa 4 directly regulate the Dux-driven zygotic transcriptional program,” *Genes & Development*, vol. 33, no. 3–4, pp. 194–208, 2019.
- [78] A. De Iaco, A. Coudray, J. Duc, and D. Trono, “DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells,” *EMBO reports*, vol. 20, no. 5, pp. 1–10, 2019.
- [79] Y. L. Yan, C. Zhang, J. Hao et al., “DPPA2/4 and SUMO E3 ligase PIAS4 oppositely regulate zygotic transcriptional program,” *PLoS Biology*, vol. 17, no. 6, pp. 1–31, 2019.
- [80] Y. J. Choi, C. P. Lin, D. Risso et al., “Deficiency of microRNAmiR-34a expands cell fate potential in pluripotent stem cells,” *Science*, vol. 355, no. 6325, p. eaag1927, 2017.
- [81] M. A. Eckersley-Maslin, V. Svensson, C. Krueger et al., “MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs,” *Cell Reports*, vol. 17, no. 1, pp. 179–192, 2016.
- [82] X. Fu, X. Wu, M. N. Djekidel, and Y. Zhang, “Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells,” *Nature Cell Biology*, vol. 21, no. 7, pp. 835–844, 2019.
- [83] D. Guallar, X. Bi, J. A. Pardavila et al., “RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells,” *Nature Genetics*, vol. 50, no. 3, pp. 443–451, 2018.
- [84] Y. Yang, B. Liu, J. Xu et al., “Derivation of pluripotent stem cells with in vivo embryonic and extraembryonic potency,” *Cell*, vol. 169, no. 2, pp. 243–257.e25, 2017.
- [85] J. Yang, D. J. Ryan, W. Wang et al., “Establishment of mouse expanded potential stem cells,” *Nature*, vol. 550, no. 7676, pp. 393–397, 2017.
- [86] S. Bao, W. W. Tang, B. Wu et al., “Derivation of hypermethylated pluripotent embryonic stem cells with high potency,” *Cell Research*, vol. 28, no. 1, pp. 22–34, 2018.
- [87] H. Shen, M. Yang, S. Li et al., “Mouse totipotent stem cells captured and maintained through spliceosomal repression,” *Cell*, vol. 184, no. 11, pp. 2843–2859.e20, 2021.
- [88] E. B. Chuong, M. A. K. Rumi, M. J. Soares, and J. C. Baker, “Endogenous retroviruses function as species-specific enhancer elements in the placenta,” *Nature Genetics*, vol. 45, no. 3, pp. 325–329, 2013.
- [89] G. J. Faulkner, Y. Kimura, C. O. Daub et al., “The regulated retrotransposon transcriptome of mammalian cells,” *Nature Genetics*, vol. 41, no. 5, pp. 563–571, 2009.
- [90] R. Rebollo, M. T. Romanish, and D. L. Mager, “Transposable elements: an abundant and natural source of regulatory sequences for host genes,” *Annual Review of Genetics*, vol. 46, no. 1, pp. 21–42, 2012.
- [91] The FANTOM Consortium, A. Fort, K. Hashimoto et al., “Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance,” *Nature Genetics*, vol. 46, no. 6, pp. 558–566, 2014.
- [92] Y. Huang, J. K. Kim, D. V. Do et al., “Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition,” *Elife*, vol. 6, no. e22345, pp. 1–23, 2017.
- [93] R. Beraldi, C. Pittoggi, I. Sciamanna, E. Mattei, and C. Spadafora, “Expression of LINE-1 retroposons is essential for murine preimplantation development,” *Molecular Reproduction and Development*, vol. 73, no. 3, pp. 279–287, 2006.
- [94] K. Kruse, N. Díaz, R. Enriquez-Gasca, X. Gaume, M. E. Torres-Padilla, and J. M. Vaquerizas, *Transposable elements drive reorganisation of 3D chromatin during early embryogenesis*, bioRxiv, 2019.
- [95] M. A. Surani, K. Hayashi, and P. Hajkova, “Genetic and Epigenetic Regulators of Pluripotency,” *Cell*, vol. 128, no. 4, pp. 747–762, 2007.
- [96] A. Bulut-Karslioglu, I. A. de la Rosa-Velázquez, F. Ramirez et al., “Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse

- Embryonic Stem Cells,” *Molecular Cell*, vol. 55, no. 2, pp. 277–290, 2014.
- [97] D. Pezic, S. A. Manakov, R. Sachidanandam, and A. A. Aravin, “piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells,” *Genes & Development*, vol. 28, no. 13, pp. 1410–1428, 2014.
- [98] I. A. Maksakova, P. J. Thompson, P. Goyal et al., “Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells,” *Epigenetics & Chromatin*, vol. 6, no. 1, pp. 1–16, 2013.
- [99] K. Wu, H. Liu, Y. Wang et al., “SETDB1-mediated cell fate transition between 2C-like and pluripotent states,” *Cell Reports*, vol. 30, no. 1, pp. 25–36.e6, 2020.
- [100] H. M. Rowe, A. Kapopoulou, A. Corsinotti et al., “TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells,” *Genome Research*, vol. 23, no. 3, pp. 452–461, 2013.
- [101] H. M. Rowe, J. Jakobsson, D. Mesnard et al., “KAP1 controls endogenous retroviruses in embryonic stem cells,” *Nature*, vol. 463, no. 7278, pp. 237–240, 2010.
- [102] T. Matsui, D. Leung, H. Miyashita et al., “Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET,” *Nature*, vol. 464, no. 7290, pp. 927–931, 2010.
- [103] J. Kim, H. Zhao, J. Dan et al., “Maternal Setdb 1 is required for meiotic progression and preimplantation development in mouse,” *PLOS Genetics*, vol. 12, no. 4, pp. 1–26, 2016.
- [104] C. Wang, X. Liu, Y. Gao et al., “Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development,” *Nature Cell Biology*, vol. 20, no. 5, pp. 620–631, 2018.
- [105] A. C. Groner, S. Meylan, A. Ciuffi et al., “KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading,” *PLoS Genetics*, vol. 6, no. 3, pp. 1–14, 2010.
- [106] D. C. Schultz, K. Ayyanathan, D. Negorev, G. G. Maul, and F. J. Rauscher, “SETDB1: A novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins,” *Genes & Development*, vol. 16, no. 8, pp. 919–932, 2002.
- [107] S. P. Sripathy, J. Stevens, and D. C. Schultz, “The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression,” *Molecular and Cellular Biology*, vol. 26, no. 22, pp. 8623–8638, 2006.
- [108] M. Wiznerowicz, J. Jakobsson, J. Szulc et al., “The Kruppel-associated Box Repressor Domain Can Trigger *de Novo* Promoter Methylation during Mouse Early Embryogenesis,” *Journal of Biological Chemistry*, vol. 282, no. 47, pp. 34535–34541, 2007.
- [109] D. C. Schultz, J. R. Friedman, and F. J. Rauscher, “Targeting histone deacetylase complexes via KRAB-zinc finger proteins: The PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD,” *Genes & Development*, vol. 15, no. 4, pp. 428–443, 2001.
- [110] C. Underhill, M. S. Qutob, S. P. Yee, and J. Torchia, “A Novel Nuclear Receptor Corepressor Complex, N-CoR, Contains Components of the Mammalian SWI/SNF Complex and the Corepressor KAP-1\*,” *Journal of Biological Chemistry*, vol. 275, no. 51, pp. 40463–40470, 2000.
- [111] J. Schoorlemmer, R. Pérez-Palacios, M. Climent, D. Guallar, and P. Muniesa, “Regulation of mouse retroelement MuERV-L/MERVL expression by REX1 and epigenetic control of stem cell potency,” *Frontiers in oncology*, vol. 4, pp. 1–18, 2014.
- [112] G. Hu, J. Kim, Q. Xu, Y. Leng, S. H. Orkin, and S. J. Elledge, “A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal,” *Genes & Development*, vol. 23, no. 7, pp. 837–848, 2009.
- [113] X. Li, M. Ito, F. Zhou et al., “A Maternal-Zygotic Effect Gene, *\_Zfp57\_*, Maintains Both Maternal and Paternal Imprints,” *Developmental Cell*, vol. 15, no. 4, pp. 547–557, 2008.
- [114] D. Wolf and S. P. Goff, “Embryonic stem cells use ZFP809 to silence retroviral DNAs,” *Nature*, vol. 458, no. 7242, pp. 1201–1204, 2009.
- [115] G. Wolf, P. Yang, A. C. Fuchtbauer et al., “The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses,” *Genes & Development*, vol. 29, no. 5, pp. 538–554, 2015.
- [116] F. Chen, W. Zhang, D. Xie, T. Gao, Z. Dong, and X. Lu, “Histone chaperone FACT represses retrotransposon MERVL and MERVL-derived cryptic promoters,” *Nucleic Acids Research*, vol. 48, no. 18, pp. 10211–10225, 2020.
- [117] K. Ancelin, L. Syx, M. Borensztein et al., “Maternal LSD1/KDM1A is an essential regulator of chromatin and transcription landscapes during zygotic genome activation,” *Elife*, vol. 5, no. e08851, pp. 1–24, 2016.
- [118] C. T. Foster, O. M. Dovey, L. Lezina et al., “Lysine-specific demethylase 1 regulates the embryonic transcriptome and CoREST stability,” *Molecular and Cellular Biology*, vol. 30, no. 20, pp. 4851–4863, 2010.
- [119] J. A. Law and S. E. Jacobsen, “Establishing, maintaining and modifying DNA methylation patterns in plants and animals,” *Nature Reviews Genetics*, vol. 11, no. 3, pp. 204–220, 2010.
- [120] Z. D. Smith and A. Meissner, “DNA methylation: Roles in mammalian development,” *Nature Reviews Genetics*, vol. 14, no. 3, pp. 204–220, 2013.
- [121] T. H. Bestor, “The DNA methyltransferases of mammals,” *Human Molecular Genetics*, vol. 9, no. 16, pp. 2395–2402, 2000.
- [122] M. Saitou, S. Kagiwada, and K. Kurimoto, “Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells,” *Development*, vol. 139, no. 1, pp. 15–31, 2012.
- [123] Y. Zhang, Y. Xiang, Q. Yin et al., “Dynamic epigenomic landscapes during early lineage specification in mouse embryos,” *Nature Genetics*, vol. 50, no. 1, pp. 96–105, 2018.
- [124] A. Inoue, L. Shen, Q. Dai, C. He, and Y. Zhang, “Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development,” *Cell Research*, vol. 21, no. 12, pp. 1670–1676, 2011.
- [125] T. P. Gu, F. Guo, H. Yang et al., “The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes,” *Nature*, vol. 477, no. 7366, pp. 606–610, 2011.
- [126] L. Shen, A. Inoue, J. He, Y. Liu, F. Lu, and Y. Zhang, “Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes,” *Cell Stem Cell*, vol. 15, no. 4, pp. 459–471, 2014.

- [127] C. Y. Howell, T. H. Bestor, F. Ding et al., “Genomic Imprinting Disrupted by a Maternal Effect Mutation in the *\_Dnmt1\_* Gene,” *Cell*, vol. 104, no. 6, pp. 829–838, 2001.
- [128] M. Wossidlo, T. Nakamura, K. Lepikhov et al., “5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming,” *Nature Communications*, vol. 2, no. 1, 2011.
- [129] R. Amouroux, B. Nashun, K. Shirane et al., “\_De novo\_ DNA methylation drives 5hmC accumulation in mouse zygotes,” *Nature Cell Biology*, vol. 18, no. 2, pp. 225–233, 2016.
- [130] C. P. Walsh, J. R. Chaillet, and T. H. Bestor, “Transcription of IAP endogenous retroviruses is constrained by cytosine methylation,” *Nature Genetics*, vol. 20, no. 2, pp. 116–117, 1998.
- [131] T. Dahlet, A. Argüeso Lleida, H. al Adhami et al., “Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity,” *Nature Communications*, vol. 11, no. 1, pp. 1–14, 2020.
- [132] J. Sharif, T. A. Endo, M. Nakayama et al., “Activation of Endogenous Retroviruses in *\_Dnmt1\_*<sup>-/-</sup> ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA,” *Cell Stem Cell*, vol. 19, no. 1, pp. 81–94, 2016.
- [133] M. M. Karimi, P. Goyal, I. A. Maksakova et al., “DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mescs,” *Cell Stem Cell*, vol. 8, no. 6, pp. 676–687, 2011.
- [134] M. Walter, A. Teissandier, R. Pérez-Palacios, and D. Bourc’his, “An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells,” *Elife*, vol. 5, pp. 1–30, 2016.
- [135] H. G. Leitch, K. R. McEwen, A. Turp et al., “Naive pluripotency is associated with global DNA hypomethylation,” *Nature Structural & Molecular Biology*, vol. 20, no. 3, pp. 311–316, 2013.
- [136] F. von Meyenn, M. Iurlaro, E. Habibi et al., “Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells,” *Molecular Cell*, vol. 62, no. 6, pp. 848–861, 2016.
- [137] S. B. Rothbart, K. Krajewski, N. Nady et al., “Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation,” *Nature Structural & Molecular Biology*, vol. 19, no. 11, pp. 1155–1160, 2012.
- [138] X. Liu, Q. Gao, P. Li et al., “UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemimethylated DNA and methylated H3K9,” *Nature communications*, vol. 4, no. 1, pp. 1–13, 2013.
- [139] S. Maenohara, M. Unoki, H. Toh et al., “Role of UHRF1 in de novo DNA methylation in oocytes and maintenance methylation in preimplantation embryos,” *PLOS Genetics*, vol. 13, no. 10, pp. 1–20, 2017.
- [140] J. Kang, M. Lienhard, W. A. Pastor et al., “Simultaneous deletion of the methylcytosine oxidases Tet 1 and Tet 3 increases transcriptome variability in early embryogenesis,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 31, pp. E4236–E4245, 2015.
- [141] G. Barreto, A. Schäfer, J. Marhold et al., “\_Gadd45a\_ promotes epigenetic gene activation by repair-mediated DNA demethylation,” *Nature*, vol. 445, no. 7128, pp. 671–675, 2007.
- [142] L. de la Rica, Ö. Deniz, K. C. L. Cheng et al., “TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells,” *Genome Biology*, vol. 17, no. 1, pp. 1–14, 2016.
- [143] F. Jafarpour, S. M. Hosseini, S. Ostadhosseini, H. Abbasi, A. Dalman, and M. H. Nasr-Esfahani, “Comparative dynamics of 5-methylcytosine reprogramming and TET family expression during preimplantation mammalian development in mouse and sheep,” *Theriogenology*, vol. 89, pp. 86–96, 2017.
- [144] R. Lister, M. Pelizzola, R. H. Dowen et al., “Human DNA methylomes at base resolution show widespread epigenomic differences,” *Nature*, vol. 462, no. 7271, pp. 315–322, 2009.
- [145] R. Lister, M. Pelizzola, Y. S. Kida et al., “Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells,” *Nature*, vol. 471, no. 7336, pp. 68–73, 2011.
- [146] W. Xie, C. L. Barr, A. Kim et al., “Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome,” *Cell*, vol. 148, no. 4, pp. 816–831, 2012.
- [147] R. Lister, E. A. Mukamel, J. R. Nery et al., “Global epigenomic reconfiguration during mammalian brain development,” *Science*, vol. 341, no. 6146, 2013.
- [148] K. Shirane, H. Toh, H. Kobayashi et al., “Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of Non-CpG methylation and role of DNA methyltransferases,” *PLoS Genetics*, vol. 9, no. 4, article e1003439, 2013.
- [149] H. Guo, P. Zhu, L. Yan et al., “The DNA methylation landscape of human early embryos,” *Nature*, vol. 511, no. 7511, pp. 606–610, 2014.
- [150] F. Guo, L. Yan, H. Guo et al., “The transcriptome and DNA methylome landscapes of human primordial germ cells,” *Cell*, vol. 161, no. 6, pp. 1437–1452, 2015.
- [151] S. Seisenberger, S. Andrews, F. Krueger et al., “The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells,” *Molecular Cell*, vol. 48, no. 6, pp. 849–862, 2012.
- [152] M. D. Schultz, Y. He, J. W. Whitaker et al., “Human body epigenome maps reveal noncanonical DNA methylation variation,” *Nature*, vol. 523, no. 7559, pp. 212–216, 2015.
- [153] S. Tomizawa, H. Kobayashi, T. Watanabe et al., “Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes,” *Development*, vol. 138, no. 5, pp. 811–820, 2011.
- [154] T. Ichiyanagi, K. Ichiyanagi, M. Miyake, and H. Sasaki, “Accumulation and loss of asymmetric non-CpG methylation during male germ-cell development,” *Nucleic Acids Research*, vol. 41, no. 2, pp. 738–745, 2013.
- [155] W. Guo, W. Y. Chung, M. Qian, M. Pellegrini, and M. Q. Zhang, “Characterizing the strand-specific distribution of non-CpG methylation in human pluripotent cells,” *Nucleic Acids Research*, vol. 42, no. 5, pp. 3009–3016, 2014.
- [156] W. Guo, M. Q. Zhang, and H. Wu, “Mammalian non-CG methylations are conserved and cell-type specific and may have been involved in the evolution of transposon elements,” *Scientific Reports*, vol. 6, no. 1, 2016.
- [157] G. Vlachogiannis, C. E. Niederhuth, S. Tuna et al., “The Dnmt3L ADD Domain Controls Cytosine Methylation Establishment during Spermatogenesis,” *Cell Reports*, vol. 10, no. 6, pp. 944–956, 2015.
- [158] D. Bourc’his and T. H. Bestor, “Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L,” *Nature*, vol. 431, no. 7004, pp. 96–99, 2004.

- [159] M. J. Ziller, F. Müller, J. Liao et al., “Genomic distribution and inter-sample variation of Non-CpG methylation across human cell types,” *PLoS Genetics*, vol. 7, no. 12, 2011.
- [160] P. Li, L. Wang, B. D. Bennett et al., “Rif 1 promotes a repressive chromatin state to safeguard against endogenous retrovirus activation,” *Nucleic Acids Research*, vol. 45, no. 22, pp. 12723–12738, 2017.
- [161] B. X. Yang, C. A. el Farran, H. C. Guo et al., “Systematic identification of factors for provirus silencing in embryonic stem cells,” *Cell*, vol. 163, no. 1, pp. 230–245, 2015.
- [162] A. Kapusta, Z. Kronenberg, V. J. Lynch et al., “Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs,” *PLoS Genetics*, vol. 9, no. 4, 2013.
- [163] C. Chen, W. Liu, J. Guo et al., “Nuclear m6A reader YTHDC1 regulates the scaffold function of LINE1 RNA in mouse ESCs and early embryos,” *Protein Cell*, vol. 12, no. 6, pp. 455–474, 2021.
- [164] J. L. Whiddon, A. T. Langford, C. J. Wong, J. W. Zhong, and S. J. Tapscott, “Conservation and innovation in the DUX4-family gene network,” *Nature Genetics*, vol. 49, no. 6, pp. 935–940, 2017.
- [165] K. Sugie, S. Funaya, M. Kawamura, T. Nakamura, M. G. Suzuki, and F. Aoki, “Expression of *\_Dux\_* family genes in early preimplantation embryos,” *Scientific Reports*, vol. 10, no. 1, pp. 1–10, 2020.
- [166] W. Zeng, J. C. de Greef, Y. Y. Chen et al., “Specific loss of histone H3 lysine 9 trimethylation and HP1 $\gamma$ /cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD),” *PLoS Genetics*, vol. 5, no. 7, pp. 1–14, 2009.
- [167] E. B. Chuong, N. C. Elde, and C. Feschotte, “Regulatory activities of transposable elements: From conflicts to benefits,” *Nature Reviews Genetics*, vol. 18, no. 2, pp. 71–86, 2017.
- [168] C. Alda-Catalinas, D. Bredikhin, I. Hernando-Herraez et al., “A single-cell transcriptomics CRISPR-activation screen identifies epigenetic regulators of the zygotic genome activation program,” *Cell Systems*, vol. 11, no. 1, pp. 25–41.e9, 2020.
- [169] K. H. Gretarsson and J. A. Hackett, “*\_Dppa2\_* and *\_Dppa4\_* counteract de novo methylation to establish a permissive epigenome for development,” *Nature Structural & Molecular Biology*, vol. 27, no. 8, pp. 706–716, 2020.
- [170] I. Theurillat, I. A. Hendriks, J. C. Cossec, A. Andrieux, M. L. Nielsen, and A. Dejean, “Extensive SUMO modification of repressive chromatin factors distinguishes pluripotent from somatic cells,” *Cell Reports*, vol. 32, no. 11, article 108146, 2020.
- [171] J. C. Cossec, I. Theurillat, C. Chica et al., “SUMO safeguards somatic and pluripotent cell identities by enforcing distinct chromatin states,” *Cell Stem Cell*, vol. 23, no. 5, pp. 742–757.e8, 2018.
- [172] B. Payer, M. Saitou, S. C. Barton et al., “*\_stella\_* Is a Maternal Effect Gene Required for Normal Early Development in Mice,” *Current Biology*, vol. 13, no. 23, pp. 2110–2117, 2003.
- [173] Y. Li, Z. Zhang, J. Chen et al., “Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1,” *Nature*, vol. 564, no. 7734, pp. 136–140, 2018.
- [174] T. Nakamura, Y. J. Liu, H. Nakashima et al., “PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos,” *Nature*, vol. 486, no. 7403, pp. 415–419, 2012.
- [175] T. Nakamura, Y. Arai, H. Umehara et al., “PGC7/Stella protects against DNA demethylation in early embryogenesis,” *Nature Cell Biology*, vol. 9, no. 1, pp. 64–71, 2007.
- [176] M. M. Denomme and M. R. W. Mann, “Maternal control of genomic imprint maintenance,” *Reproductive BioMedicine Online*, vol. 27, no. 6, pp. 629–636, 2013.
- [177] Y. I. Tsukada, T. Akiyama, and K. I. Nakayama, “Maternal TET3 is dispensable for embryonic development but is required for neonatal growth,” *Scientific Reports*, vol. 5, no. 1, pp. 1–13, 2015.
- [178] L. Daxinger, S. J. Tapscott, and S. M. van der Maarel, “Genetic and epigenetic contributors to FSHD,” *Current Opinion in Genetics & Development*, vol. 33, pp. 56–61, 2015.
- [179] M. L. Ruebel, K. A. Vincent, P. Z. Schall, K. Wang, and K. E. Latham, “SMCHD1 terminates the first embryonic genome activation event in mouse two-cell embryos and contributes to a transcriptionally repressive state,” *American Journal of Physiology-Cell Physiology*, vol. 317, no. 4, pp. C655–C664, 2019.
- [180] M. M. Wong, M. D. Belew, A. Kwieraga, J. D. Nhan, and W. M. Michael, “Programmed DNA Breaks Activate the Germline Genome in *\_Caenorhabditis elegans\_*,” *Developmental Cell*, vol. 46, no. 3, pp. 302–315.e5, 2018.
- [181] C. Ziegler-Birling, A. Helmrich, L. Tora, and M. E. Torres-Padilla, “Distribution of p 53 binding protein 1 (53BP1) and phosphorylated H2A.X during mouse preimplantation development in the absence of DNA damage,” *The International Journal of Developmental Biology*, vol. 53, no. 7, pp. 1003–1011, 2009.
- [182] L. Liu, S. M. Bailey, M. Okuka et al., “Telomere lengthening early in development,” *Nature Cell Biology*, vol. 9, no. 12, pp. 1436–1441, 2007.
- [183] M. Wossidlo, J. Arand, V. Sebastiano et al., “Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes,” *The EMBO Journal*, vol. 29, no. 11, pp. 1877–1888, 2010.
- [184] M. Zalzman, G. Falco, L. V. Sharova et al., “*\_Zscan4\_* regulates telomere elongation and genomic stability in ES cells,” *Nature*, vol. 464, no. 7290, pp. 858–863, 2010.
- [185] M. S. H. Ko, J. R. Kitchen, X. Wang et al., “Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development,” *Development*, vol. 127, no. 8, pp. 1737–1749, 2000.
- [186] T. Hirata, T. Amano, Y. Nakatake et al., “*Zscan4* transiently reactivates early embryonic genes during the generation of induced pluripotent stem cells,” *Scientific Reports*, vol. 2, no. 1, pp. 1–11, 2012.
- [187] W. Zhang, F. Chen, R. Chen et al., “*Zscan4c* activates endogenous retrovirus MERV1 and cleavage embryo genes,” *Nucleic acids research*, vol. 47, no. 16, pp. 8485–8501, 2019.
- [188] J. E. Phillips and V. G. Corces, “CTCF: master weaver of the genome,” *Cell*, vol. 137, no. 7, pp. 1194–1211, 2009.
- [189] Y. Ke, Y. Xu, X. Chen et al., “3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis,” *Cell*, vol. 170, no. 2, pp. 367–381.e20, 2017.
- [190] D. A. Ray, C. Feschotte, H. J. T. Pagan et al., “Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*,” *Genome Research*, vol. 18, no. 5, pp. 717–728, 2008.
- [191] R. A. Grahn, T. A. Rinehart, M. A. Cantrell, and H. A. Wichman, “Extinction of LINE-1 activity coincident with a major

- mammalian radiation in rodents,” *Cytogenetic and Genome Research*, vol. 110, no. 1–4, pp. 407–415, 2005.
- [192] R. Belshaw, A. L. A. Dawson, J. Woolven-Allen, J. Redding, A. Burt, and M. Tristem, “Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K (HML2): implications for present-day activity,” *Journal of Virology*, vol. 79, no. 19, pp. 12507–12514, 2005.
- [193] A. De Iaco, S. Verp, S. Offner, and D. Trono, “DUX is a non-essential synchronizer of zygotic genome activation,” *Development*, vol. 147, no. 2, article dev177725, 2019.
- [194] Z. Chen and Y. Zhang, “Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development,” *Nature Genetics*, vol. 51, no. 6, pp. 947–951, 2019.
- [195] D. Guallar, R. Pérez-Palacios, M. Climent et al., “Expression of endogenous retroviruses is negatively regulated by the pluripotency marker Rex 1/Zfp 42,” *Nucleic Acids Research*, vol. 40, no. 18, pp. 8993–9007, 2012.