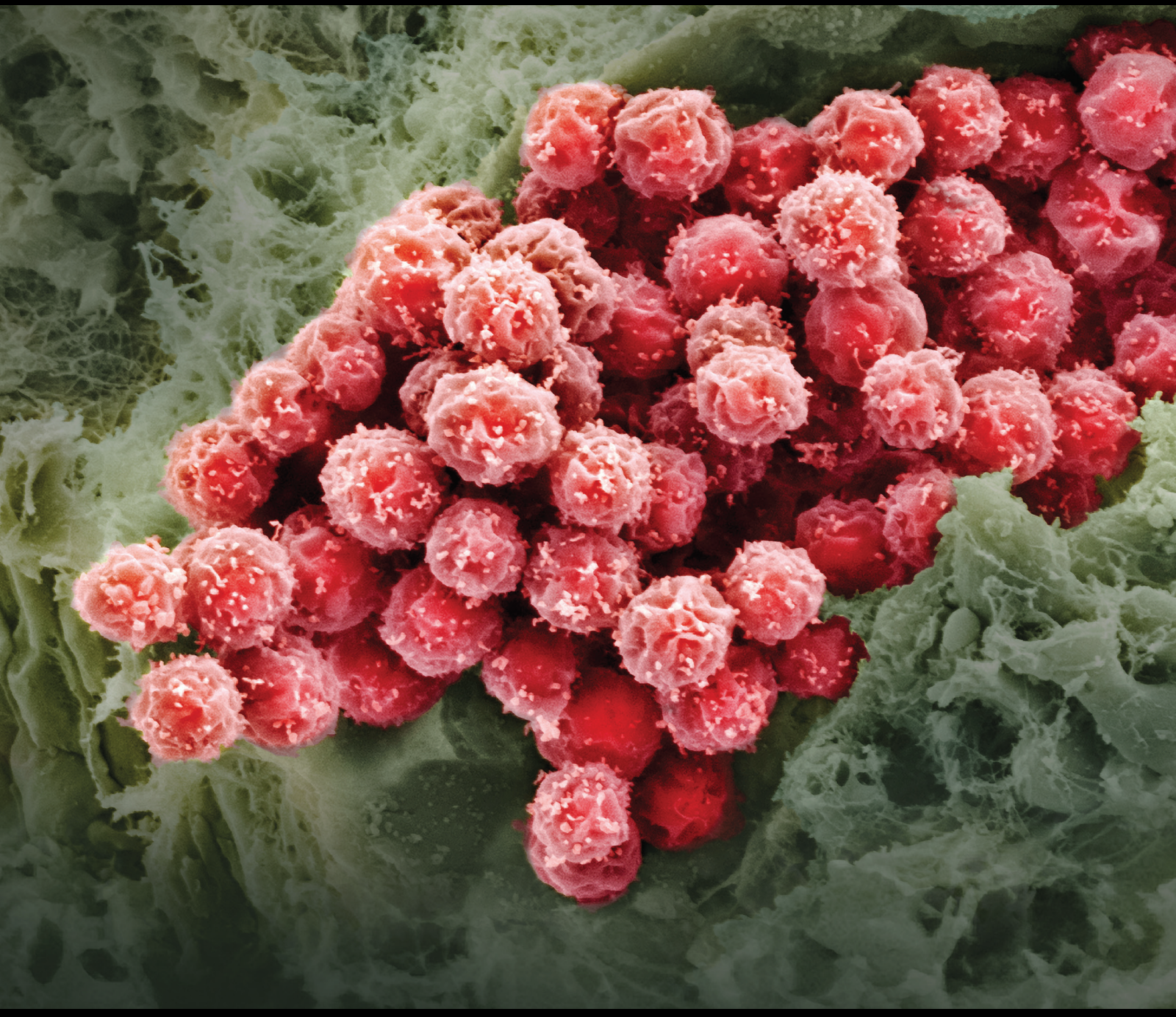
Stem Cells International

The Function and Regulation of Transposons in Stem Cells

Lead Guest Editor: Xinyi Lu
Guest Editors: Andrew Hutchins, Chengqi Lin, and Yun-Shen Chan



The Function and Regulation of Transposons in Stem Cells

Stem Cells International

The Function and Regulation of Transposons in Stem Cells

Lead Guest Editor: Xinyi Lu
Guest Editors: Andrew Hutchins, Chengqi Lin, and Yun-Shen Chan

Contents

Hindawi
Stem Cells International
Volume 2022, Article ID 8051717, 19 pages
https://doi.org/10.1155/2022/8051717

Review Article

Long Noncoding RNA Mediated Regulation in Human Embryogenesis, Pluripotency, and Reproduction

Lei Liu and Fang Fang

The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, China 230001

Correspondence should be addressed to Fang Fang; fangfang0724@gmail.com

Received 1 September 2021; Accepted 27 December 2021; Published 22 January 2022

Academic Editor: Xinyi Lu

Long noncoding RNAs (lncRNAs), a class of noncoding RNAs with more than 200 bp in length, are produced by pervasive transcription in mammalian genomes and regulate gene expression through various action mechanisms. Accumulating data indicate that lncRNAs mediate essential biological functions in human development, including early embryogenesis, induction of pluripotency, and germ cell development. Comprehensive analysis of sequencing data highlights that lncRNAs are expressed in a stage-specific and human/primate-specific pattern during early human development. They contribute to cell fate determination through interacting with almost all classes of cellular biomolecules, including proteins, DNA, mRNAs, and microRNAs. Furthermore, the expression of a few of lncRNAs is highly associated with the pathogenesis and progression of many reproductive diseases, suggesting that they could serve as candidate biomarkers for diagnosis or novel targets for treatment. Here, we review research on lncRNAs and their roles in embryogenesis, pluripotency, and reproduction. We aim to identify the underlying molecular mechanisms essential for human development and provide novel insight into the causes and treatments of human reproductive diseases.

1. Introduction

Identification and functional characterization of noncoding RNAs (ncRNAs) have revolutionized our traditional view of RNA biology, as well as developmental biology [1]. Before discovering microRNAs (miRNAs) and small interfering RNAs, mRNAs that are transcribed from the coding region of the genome and translated as proteins are considered the primary regulators of the gene expression program in the cells [2]. The vast majority of the genome that is not translated into protein is junk DNA regions [3]. With the rapid development of microarray and high-throughput sequencing technology, a comprehensive annotation of the mammalian genome demonstrates that most mammalian genome is actively transcribed into RNAs, and thousands of ncRNAs have been identified [4, 5]. ncRNAs are divided into two main types according to the length of the transcripts: small noncoding RNAs (sncRNAs), which are composed of less than 200 nucleotides, and long noncoding RNAs (lncRNAs), which consist of more than 200 nucleotides [6]. In this review, we focused on the discussion of lncRNAs. There are five different sources of lncRNAs: (1) a protein-coding gene was mutated and transformed into a noncoding RNA sequence. (2) Following chromosome rearrangement, two separate nontranscribed sequence regions are juxtaposed together to produce expressed noncoding sequences. (3) lncRNAs without a protein-coding function are produced by duplicating noncoding genes by retrotransposition. (4) Local two tandem duplication produces adjacent repeat sequences, which increases the size of lncRNs. (5) The insertion of transposable elements (TEs) can produce functional lncRNAs [7, 8].

It was questionable whether lncRNAs have putative functions in cells, as they are present in relatively low levels. It is estimated that total lncRNAs are present at two magnitudes less than total mRNAs. However, recent research suggests that lncRNAs may function at a very low level as a molecular scaffold or a catalytic molecule [9]. A growing

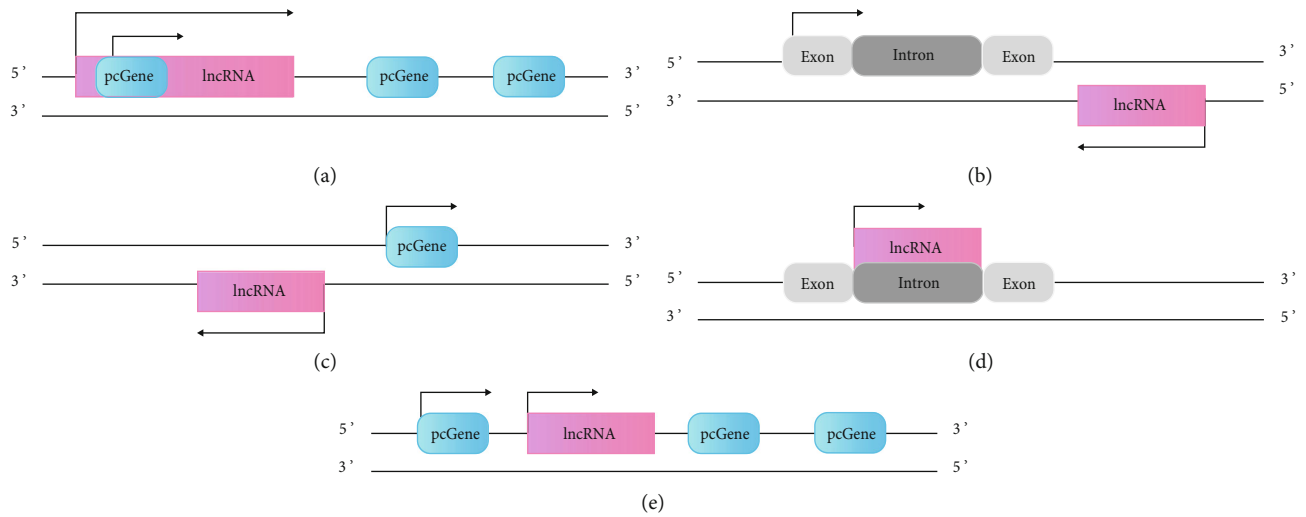


Figure 1: Schematic diagram of lncRNA classification. Classification of lncRNAs into five classes: (a) sense, (b) antisense, (c) bidirectional, (d) intronic, and (e) intergenic.

number of lncRNAs are found to play essential roles in regulating cell proliferation, survival, cell cycle, differentiation, and apoptosis [10]. They are also indicated as vital regulators in initiating and developing many diseases, including reproductive diseases [11]. X-inactive specific transcript (XIST), located on the X-chromosome of mammalian cells, is the first reported lncRNA. It has been proven to be a major regulator of the X-inactivation process [12]. Another well-established example of functional lncRNAs is H19, which is highly expressed in many tissues derived from endoderm and mesoderm. It regulates the network of imprinted genes that regulate fetal and postnatal growth [13], and it is differentially expressed in many disease tissues.

lncRNAs can be divided into five categories based on their genome localization and the direction of transcription relative to the protein-coding genes (pcGenes) in the genome: sense, antisense, bidirectional, intronic, and long intergenic (Figure 1) [14]. Sense lncRNAs are transcribed from the same strand and direction as pcGenes, and antisense lncRNAs are transcribed from the opposite strand of pcGenes. Sense and antisense lncRNAs are located within the regions of their surrounding pcGenes. Bidirectional lncRNA is located less than 1 kb from the surrounding pcGenes, sharing the same promoter as the protein-coding gene, but transcribed from the opposite direction [15]. Long intergenic noncoding RNAs (lincRNAs) are located within the intergenic regions of pcGenes, and they do not overlap with protein-coding regions.

lncRNAs could control transcription in *cis* or *trans*, regulate essential proteins or nucleic acid molecules, and are also involved in the organization of the nuclear domains [16]. The mechanisms of action vary depending on their structural conformations, biochemical properties, and specific subcellular localization [17, 18] (Figure 2). (1) They could function as signal molecules. In this case, lncRNAs respond to the environmental stimuli and then are transcribed at a specific time and space. This property makes them act as biomarkers for specific biological events. (2) They could act as decoy molecules by binding to the regulatory factors of transcription. For example, lncRNAs could bind to RNA-binding proteins, transcription factors, or chromatin modifiers to inhibit their biological activity. (3) They could function as guide molecules to direct the localization of regulatory factors. For example, lncRNAs can directly bind to protein molecules to form ribonucleoprotein complexes and mediate their precise localization to specific targets to regulate gene expression [19]. (4) lncRNAs could serve as scaffold molecules to assemble various effector molecules into macromolecules to achieve precise and specific control of biological events [19]. Finally, (5) lncRNAs could function as competing endogenous RNAs (ceRNAs) to sequester miRNAs, leading to the active transcription of their mRNA targets [20]. Several studies have shown that when TEs were embedded in lncRNAs, they may function in the processing, stability, and localization of lncRNAs. More importantly, TEs are often found to be the functional domains of lncRNAs [21]. For example, 73% of Linc-ROR sequences that have miRNA binding sites are derived from TE, and these sequences are essential for maintaining the pluripotency and self-renewal of embryonic stem cells [22]. Another example is XIST, which is important in early embryonic development and reproductive diseases [23]. XIST contains three functional repeat domains that are derived from TE. A-repeats that originated from ERVB5 TE are responsible for recruiting SPEN to silence the X chromosome; C-repeats, originating from ERVB4 TE, are required for the localization of XIST; and F-repeats, which are derived from a DNA transposon, are found to interact with JARID2 [24–28].

In mammals, development starts from the fusion of mature germ cells, sperms, and eggs, generating a totipotent zygote. Then, the zygote differentiates to form pluripotent stem cells that have the potential to give rise to an entire organism, including germ cells [29]. Thus, germ cells are the most remarkable cell type capable of reestablishing totipotency and transmitting heritable genetic and epigenetic information between generations [30]. Understanding the unique cell fate change from totipotent embryos to

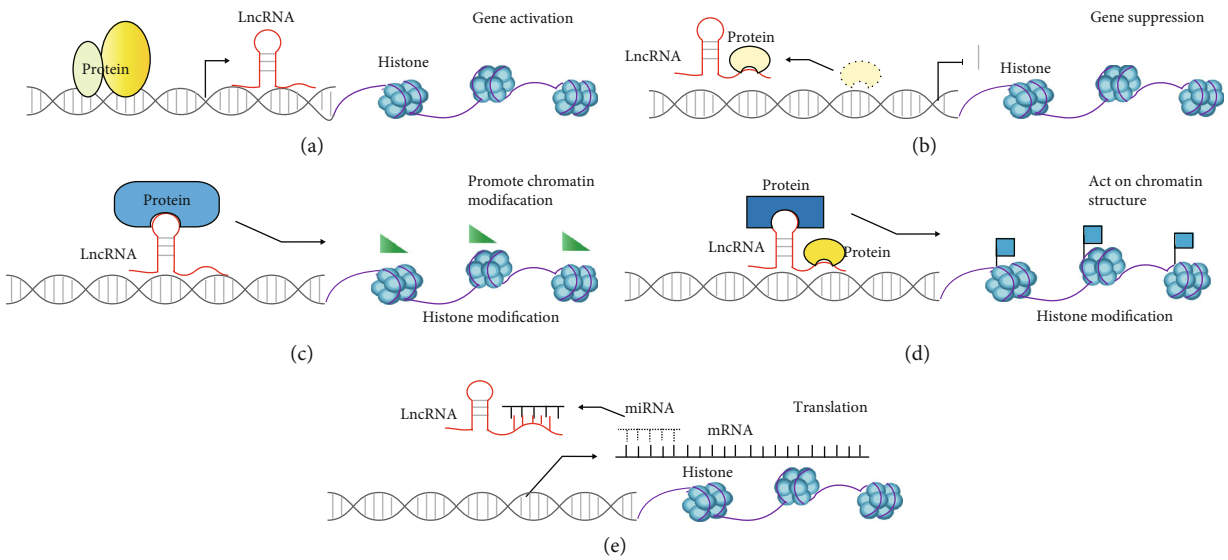


Figure 2: Schematic diagram of lncRNA mechanisms of action. Mechanisms of action: (a) signaling, (b) decoy, (c) guides, (d) scaffold, and (e) miRNA sponge.

pluripotent stem cells and germ cells will enable us to develop novel strategies for disease treatments, particularly in regenerative medicine [31]. Although substantial progress has been made to dissect the molecular mechanism underpinning this cell fate change, the role of lncRNAs remains largely unknown. In this article, we have reviewed the recent progress of lncRNAs studies in embryogenesis, pluripotency, and reproduction, aiming to shed light on future research to probe the genetic program that drives the multistep developmental processes.

2. lncRNAs in Early Human Embryonic Development

lncRNAs are present from the beginning of human embryo development. After embryonic gene activation (EGA), lncRNAs become the main category of transcripts [14]. RNA-seq and hierarchical clustering analysis demonstrated that lncRNAs show distinct developmental stage-specific expression patterns [32]. Furthermore, the epigenetic signatures of lncRNAs are similar to those of protein-coding genes, including methylation distribution at the transcription start site (TSS), methylation dynamics, and negative correlation between gene expression and promoter methylation level. Collectively, these data suggest that lncRNAs may play essential roles in early human embryonic development by regulating gene expression [33].

Human endogenous retroviruses (HERV) are remnants from ancient germline infections by exogenous retroviruses and account for 8% of the human genome [34]. HERV-derived lncRNAs are found to express at specific stages and function in human-specific or even individual-specific aspects of early human embryo development [35]. HERVK is activated by the master transcription regulator of pluripotency, OCT4, from embryonic genome activation at the eight-cell stage to human embryonic stem cell derivation. It is involved in the immunoprotective process of human embryos against exogenous viral infection [36]. Another species of HERV, HERVH, is considered the most successful endogenous retrovirus in the human genome. It is expressed during human preimplantation embryogenesis and regulates human pluripotency by providing alternative binding sites for key transcription factors, functioning as a long-range enhancer, and producing pluripotency-specific lncRNAs [37].

Human pluripotency-associated transcripts 2, 3, and 5 (HPAT2, HPAT3, and HPAT5) are derived from transposable elements (TEs) and are essential for preimplantation embryo development by modulating the acquisition of pluripotency and the formation of the inner cell mass [38].

In addition, the activity of the X chromosome is regulated by the antagonistic action of lncRNAs XIST and XACT in the early development of human embryogenesis [39].

3. lncRNAs in Pluripotent Stem Cells

Pluripotent stem cells (PSCs) cultured *in vitro* provide a unique model for studying the molecular mechanisms of human embryogenesis [40] and are considered the seed cells to differentiate into functional cells for cellular therapeutics [41]. The core regulatory network for self-renewal and pluripotency involves transcription factors, chromatin modifiers, and lncRNAs [42, 43](Figure 3). PSCs express a characteristic set of lncRNAs that interact with the other members of the core regulatory network to (1) regulate gene expression, (2) modulate signaling pathways, (3) maintain epigenetic signatures, and (4) direct differentiation.

Linc-RoR, HERVH (human endogenous retrovirus subfamily H), HPAT5, and GAS5 (growth arrest-specific transcript 5) are found to be preferentially expressed in PSCs and interact with the core regulatory transcription factor network (OCT4, NANOG, SOX2, and SALL4) to regulate the gene expression profiles and safeguard pluripotency [22, 38, 44, 45]. Mechanically, Linc-RoR works as a competing endogenous RNA to connect the network of miRNAs

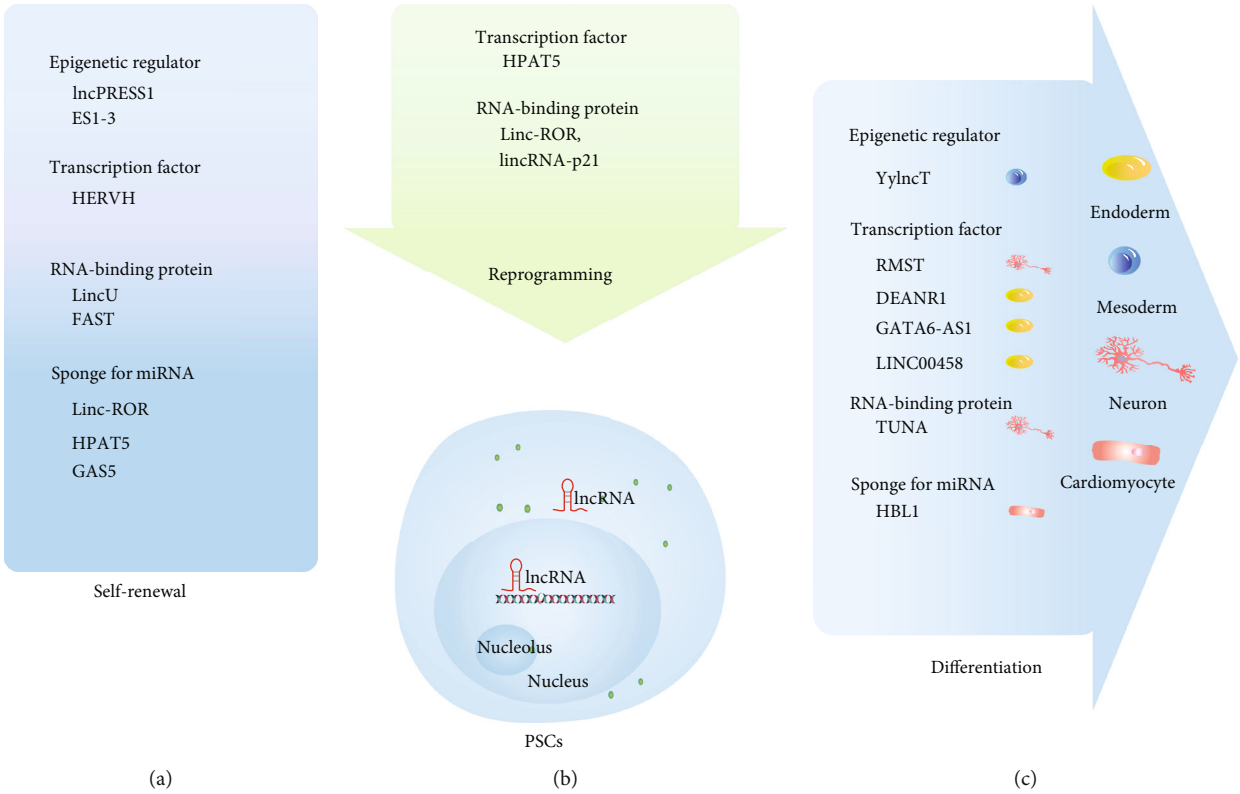


Figure 3: Mechanisms of lncRNAs in pluripotency, differentiation, and reprogramming of PSCs by interacting with different partners. Epigenetic regulator: recruit chromatin modification factors to affect chromatin status; transcription factor: binding transcription factors to regulate gene transcriptional activity; RNA-binding protein: interacting with RNA-binding protein to directly regulate protein activity; sponge for miRNA: functioning as the sponge of endogenous miRNA, preventing miRNA targets from degradation.

with core transcription factors in PSCs. Linc-ROR prevents the core transcription factors from miRNA-mediated suppression in PSCs, thus regulating the self-renewal and pluripotency of PSCs [22]. HPAT5 acts as a miRNA sponge to modulate the balance between pluripotency and differentiation by counteracting the activity of let-7 [38].

Another group of lncRNAs, such as LincU, FAST, and GAS5, maintains the pluripotency of PSCs by modulating signaling pathways that are essential for PSCs [45–47]. Mechanistically, LincU binds to DUSP9 protein, an ERK-specific phosphatase, and stabilizes its expression, thereby inhibiting the MAPK/ERK signal pathway and maintaining the naive state of ESCs [46].

Examples of lncRNAs that modulate the epigenetic status of PSCs include ES1-3 and IncPRESS1. They are shown to function as molecular scaffolds that bridge different chromatin modifiers to maintain the epigenetic signatures of PSCs. ES1-3 are highly expressed in undifferentiated hESCs. As a modular scaffold, they recruit the suppressive PRC2 component SUZ12 to silence the SOX2 neural targets in PSCs, thus maintaining pluripotency [48–50].

lncRNAs are also involved in the differentiation of PSCs into three germ layers. RMST and TUNA (Tcl1 upstream neuron-associated lincRNA) promote neuronal differentiation of human PSCs [48, 49], while DEANR1, GATA6-AS1, and LINC00458 promote endodermal lineage specification [51–53]. For example, RMST interacts with SOX2 and binds to the promoter regions of neurogenic target genes to promote neuronal differentiation [48, 49]. DEANR1, an endoderm-specific lncRNA, interacts with SMAD2/3 to activate the expression of FOXA2, thus enabling the differentiation towards endoderm [51]. In addition, HBL1, BANCR, and YyIncT are identified as critical regulators for mesoderm development [54–56].

lncRNAs are also involved in reprogramming. Linc-ROR, as a negative regulator of p53, directly binds to heterogeneous nuclear ribonucleoprotein I (hnRNP I) to inhibit the expression of p53, thereby inhibiting p53-mediated cell cycle arrest and apoptosis and promoting cell reprogramming [57]. HERVH is significantly upregulated in the reprogramming process of fibroblasts to induce pluripotent stem cells (iPSCs). By recruiting P300 and OCT4 to the HERVH LTR7 region, HERVH regulates the expression of neighboring genes, as well as pluripotency-associated transcripts. It is suggested that HERVH plays an essential role in the acquisition of somatic pluripotency [44]. lincRNA-p21 (P53-induced large intergenic noncoding RNA p21) interacts with the H3K9 methyltransferase SETDB1 and the DNA methyltransferase DNMT1 through the RNA-binding protein HNRNPK to maintain high levels of H3K9me3 modification and/or CpG methylation at the pluripotency gene promoter, thus hindering somatic cell reprogramming [58].

Knockdown of HPAT5 impairs reprogramming, indicating that it contributes directly to reprogramming and acquisition of pluripotency [38].

4. lncRNAs in Human Germ Cell Development

Germ cell development is a complex differentiation process essential for the generation of gametes, which pass on the genetic information between generations [59]. Disruption of germ cell development or misregulation of gene expression in germline-related cells leads to infertility or reproductive diseases [60]. This dynamic developmental process is precisely regulated by a tissue- or cell-specific gene network [61]. As a new regulator in gene expression networks, cell type-specific lncRNAs have recently been discovered and suggested to be involved in many cellular processes during human germ cell development [62]. Several lncRNAs show differential expression or regulatory roles in the development of human primordial germ cells (hPGCs), the first progenitor cells of the germline [63]. For example, HIPSTR (heterogeneously expressed from the Intronic Plus Strand of the TFAP2A-locus RNA) has been identified as a novel lncRNA transcribed from the TFAP2A locus and shows differential expression in human primordial germ cells [64]. In addition, XACT and XIST are expressed to regulate X-chromosome dosage in hPGCs before meiosis [65]. RNA-seq analysis of human testicular cells has identified thousands of syntenic lncRNAs associated with spermatogenesis [66–71]. The narcolepsy candidate-region 1 gene (NLC1-C), a lncRNA expressed in the cytoplasm of spermatogonia and early spermatocytes, is found to be associated with male infertility and promotes testicular embryonal carcinoma cell proliferation [71]. Single-cell RNA-seq profiling of metaphase II oocytes also found 8,700 maternal lncRNAs expressed in the preimplantation embryos [32]. Note that a large number of RNA-binding proteins are found to be critical for germ cell development across species, including VASA (DDX4) and DAZL (Deleted in Azoospermia Like) [72]. These proteins might function by influencing lncRNA action to reinforce germ cell fate.

5. lncRNAs in Reproductive Diseases

Besides the roles in development, differential expression of many lncRNAs has been identified using microarray or RNA-seq between control and reproductive disease samples [73], indicating potential roles in pathogenesis. Although most of their functions and mechanisms of action need to be further annotated and characterized, these lncRNAs could serve as potential targets for the diagnosis and treatment [74] (Table 1).

5.1. lncRNAs Associated with Male Infertility. Spermatogenesis is a complex developmental process that is essential for male fertility [75]. The process is classified into three major phases: (1) mitotic proliferation of spermatogonia, (2) the meiosis of spermatocytes, and (3) spermiogenesis and maturation of spermatocytes to spermatozoa [76]. Each phase is strictly regulated by transcriptional factors, hormones, epigenetic regulators, and lncRNAs. Disruption of any steps of spermatogenesis, referred to as maturation arrest (MA), causes male infertility [77]. Nonobstructive azoospermia (NOA) is considered the most severe case of male infertility, and it is characterized as no sperm in the ejaculate due to failure of spermatogenesis [78]. Several lncRNAs have been indicated to play roles in the process of spermatogenesis and NOA.

The *narcolepsy candidate-region 1 gene* (*NLC1-C*, *also known as LINC00162*) is expressed in spermatogonia and primary spermatocytes. Compared with fertile controls, its expression is significantly downregulated in the cytoplasm and accumulated in the nucleus in the testis of infertile MA patients. NLC1-C forms a regulatory feedback loop with miR-320a and miR-383 to control the survival and proliferation of the germ cells in the process of spermatogenesis. In the cytoplasm, NLC1-C is the target of miR-320a and miR-383; while accumulated in the nucleus of spermatogonia and primary spermatocytes, it is suggested to repress the expression of miR-320a and miR-383 by direct binding to nucleolin, resulting in the hyperactive proliferation of germ cells, which leads to male infertility [71].

GM2044 is indicated to play an essential role in NOA and specific in reproductive diseases. It is the miR-202 host gene, and its expression is significantly increased with its host gene miR202 in NOA of spermatogonial arrest. lncRNA Gm2044 inhibits the proliferation of the human testicular embryonic carcinoma cell NCCIT through the miR-202-Rbfox2 molecular signal pathway [79].

The expression of *Hox transcript antisense intergenic RNA* (*HOTAIR*) is decreased in asthenozoospermic and oligoasthenozoospermic patients [80]. The low expression of HOTAIR was also observed to be associated with specific sperm function parameters, including motility and vitality. It is found that low HOTAIR leads to downregulation of nuclear factor erythroid 2-related factor 2 (NRF2), a gene related to the expression of antioxidant genes and the quality of spermatozoa [81]. This eventually results in reactive oxygen species- (ROS-) related defects in sperm function.

lncRNA growth-arrested DNA damage-inducible gene 7 (*Gadd7*) is indicated in the regulation of the oxidative stress response and specific in reproductive diseases. Its expression is upregulated in patients with varicocele compared with fertile controls. Further functional analysis in mouse cell lines indicates that overexpression of gadd7 inhibits cell growth and promotes apoptosis by upregulating the proapoptotic regulator Bax and downregulating the antiapoptotic regulator Bcl2, resulting in male infertility [82].

5.2. lncRNAs Associated with Prostate Tumors. Prostate cancer is the most common cancer among men, and the androgen receptor (AR) plays a central role in its progression by regulating the expression of genes associated with the identity and behavior of prostate cancer cells [83]. A number of lncRNAs are identified as potential regulators for disease progression and may be applied as novel therapeutic targets.

PRNCR1 and *PCGEM1* are highly expressed in aggressive prostate cancer and bind to AR successively. They enhance the activation of ligand-dependent and ligand-

Table 1: lncRNAs and their functions in reproductive diseases.

Diseases	lncRNA	Full name	Expression level	Assessed cell line	Signaling pathways and molecules	Functions	In other diseases	References
Nonobstructive azoospermia (NOA)	NLC1-C	Narcolepsy candidate-region 1 gene	Downregulated	NCCIT, NTERA-2 (NT2), HEK293 T	Sponge for miR-320a, miR-383	Inhibits miR-320a and miR-383 transcripts by binding to nucleolin, resulting in a hyperactive proliferation of germ cells	Testicular embryonal carcinoma	[71]
	GM2044	—	Upregulated	NCCIT	miR-202-Rbfox2 pathway	Inhibits the proliferation of the human testicular embryonic carcinoma cell NCCIT	—	[79]
	HOTAIR	Hox transcript antisense intergenic RNA	Downregulated	—	NRF2	Relates to defects in sperm function	Breast cancer, lung cancer, and pancreatic cancer	[80, 81, 123]
	Gadd7	lncRNA growth-arrested DNA damage-inducible gene 7	Upregulated	GC-1, GC-2	Bax, Bcl2	Inhibits cell growth and promotes apoptosis by upregulating the proapoptotic regulator Bax and downregulating the antiapoptotic regulator Bcl2	—	[82]
Prostate tumors	PRNCR1/PCGEM1	Prostate cancer-associated noncoding RNA 1/PCGEM1 prostate-specific transcript	Upregulated	LNCaP, LNCaP-cds1, LNCaP-cds2, CWR22Rv1	AR	Promotes the proliferation of prostate cancer cells	Breast cancer and lung cancer	[84, 124]
	NEAT1	Nuclear-rich transcriptase 1	Upregulated	LNCaP and PC3, RWPE1, VCaP and DU145	Estrogen receptor alpha (ERα)	Promotes the development of prostate cancer	Non-small-cell lung cancer, breast cancer, and hepatocellular carcinoma	[85, 125]
	PCAT-1	Prostate cancer-associated transcript-1	Upregulated	LNCaP	PRC2, cMyc	Promotes the proliferation of prostate cancer cells	Colorectal cancer, hepatocellular cancer, and gastric cancer	[86, 126]
	MALAT-1	Metastasis-associated lung adenocarcinoma transcript 1	Upregulated	LNCaP-AI, 22RV1	ZEB1, ZEB2, Slug	Is associated with the increase in the Gleason score, prostate-specific antigen (PSA), and tumor stage and promotes the invasion and growth of prostate cancer cells	Glioma, hepatocellular carcinoma, and multiple myeloma	[87, 127]
	SChLAP1	Second chromosome locus associated with prostate-1	Upregulated	—	—	Relates to poor prognosis and could be used as an important biomarker to identify patients with a high risk of lethal prostate cancer	Triple negative breast cancer and bladder cancer	[88, 128, 129]

Table 1: Continued.

Diseases	lncRNA	Full name	Expression level	Assessed cell line	Signaling pathways and molecules	Functions	In other diseases	References
	GAS5	Growth arrest specific 5	Downregulated	PC3, DU145, and PNT2C2	E2F1, $P27^{Kip1}$	Induces a cell cycle arrest in the G0–G1 phase and acts as a tumor suppressor	Colorectal cancer, gastric cancer, and melanoma	[89] [89, 130]
	XIST	Inactive X chromosome-specific transcripts	Downregulated	ALST, CAOV3, OVCA3, OVCA420, OVCA429, OVCA432, OVCA433, OVCA633, OVCA680, OVCA702, OVCA810, SKOV3, ES-2, TOV21G	XIAP	Downregulation of Xist may increase the expression of linked inhibitors of apoptosis protein and lead to the phenotype of drug	Non-small-cell lung cancer, breast cancer, and primary hepatocellular carcinoma	[91, 131, 132]
	H19	Imprinted maternally expressed transcript	Upregulated	SKOV3, OV90, TOV112D, ES2	Caspase-3, caspase-9, Bax, Bcl-2, cyclin B1/Cdc2	Promotes ovarian cancer cell proliferation	Head and neck cancer, pancreatic cancer, and osteosarcoma	[92, 93, 133]
	MALAT1	Metastasis-associated lung adenocarcinoma transcript 1	Upregulated	SKOV3, SKOV3.ip1, 293T	—	Promotes cell proliferation and metastasis and inhibits cell apoptosis	Glioma, hepatocellular carcinoma, and multiple myeloma	[94–96, 127]
Ovarian cancer	LINC00565	Long intergenic nonprotein coding RNA 565	Upregulated	OVCAR3, SKOV3, HO8910, A2780, and HEY	GAS6, cyclinE1, cyclinD1, CDK4 P16, P21	Relates to the FIGO (International Federation of Gynecology and Obstetrics) stage, cell cycle, and size of tumor cells and promotes cell proliferation, invasion, and migration	Gastric cancer and colorectal cancer	[97, 134, 135]
	DARS-AS1	DARS1 antisense RNA 1	Upregulated	A2780, SKOV3, and OVCAR-3	Sponge for miR-532-3p	Promotes the proliferation, migration, and invasion of ovarian cancer cells	Thyroid cancer, clear cell renal cell carcinoma, and non-small-cell lung cancer	[98, 136, 137]
	FEZF1-AS1	FEZF1 antisense RNA 1	Upregulated	SKOV-3, HO8910, HO8910PM, ES2, and HG-SOC	JAK-STAT3 pathway	Relates to poor prognosis, promotes cell proliferation, and inhibits cell apoptosis	Colorectal cancer, gastric neoplasia, and hepatocellular carcinoma	[99, 138]
	LEF1-AS1	LEF1 antisense RNA 1	Upregulated	SKOV3, OVCAR3	miR-1285-3p	The absence of LEF1-AS1 results in inhibiting proliferation, migration, and invasion of ovarian cancer cells	Glioblastoma, colorectal cancer, and retinoblastoma	[100, 139–141]
	H19		Upregulated	—	—			[102, 133]

Table 1: Continued.

Diseases	lncRNA	Full name	Expression level	Assessed cell line	Signaling pathways and molecules	Functions	In other diseases	References
		Imprinted maternally expressed transcript				Regulates migration and invasion of the tumor cells	Head and neck cancer, pancreatic cancer, and osteosarcoma	
Endometrial carcinoma (EC)	CCAT1	Colon cancer-associated transcript 1	Upregulated	HEC-1-A, KLE, Ishikawa	Sponge for miR-181a-5p	Promotes the proliferation and migration of endometrial cancer cells	Breast cancer and multiple myeloma	[103, 142]
	MIR22HG	MIR22 host gene	Downregulated	HEC-1 A, KLE	Sponge for miR-141-3p	Inhibits the proliferation and migration and promotes the apoptosis of cancer cells	Esophageal cancer, lung cancer, and hepatocellular carcinoma	[104, 143]
	MEG3	Maternal expression gene 3	Downregulated	Ishikawa, HEC-1B	PI3K/m-TOR pathway, BclxL, VEGFA	High expression of MEG3 inhibits the migration, invasion, and proliferation of EC cells and increases apoptosis	Gastric cancer, osteosarcoma, and breast cancer	[105, 144]
	AC002454.1	—	Upregulated	—	CDK6	Promotes the migration, invasion, and proliferation of cells and regulates the cell cycle	Bladder cancer	[107, 145]
	MALAT1	Metastasis-associated lung adenocarcinoma transcript 1	Upregulated	—	NF-κB/iNOS pathway, MMP-9, caspase-3	Promotes the proliferation and migration of endometrial cells	Glioma, hepatocellular carcinoma, and multiple myeloma	[108, 127]
Endometriosis	AFAP1-AS1	Actin filament-associated protein 1 Antisense RNA1	Upregulated	Ishikawa	ZEB1	Promotes the EMT process of endometriosis	Esophageal cancer, pancreatic ductal adenocarcinoma	[109, 146]
	CCDC144NL-AS1	CCDC144NL antisense RNA 1	Upregulated	hEM15A	MMP-9, F-actin, vimentin	Affects the cytoskeleton structure and promotes cell invasion and migration	Osteosarcoma, gastric cancer, non-small-cell lung cancer, and hepatocellular carcinoma	[110, 147–150]
	TC0101441	—	Upregulated	ECSCs	TCF8/ZEB1, slug, snail, and N-cadherin	EV shuttling of TC0101441 promotes invasion and migration of endometriosis	Gastric cancer	[151, 152]

Table 1: Continued.

Diseases	lncRNA	Full name	Expression level	Assessed cell line	Signaling pathways and molecules	Functions	In other diseases	References
	UCA1	Urothelial carcinoma-associated-1	Downregulated	—	—	Is involved in the pathogenesis of endometriosis and can be used as a biomarker for diagnosis and prognosis	Urothelial carcinoma-associated 1 gastric cancer and colorectal cancer	[153, 154]
	H19	Imprinted maternally expressed transcript	Downregulated	293T, HESCs	H19/Let-7/IGF1R, H19/miR-216a-5p/ACTA2 pathway	Regulates endometrial stromal cell proliferation, invasion, and migration	Head and neck cancer, pancreatic cancer, and osteosarcoma	[133, 155, 156]
	aHIF	Antisense hypoxia-inducible factor	Upregulated	ECSCs, HUVECs	(VEGF)-A, VEGF-D	Facilitates endometriosis angiogenesis and is used as a potential biomarker and therapeutic target for endometriosis	Gastric cancer, glioblastoma multiforme, and paraganglioma	[157–159]
	MALAT1	Metastasis-associated lung adenocarcinoma transcript 1	Upregulated	—	HeLa, CaSki	Promotes the proliferation and invasion of cervical cancer cells and reduces apoptosis	Glioma, hepatocellular carcinoma, and multiple myeloma	[112, 127]
	HOTAIR	Hox transcript antisense intergenic RNA	Upregulated	SiHa, HeLa, Caski	VEGF, MMP-9, E-cadherin, β-catenin, vimentin, snail, twist	Promotes metastasis and invasion of tumor cells	Breast cancer, lung cancer, and pancreatic cancer	[113, 123]
Cervical cancer	RP11-480I12.5	—	Upregulated	PCS-480-011, SiHa (HTB-35), HeLa229 (CCL-2.1), and MS751	Wnt/β-catenin pathway	Induces EMT through the Wnt/β-catenin pathway and promotes migration, invasion, and proliferation of cervical cancer cell lines	Breast cancer	[114, 160, 161]
	RP1-93H18.6	—	Upregulated	SiHa, HeLa, CaSki, and C-33A	PI3K/Akt/mTOR pathway	Promotes growth and metastasis of tumor cells and reduces apoptosis	—	[115]
	DSCAM-AS1	DSCAM antisense RNA 1	Upregulated	SiHa, HeLa, C-33A, and CaSki	Sponge for miR-361-5p	Enhances the ability of cells to migrate, invade, and proliferate and promotes the development of cervical cancer	Non-small-cell lung cancer, colorectal cancer, and osteosarcoma	[116, 162]
	GAS5-AS1	GAS5 antisense RNA 1	Downregulated	Caski, SiHa, C33A, and HeLa	GAS5	Relates to the FLGO stage, lymphatic metastasis, distant	Glioma, non-small-cell lung	[117, 153, 163, 164]

Table 1: Continued.

Diseases	lncRNA	Full name	Expression level	Assessed cell line	Signaling pathways and molecules	Functions	In other diseases	References
						metastasis, and poor prognosis and promotes proliferation, migration, and invasion	cancer, and hepatocellular carcinoma	
	PVT1	Plasmacytoma variant translocation-1	Upregulated	HeLa and SiHa	Sponge for miR-140-5p, Smad3	Promotes the proliferation and metastasis of cervical cancer	Clear cell renal cell carcinoma and thyroid cancer	[118, 165]
	H19	Imprinted maternally expressed transcript	Upregulated	Peripheral blood leukocytes	FPG	May be a key factor in endocrine and metabolic diseases in patients with PCOS	Head and neck cancer, pancreatic cancer, and osteosarcoma	[120, 133]
	PVT1	Plasmacytoma variant translocation-1	Upregulated	—	Sponge for miR-17-5p	Regulates the apoptosis and the proliferation of ovarian granulosa cells	Clear cell renal cell carcinoma and thyroid cancer	[165, 166]
Polycystic ovary syndrome (PCOS)	LET	Low expression in tumor	Downregulated	KGN	Wnt/β-catenin and Notch pathways, TIMP2	Promotes cell migration and survival and inhibits cell apoptosis	Hepatocellular carcinoma, colorectal cancer, and squamous cell lung carcinoma tissues	[167, 168]
	TMPO-AS1	TMPO antisense RNA 1	Upregulated	COV434	Sponge for miR-355-5p	Serve as a potential target to treat PCOS	Lung cancer, breast cancer, and colorectal cancer	[169, 170]
	NEAT1	Nuclear-rich transcriptase 1	Upregulated	Ovarian tissue in rats	Sponge for miR-381, IGF1	Promotes cell proliferation and represses cell apoptosis	Non-small-cell lung cancer, breast cancer, and hepatocellular carcinoma	[125, 171]
	LINC00477	Long intergenic nonprotein coding RNA 477	Upregulated	Sponge for miR-128	Sponge for miR-128	LINC00477/miR-128 axis may represent a potential method for the treatment of PCOS	Gastric cancer	[172, 173]

independent AR-mediated genes and promote the proliferation of prostate cancer cells [84].

Nuclear-rich transcriptase 1 (*NEAT1*), a potential target of estrogen receptor alpha (ERα), is significantly overexpressed in prostate cancer. NEAT1 is shown to regulate the expression of prostate cancer genes and promotes the development of prostate cancer by changing the epigenetic landscape of the target gene promoter [85].

PCAT-1 is upregulated in prostate cancer and promotes the proliferation of prostate cancer cells through PRC2 and cMyc proteins [86].

MALAT-1 is upregulated in prostate cancer and is associated with the increase in the Gleason score, prostate-specific antigen (PSA), and tumor stage. Downregulating the expression of MALAT-1 inhibits the migration, invasion, and growth of prostate cancer cells, increases the rate of apoptosis, and blocks the cell cycle [87].

SChLAP1 is highly expressed in prostate cancer and is associated with a poor prognosis. Thus, it could be used as an essential biomarker to identify patients with a high risk of lethal prostate cancer [88].

GAS5 is downregulated in prostate cancer cells compared with prostate epithelial cells. GAS5 inhibits prostate cancer cell proliferation. It can bind directly to E2F1 and activate the $P27^{Kip1}$ which is a regulator of the cell cycle. Thus, GAS5 induces a cell cycle arrest in the G0–G1 phase and acts as a tumor suppressor [89].

5.3. lncRNAs Associated with Ovarian Cancer. Ovarian cancer is one of the most common gynecological cancers that affect women's health worldwide. As there has been no effective method to detect ovarian cancer at an early stage, most patients are diagnosed in an advanced stage, which has developed resistance to multiple treatment modalities [90]. Despite the revolutionary role of surgery and chemotherapy in curing ovarian cancer, the overall prognosis of ovarian cancer is poor. Thus, improving our understanding of the pathogenesis of ovarian cancer is essential for developing more effective treatments.

XIST encodes a specific spliced lncRNA, and it is a vital regulator of X chromosome inactivation. It is identified to be the most differentially expressed gene and downregulated in recurrent ovarian tumors. Downregulation of Xist may increase the expression of linked inhibitors of apoptosis protein (X-linked Inhibitor of Apoptosis Protein (XIAP)) and lead to the phenotype of drug resistance [91].

H19 is significantly increased in ovarian cancer cells and ovarian cancer tissues. Ectopic expression of H19 promotes cell proliferation while silencing the expression of H19 by RNA interference inhibits the growth of ovarian cancer cells and induces cell cycle arrest and apoptosis [92]. Moreover, overexpression of H19 enhances the ability of tumor cells to invade *in vitro* and metastasize *in vivo* [93].

Metastasis-associated lung adenocarcinoma transcript 1 (*MALAT1*) is one of the earliest cancer-related lncRNAs identified to be related to ovarian cancer [94]. The expression level of MALAT1 is associated with ovarian cancer cells with different metastatic potentials. MALAT1 may play a role in the metastasis of epithelial ovarian cancer cells, but its mechanism needs to be further studied [95]. Knockdown of MALAT1 in ovarian cancer cells changes the expression of many genes related to cell proliferation, metastasis, and apoptosis, and inhibition of MALAT1 can significantly inhibit the tumorigenicity of SKOV3 cells [96].

LINC00565 is highly expressed in ovarian cancer tissues, and its expression level was negatively correlated with the prognosis of patients with ovarian cancer. It has been found that the expression level of LINC00565 is related to the FIGO (International Federation of Gynecology and Obstetrics) stage and the size of tumor cells. Knockdown of LINC00565 in ovarian cancer cells inhibits the proliferation, invasion, and migration of the cells and induces cell cycle arrest. *In vivo* studies have shown that downregulating the expression of LINC00565 has an inhibitory effect on the growth of ovarian cancer cells by mediating the expression of cell cycle-related genes [97].

DARS-AS1 is expressed higher in ovarian cancer tissues than in adjacent normal tissues. It promotes the migration and invasion of ovarian cancer cells. MicroRNA-532-3p (miR-532-3p) is identified as the direct target of DARS-AS1 in ovarian cancer, and DARS-AS1 via sponging miR-532-3p promotes the proliferation, migration, and invasion of ovarian cancer cells [98].

FEZF1-AS1 is identified as a carcinogenic gene in ovarian cancer, as it is highly expressed in ovarian cancer tissues compared with adjacent normal tissues. Its expression is associated with a poor prognosis. After knocking down FEZF1-AS1, the proliferation of ovarian cancer cells was inhibited, and apoptosis was promoted. The mechanistic analysis found that FEZF1-AS1 regulated the JAK-STAT3 signal pathway by regulating the phosphorylation of STAT3 [99].

LEF1-AS1 is upregulated in ovarian cancer and is related to poor prognosis. The absence of LEF1-AS1 results in the inhibition of proliferation, migration, and invasion of ovarian cancer cells. LEF1-AS1 interacts with miR-1285-3p, a tumor suppressor in ovarian cancer, to inhibit the expression of miR-1285-3p and promote the growth and metastasis of ovarian cancer cells [100].

5.4. lncRNAs Associated with Endometrial Carcinoma (EC). Endometrial carcinoma is the most common cancer in the uterus. It is formed by the outgrowth of the cells that develop the glands in the endometrium. Although it tends to have a favorable prognosis if an early sign of abnormal uterine bleeding is presented, once it develops into metastasis or recurrence, the patients are at a significantly higher risk of mortality, with a median overall survival time of <16 weeks [101]. The genetic factors that cause endometrial carcinoma remain unclear, and a growing number of studies have associated lncRNAs with its initiation and progression.

H19 is expressed higher in EC and tumor tissues than in the normal endometrial epithelium, and it regulates migration and invasion of the tumor cells [102].

Colon cancer-associated transcript 1 (*CCAT1*) is expressed significantly higher in EC and tumor tissues than in normal endometrial tissue. Downregulation of CCAT1 expression leads to the inhibition of tumor cell growth and

metastasis. In addition, it was found that CCAT1 was the direct target of miR181a-5p in endometrial carcinoma cells. It promotes the proliferation and migration of endometrial cancer cells by negatively regulating the expression of miR-181a-5p [103].

MIR22HG has been identified as a tumor repressor in EC. Its expression is significantly downregulated in endometrial carcinoma tissue. Functional tests *in vitro* showed that increased expression of MIR22HG could inhibit the proliferation and promote the apoptosis of cancer cells. In addition, the study proposed that MIR22HG inhibits the proliferation and migration of cancer cells by regulating the miR-141-3p/DAPK1 axis [104].

Maternal expression gene 3 (*MEG3*) is a tumor suppressor gene, and its expression level in EC tissue is significantly lower than that in normal endometrial tissue. High expression of MEG3 inhibits the migration, invasion, and proliferation of EC cells and increases apoptosis, probably through the PI3K/mTOR signal transduction pathway [105].

5.5. lncRNAs Associated with Endometriosis. Endometriosis is a benign gynecological disorder characterized by the presence of endometrial cells from the lining of the uterus outside of the uterine cavity. Although research efforts have been devoted to uncovering the underlying cause of endometriosis, the pathophysiological mechanisms causing this disease remained obscure. Recent studies, especially the results from high-throughput RNA sequencing [106], have shown differential expression of lncRNAs in endometriosis-related tissues and indicate the contribution of lncRNAs to the pathogenesis of endometriosis.

AC002454.1 is upregulated with cyclin-dependent kinase-6 (CDK6) in patients with endometriosis, and there was a significant positive correlation between them. After downregulating the expression of AC002454.1 and CDK6, the ability of cells to migrate, invade, and proliferate decreased, the proportion of cells in the S phase decreased, and the proportion of cells in the G0/G1 phase increased. Therefore, AC002454.1 and CDK6 have a synergistic effect on the biological behavior of endometrial cells [107].

MALAT1 plays a vital role in endometriosis. Compared with normal tissues, the expression of MALAT1 in endometriosis is upregulated. Knockdown of MALAT1 inhibits the proliferation and migration of endometrial cells, enhances the activity of caspase-3, and induces apoptosis by inhibiting the NF-κB/iNOS signal pathway [108].

AFAP1-AS1 is significantly upregulated in ectopic endometrial tissues and is positively correlated with epithelial-mesenchymal transition (EMT). Knocking down AFAP1-AS1 can inhibit the activity of the EMT-related transcription factor ZEB1, thus inhibiting the EMT process of endometriosis [109].

CCDC144NL-AS1 is a newly identified lncRNA whose expression is upregulated in ectopic endometrium tissues. Downregulation of CCDC144NL-AS1 inhibited the migration and invasion of EC cell lines. Mechanism studies have shown that the knockdown of CCDC144NL-AS1 leads to changes in the distribution of filamentous actin (F-actin) stress fibers in the cytoskeleton and affects the cytoskeleton structure. In addition, the expression of the CCDC144NL-AS1 gene promotes the protein expression of vimentin filament and matrix metalloproteinase-9 (MMP-9), which promotes cell invasion and migration [110].

5.6. lncRNAs Associated with Cervical Cancer. Cervical cancer is one of the most frequently diagnosed malignant gynecological cancers that endanger women's health and lives [111]. Increasing data have shown the regulatory roles of lncRNAs in the pathogenesis of cervical cancer, with the prospective clinical application in the diagnosis and treatment of cervical cancers.

In cervical cancer, the expression of IGF2 was significantly increased, and the expression of *H19* was decreased considerably. However, the mechanism of this disorder is not precise, and further research is needed [102].

MALAT1 is identified as an essential regulatory factor involved in the occurrence of cervical cancer. Its expression in cervical cancer tissues is significantly higher than that in normal tissues. When endogenous MALAT1 is knocked out, it reduces the proliferation and invasion of cervical cancer cells and promotes apoptosis [112].

The expression of *HOTAIR* in cervical cancer is higher than that in normal tissues. HOTAIR has indicated a role in metastasis and invasion of tumor cells by regulating the expression of vascular endothelial growth factor, matrix metalloprotein-9, and epithelial-to-mesenchymal transformation- (EMT-) related genes [113].

The expression level of *RP11-480I12.5* in the cervical carcinoma cell line is higher than that in normal tissue. RP11-480I12.5 induces EMT through the Wnt/β-catenin pathway and promotes cervical cancer cell lines' migration, invasion, and proliferation [114].

lncRNARP1-93H18.6 is expressed higher in paracancerous tissues in cervical cancer and specific in cervical cancer. Overexpression of RP1-93H18.6 promotes growth and metastasis of tumor cells and reduces apoptosis. Knocking down the expression of lncRNARP1-93H18.6 promotes apoptosis and inhibits the development of cervical carcinoma cells by blocking the PI3K/Akt/mTOR pathway [115].

DSCAM-AS1 is related to the occurrence and development of various tumors, and its role in cervical cancer has recently been studied. The expression of DSCAM-AS1 in cervical carcinoma is increased. DSCAM-AS1 enhances the ability of cells to migrate, invade, and proliferate and promotes the development of cervical cancer through regulating the miR-877-5p/ATXN7L3 axis [116].

GAS5 is a tumor suppressor factor that inhibits proliferation, EMT, invasion, and metastasis of tumor cells. *GAS5-AS1* is the antisense RNA of GAS5, located on chromosome 1q25.1. Compared with normal tissues adjacent to cancer, the expression of GAS5-AS1 in cervical cancer is downregulated, and its expression is related to the FLGO stage, lymphatic metastasis, distant metastasis, and poor prognosis in patients with cervical cancer. Mechanistically, GAS5-AS1 regulates the tumor suppressor GAS5 in an ALKBH5-m6A-YTHDF2-dependent manner. Specifically, GAS5-AS1 reduced the level of GAS5N6-methyladenosine (M6A) modification and improved the stability of GAS5 through the

interaction of RNA demethylase and ALKBH5. In addition, YTHDF2 specifically recognizes and binds to the RNA containing M6A and degrades M6A-modified transcript [117].

Plasmacytoma variant translocation-1 (*PVT1*) promotes the proliferation and metastasis of cervical cancer. The expression of PVT1 is upregulated in cervical cancer cells, and PVT1 binds directly to miR-140-5p, which promotes the expression of Smad3 and then promotes the development of cervical cancer [118].

5.7. lncRNAs Associated with Polycystic Ovary Syndrome (PCOS). Polycystic ovary syndrome (PCOS) is one of the most common metabolic and reproductive disorders that has been estimated to affect approximately 5 to 20% of reproductive-aged women worldwide [119]. Although the etiology of PCOS remains unclear, most researchers believe that the causes are multifactorial, and lncRNAs have recently been suggested to play pivotal roles in its pathogenesis and prognosis.

H19 is suggested to be involved in the occurrence and development of PCOS. In patients with PCOS, the expression of H19 is increased. The expression level of fasting plasma glucose (FPG), a sensitive indicator in the early stage of metabolic disease, is positively correlated with H19 in PCOS patients. These results suggest that the expression of H19 may be a critical factor in endocrine and metabolic disorders in patients with PCOS [120].

Taken together, many lncRNAs, including H19, NEAT1, MALAT1, HOTAIR, and PVT1, are upregulated in the progression of many reproductive diseases. Interestingly, the expression of several lncRNAs, which is highly expressed in embryonic development, is reactivated in the development of reproductive cancer. For example, H19 is highly expressed in embryonic stem cells and essential for early human embryonic development. While its expression is downregulated after birth, the expression of H19 is significantly upregulated in endometrial carcinoma and ovarian cancer [121]. Recently, the reemergence of fetal-associated features in the tumor ecosystem is getting much attention and is referred to as oncofetal reprogramming [122]. Upregulation of specific lncRNAs in reproductive cancer development could be one of the features reminiscent of fetal development and serves as one of the potential targets for therapeutic interventions.

6. Conclusion and Future Perspectives

With the advances in sequencing technology, especially at the single-cell level, more and more lncRNAs have been identified at specific stages or within a particular type of cells, during human embryo and reproductive development. While expanding the repositories of lncRNAs, we notice that a unique subset of lncRNAs is expressed during human development. Dissection of the function of human-specific lncRNAs may be of preeminent importance for understanding the unique specifics of human development.

As a newly discovered role in gene regulatory networks, lncRNAs provide an additional layer of complexity for transcriptional and posttranscriptional regulation of gene expression programs. In addition, an increasing number of lncRNAs are differentially expressed within the disease tissues. They were found to regulate the initiation and progression of reproductive diseases through mediating the gene expression program. However, most of the functional results are based on the analysis *in vitro* on disease-related cell lines. Rigorous investigations *in vivo* or in organoids that resemble the physiological environment of development or diseases are necessary to reveal the biological and physiological functions of lncRNAs.

lncRNAs are proposed as therapeutic or diagnostic targets for disease treatment, as many of their expression are restricted to a specific tissue/or cell type within a specific cellular stage, which renders superior specificity. Furthermore, the diversity of strategies to target lncRNAs offers a wide range of therapeutic options. At the transcription level, we can inhibit the expression of lncRNAs by genome editing techniques or upregulate their expression by knockdown of the corresponding natural antisense transcripts (NATs). At the posttranslational level, lncRNAs can be degraded by nucleic acid-based approaches, including siRNAs, antisense oligonucleotides (ASO), and morpholinos.

Although immense enthusiasm is aroused in the field of lncRNA-based therapy, especially nucleic acid-based approaches, several challenges must be addressed before the progression to large-scale clinical applications. First, we need to have a thorough understanding of the molecular function of lncRNAs to identify disease-determining lncRNAs. Second, robust and physiologically relevant preclinical models need to be established. As we mentioned above, a few lncRNAs associated with diseases are human/primate-specific or even patient-specific. So patient-derived xenograft models or 3D organoids have gained much interest in preclinical research. Third, for nucleic acid-based therapies, a lack of an efficient delivery system to cross the cellular plasma membrane, the risk of the overactivating innate immune response, and the possibility of the off-target effect are the main issues that need to be solved.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (32070830) and research funds from the University of Science and Technology of China (WK9110000141 and YD9100002007).

References

[1] International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, Center for Genome Research, E. S. Lander et al., "Initial sequencing and analysis of the human genome," *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.

[2] K. V. Morris and J. S. Mattick, "The rise of regulatory RNA," *Nature Reviews. Genetics*, vol. 15, no. 6, pp. 423–437, 2014.

[3] D. K. Niu and L. Jiang, "Can ENCODE tell us how much junk DNA we carry in our genome?," *Biochemical and Biophysical Research Communications*, vol. 430, no. 4, pp. 1340–1343, 2013.

[4] P. Bertone, V. Stolc, T. E. Royce et al., "Global identification of human transcribed sequences with genome tiling arrays," *Science*, vol. 306, no. 5705, pp. 2242–2246, 2004.

[5] P. Carninci, T. Kasukawa, S. Katayama et al., "The transcriptional landscape of the mammalian genome," *Science*, vol. 309, no. 5740, pp. 1559–1563, 2005.

[6] B. Wu, K. Wang, J. Fei et al., "Novel three-lncRNA signature predicts survival in patients with pancreatic cancer," *Oncology Reports*, vol. 40, no. 6, pp. 3427–3437, 2018.

[7] C. P. Ponting, P. L. Oliver, and W. Reik, "Evolution and functions of long noncoding RNAs," *Cell*, vol. 136, no. 4, pp. 629–641, 2009.

[8] R. Chen, G. Wang, Y. Zheng, Y. Hua, and Z. Cai, "Long noncoding RNAs in osteosarcoma," *Oncotarget*, vol. 8, no. 12, pp. 20462–20475, 2017.

[9] A. F. Palazzo and E. S. Lee, "Non-coding RNA: what is functional and what is junk?," *Frontiers in Genetics*, vol. 6, p. 2, 2015.

[10] M. Joshi and S. Rajender, "Long non-coding RNAs (lncRNAs) in spermatogenesis and male infertility," *Reproductive Biology and Endocrinology*, vol. 18, no. 1, 2020.

[11] Z. S. Razavi, V. Tajiknia, S. Majidi et al., "Gynecologic cancers and non-coding RNAs: epigenetic regulators with emerging roles," *Critical Reviews in Oncology/Hematology*, vol. 157, 2021.

[12] F. Muscatelli, D. Lena, M. G. Mettei, and M. Fontes, "A male with two contiguous inactivation centers on a single X chromosome: study of X inactivation and XIST expression," *Human Molecular Genetics*, vol. 1, no. 2, pp. 115–119, 1992.

[13] A. Gabory, H. Jammes, and L. Dandolo, "The H19 locus: role of an imprinted non-coding RNA in growth and development," *BioEssays*, vol. 32, no. 6, pp. 473–480, 2010.

[14] J. Bouckenheimer, S. Assou, S. Riquier et al., "Long non-coding RNAs in human early embryonic development and their potential in ART," *Human Reproduction Update*, vol. 23, no. 1, pp. 19–40, 2016.

[15] S. Hermans-Beijnsberger, M. van Bilsen, and B. Schroen, "Long non-coding RNAs in the failing heart and vasculature," *Non-coding RNA research*, vol. 3, no. 3, pp. 118–130, 2018.

[16] P. Yan, S. Luo, J. Y. Lu, and X. Shen, "*Cis*\- and *trans* -acting lncRNAs in pluripotency and reprogramming," *Current Opinion in Genetics & Development*, vol. 46, pp. 170–178, 2017.

[17] V. S. Akhade, D. Pal, and C. Kanduri, "Long noncoding RNA: genome organization and mechanism of action," *Advances in Experimental Medicine and Biology*, vol. 1008, pp. 47–74, 2017.

[18] M. J. Delás and G. J. Hannon, "lncRNAs in development and disease: from functions to mechanisms," *Open biology*, vol. 7, no. 7, 2017.

[19] K. C. Wang and H. Y. Chang, "Molecular mechanisms of long noncoding RNAs," *Molecular Cell*, vol. 43, no. 6, pp. 904–914, 2011.

[20] J. Zhang, T. Xu, L. Liu et al., "LMSM: a modular approach for identifying lncRNA related miRNA sponge modules in breast cancer," *PLoS Computational Biology*, vol. 16, no. 4, 2020.

[21] V. Fort, G. Khelifi, and S. M. I. Hussein, "Long non-coding RNAs and transposable elements: a functional relationship," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1868, no. 1, article 118837, 2021.

[22] Y. Wang, Z. Xu, J. Jiang et al., "Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal," *Developmental Cell*, vol. 25, no. 1, pp. 69–80, 2013.

[23] A. Wutz, T. P. Rasmussen, and R. Jaenisch, "Chromosomal silencing and localization are mediated by different domains of *Xist* RNA," *Nature Genetics*, vol. 30, no. 2, pp. 167–174, 2002.

[24] E. A. Elisaphenko, N. N. Kolesnikov, A. I. Shevchenko et al., "A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements," *PLoS One*, vol. 3, no. 6, 2008.

[25] C. Chu, Q. C. Zhang, S. T. da Rocha et al., "Systematic discovery of Xist RNA binding proteins," *Cell*, vol. 161, no. 2, pp. 404–416, 2015.

[26] G. Pintacuda, G. Wei, C. Roustan et al., "hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb- Mediated Chromosomal Silencing," *Molecular Cell*, vol. 68, no. 5, pp. 955–969.e10, 2017.

[27] G. Pintacuda, A. N. Young, and A. Cerase, "Function by structure: spotlights on Xist long non-coding RNA," *Frontiers in Molecular Biosciences*, vol. 4, p. 90, 2017.

[28] S. T. da Rocha, V. Boeva, M. Escamilla-del-Arenal et al., "Jarid2 is implicated in the initial Xist-induced targeting of PRC2 to the inactive X chromosome," *Molecular Cell*, vol. 53, no. 2, pp. 301–316, 2014.

[29] N. Irie, A. Sybirna, and M. A. Surani, "What can stem cell models tell us about human germ cell biology?," *Current Topics in Developmental Biology*, vol. 129, pp. 25–65, 2018.

[30] D. Sabour and H. R. Schöler, "Reprogramming and the mammalian germline: the Weismann barrier revisited," *Current Opinion in Cell Biology*, vol. 24, no. 6, pp. 716–723, 2012.

[31] J. Wu and I. Barbaric, "Fitness selection in human pluripotent stem cells and interspecies chimeras: implications for human development and regenerative medicine," *Developmental Biology*, vol. 476, pp. 209–217, 2021.

[32] L. Yan, M. Yang, H. Guo et al., "Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells," *Nature Structural & Molecular Biology*, vol. 20, no. 9, pp. 1131–1139, 2013.

[33] J. Li, W. Han, X. Shen, S. Han, H. Ye, and G. Huang, "DNA methylation signature of long noncoding RNA genes during human pre-implantation embryonic development," *Oncotarget*, vol. 8, no. 34, pp. 56829–56838, 2017.

[34] V. Alcazer, P. Bonaventura, and S. Depil, "Human endogenous retroviruses (HERVs): shaping the innate immune response in cancers," *Cancers*, vol. 12, no. 3, 2020.

[35] D. R. Fuentes, T. Swigut, and J. Wysocka, "Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation," *eLife*, vol. 7, 2018.

[36] 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.

[37] C. Römer, M. Singh, L. D. Hurst, and Z. Izsvák, "How to tame an endogenous retrovirus: HERVH and the evolution of human pluripotency," *Current Opinion in Virology*, vol. 25, pp. 49–58, 2017.

[38] 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.

[39] C. Vallot, C. Patrat, A. J. Collier et al., "_XACT_ Noncoding RNA Competes with _XIST_ in the Control of X Chromosome Activity during Human Early Development," *Cell Stem Cell*, vol. 20, no. 1, pp. 102–111, 2017.

[40] Y. L. Guo, "The underdeveloped innate immunity in embryonic stem cells: the molecular basis and biological perspectives from early embryogenesis," *American Journal of Reproductive Immunology*, vol. 81, no. 2, 2019.

[41] A. Jacob, M. Morley, F. Hawkins et al., "Differentiation of human pluripotent stem cells into functional lung alveolar epithelial cells," *Cell Stem Cell*, vol. 21, no. 4, pp. 472–488.e10, 2017.

[42] S. Lee, H. H. Seo, C. Y. Lee et al., "Human long noncoding RNA regulation of stem cell potency and differentiation," *Stem Cells International*, vol. 2017, Article ID 6374504, 10 pages, 2017.

[43] Z. du, X. Wen, Y. Wang et al., "Chromatin lncRNA Platr10 controls stem cell pluripotency by coordinating an intrachromosomal regulatory network," *Genome Biology*, vol. 22, no. 1, p. 233, 2021.

[44] X. Lu, F. Sachs, L. A. Ramsay et al., "The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity," *Nature Structural & Molecular Biology*, vol. 21, no. 4, pp. 423–425, 2014.

[45] C. Xu, Y. Zhang, Q. Wang et al., "Long non-coding RNA GAS5 controls human embryonic stem cell self-renewal by maintaining NODAL signalling," *Nature Communications*, vol. 7, no. 1, 2016.

[46] Z. Jiapaer, G. Li, D. Ye et al., "LincU Preserves Naïve Pluripotency by Restricting ERK Activity in Embryonic Stem Cells," *Stem Cell Reports*, vol. 11, no. 2, pp. 395–409, 2018.

[47] C.-J. Guo, X. K. Ma, Y. H. Xing et al., "Distinct processing of lncRNAs contributes to non-conserved functions in stem cells," *Cell*, vol. 181, no. 3, pp. 621–636.e22, 2020.

[48] S.-Y. Ng, G. K. Bogu, B. S. Soh, and L. W. Stanton, "The Long Noncoding RNA *RMST* Interacts with SOX2 to Regulate Neurogenesis," *Molecular Cell*, vol. 51, no. 3, pp. 349–359, 2013.

[49] S.-Y. Ng, R. Johnson, and L. W. Stanton, "Human long noncoding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors," *The EMBO Journal*, vol. 31, no. 3, pp. 522–533, 2012.

[50] A. K. Jain, Y. Xi, R. McCarthy et al., "*LncPRESS1* Is a p53-Regulated LncRNA that Safeguards Pluripotency by Disrupting SIRT6-Mediated De-acetylation of Histone H3K56," *Molecular Cell*, vol. 64, no. 5, pp. 967–981, 2016.

[51] W. Jiang, Y. Liu, R. Liu, K. Zhang, and Y. Zhang, "The lncRNA *DEANR1* Facilitates Human Endoderm Differentiation by Activating *FOXA2* Expression," *Cell Reports*, vol. 11, no. 1, pp. 137–148, 2015.

[52] J. Yang, P. Lu, M. Li, C. Yan, T. Zhang, and W. Jiang, "_GATA6-AS1_ Regulates GATA6 Expression to Modulate Human Endoderm Differentiation," *Stem Cell Reports*, vol. 15, no. 3, pp. 694–705, 2020.

[53] Y.-F. Chen, Y. S. J. Li, C. H. Chou et al., "Control of matrix stiffness promotes endodermal lineage specification by regulating SMAD2/3 via lncRNA LINC00458," *Science Advances*, vol. 6, no. 6, 2020.

[54] J. Liu, Y. Li, B. Lin, Y. Sheng, and L. Yang, "*HBL1* Is a Human Long Noncoding RNA that Modulates Cardiomyocyte Development from Pluripotent Stem Cells by Counteracting *MIR1*," *Developmental Cell*, vol. 42, no. 4, pp. 333–348.e5, 2017.

[55] K. D. Wilson, M. Ameen, H. Guo et al., "Endogenous Retrovirus-Derived lncRNA _BANCR_ Promotes Cardiomyocyte Migration in Humans and Non-human Primates," *Developmental Cell*, vol. 54, no. 6, pp. 694–709.e9, 2020.

[56] S. Frank, G. Ahuja, D. Bartsch et al., "*yylncT* Defines a Class of Divergently Transcribed lncRNAs and Safeguards the T-mediated Mesodermal Commitment of Human PSCs," *Cell Stem Cell*, vol. 24, no. 2, pp. 318–327.e8, 2019.

[57] A. Zhang, N. Zhou, J. Huang et al., "The human long noncoding RNA-RoR is a p53 repressor in response to DNA damage," *Cell Research*, vol. 23, no. 3, pp. 340–350, 2013.

[58] X. Bao, H. Wu, X. Zhu et al., "The p53-induced *lincRNA-p21* derails somatic cell reprogramming by sustaining H3K9me3 and CpG methylation at pluripotency gene promoters," *Cell Research*, vol. 25, no. 1, pp. 80–92, 2015.

[59] X. Wang, Y. Shi, Z. Liu, and X. Y. Zhao, "Gamete differentiation from pluripotent stem cells," *National Science Review*, vol. 4, no. 4, pp. 525–528, 2017.

[60] I. Kawashima and K. Kawamura, "Disorganization of the germ cell pool leads to primary ovarian insufficiency," *Reproduction*, vol. 153, no. 6, pp. R205–r213, 2017.

[61] J. T. Kwon, S. Ham, S. Jeon, Y. Kim, S. Oh, and C. Cho, "Expression of uncharacterized male germ cell-specific genes and discovery of novel sperm-tail proteins in mice," *PLoS One*, vol. 12, no. 7, 2017.

[62] K. S. Liu, T. P. Li, H. Ton, X. D. Mao, and Y. J. Chen, "Advances of long noncoding RNAs-mediated regulation in reproduction," *Chinese Medical Journal*, vol. 131, no. 2, pp. 226–234, 2018.

[63] D. H. Taylor, E. T. J. Chu, R. Spektor, and P. D. Soloway, "Long non-coding RNA regulation of reproduction and development," *Molecular Reproduction and Development*, vol. 82, no. 12, pp. 932–956, 2015.

[64] D. Yunusov, L. Anderson, L. F. DaSilva et al., "_HIPSTR_ and thousands of lncRNAs are heterogeneously expressed in human embryos, primordial germ cells and stable cell lines," *Scientific Reports*, vol. 6, no. 1, 2016.

[65] T. Chitiashvili, I. Dror, R. Kim et al., "Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation," *Nature Cell Biology*, vol. 22, no. 12, pp. 1436–1446, 2020.

[66] E. Sendler, G. D. Johnson, S. Mao et al., "Stability, delivery and functions of human sperm RNAs at fertilization," *Nucleic Acids Research*, vol. 41, no. 7, pp. 4104–4117, 2013.

[67] Z. Zhu, C. Li, S. Yang et al., "Dynamics of the transcriptome during human spermatogenesis: predicting the potential key genes regulating male gametes generation," *Scientific Reports*, vol. 6, no. 1, 2016.

[68] S. Z. Jan, T. L. Vormer, A. Jongejan et al., "Unraveling transcriptome dynamics in human spermatogenesis," *Development*, vol. 144, no. 20, pp. 3659–3673, 2017.

[69] A. D. Rolland, B. Evrard, T. A. Darde et al., "RNA profiling of human testicular cells identifies syntenic lncRNAs associated with spermatogenesis," *Human Reproduction*, vol. 34, no. 7, pp. 1278–1290, 2019.

[70] X. Zhang, P. Zhang, D. Song et al., "Expression profiles and characteristics of human lncRNA in normal and asthenozoospermia sperm†," *Biology of Reproduction*, vol. 100, no. 4, pp. 982–993, 2019.

[71] M. Lü, H. Tian, Y. X. Cao et al., "Downregulation of miR-320a/383-sponge-like long non-coding RNA NLC1-C (narcolepsy candidate-region 1 genes) is associated with male infertility and promotes testicular embryonal carcinoma cell proliferation," *Cell Death & Disease*, vol. 6, no. 11, 2015.

[72] J. V. Medrano, C. Ramathal, H. N. Nguyen, C. Simon, and R. A. Reijo Pera, "Divergent RNA-binding proteins, DAZL and VASA, induce meiotic progression in human germ cells derived in vitro," *Stem Cells*, vol. 30, no. 3, pp. 441–451, 2012.

[73] J. Huang, T. Liu, C. Shang et al., "Identification of lncRNAs by microarray analysis reveals the potential role of lncRNAs in cervical cancer pathogenesis," *Oncology Letters*, vol. 15, no. 4, pp. 5584–5592, 2018.

[74] H. Dong, S. Jiang, Y. Fu, Y. Luo, R. Gui, and J. Liu, "Upregulation of lncRNA NR_046683 serves as a prognostic biomarker and potential drug target for multiple myeloma," *Frontiers in Pharmacology*, vol. 10, p. 45, 2019.

[75] D. Khalife, A. Khalil, and G. Ghazeeri, "The developmental process of spermatogenesis," *Journal of Andrology & Gynaecology*, vol. 7, 2019.

[76] R. González and I. Dobrinski, "Beyond the mouse monopoly: studying the male germ line in domestic animal models," *ILAR Journal*, vol. 56, no. 1, pp. 83–98, 2015.

[77] W. Tang, Y. Zhu, W. Qin et al., "Ran-binding protein 3 is associated with human spermatogenesis and male infertility," *Andrologia*, vol. 52, no. 2, 2020.

[78] Q. X. Liang, Z. B. Wang, W. L. Lei et al., "Deletion of *Ck2β* gene causes germ cell development arrest and azoospermia in male mice," *Cell Proliferation*, vol. 53, no. 1, 2020.

[79] M. Liang, K. Hu, C. He, J. Zhou, and Y. Liao, "Upregulated lncRNA Gm2044 inhibits male germ cell development by acting as miR-202 host gene," *Animal Cells and Systems*, vol. 23, no. 2, pp. 128–134, 2019.

[80] L. Zhang, Z. Liu, X. Li et al., "Low long non-coding RNA HOTAIR expression is associated with down-regulation of Nrf2 in the spermatozoa of patients with asthenozoospermia or oligoasthenozoospermia," *International Journal of Clinical and Experimental Pathology*, vol. 8, no. 11, pp. 14198–14205, 2015.

[81] K. Chen, Z. Mai, Y. Zhou, X. Gao, and B. Yu, "Low NRF2 mRNA expression in spermatozoa from men with low sperm motility," *The Tohoku Journal of Experimental Medicine*, vol. 228, no. 3, pp. 259–266, 2012.

[82] J. Zhao, H. Li, H. Deng et al., "LncRNA gadd7, increased in varicocele patients, suppresses cell proliferation and promotes cell apoptosis," *Oncotarget*, vol. 9, no. 4, pp. 5105–5110, 2018.

[83] X. Dai, L. Liu, Z. Liang, K. Guo, S. Xu, and H. Wang, "Silencing of lncRNA MALAT1 inhibits cell cycle progression via androgen receptor signaling in prostate cancer cells," *Pathology, Research and Practice*, vol. 215, no. 4, pp. 712–721, 2019.

[84] L. Yang, C. Lin, C. Jin et al., "lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs," *Nature*, vol. 500, no. 7464, pp. 598–602, 2013.

[85] D. Chakravarty, A. Sboner, S. S. Nair et al., "The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer," *Nature Communications*, vol. 5, no. 1, 2014.

[86] J. R. Prensner, W. Chen, S. Han et al., "The Long Non-Coding RNA *PCAT-1* Promotes Prostate Cancer Cell Proliferation through cMyc," *Neoplasia*, vol. 16, no. 11, pp. 900–908, 2014.

[87] S. Ren, Y. Liu, W. Xu et al., "Long noncoding RNA MALAT-1 is a new potential therapeutic target for castration resistant prostate cancer," *The Journal of Urology*, vol. 190, no. 6, pp. 2278–2287, 2013.

[88] R. Mehra, A. M. Udager, T. U. Ahearn et al., "Overexpression of the long non-coding RNA SChLAP1 independently predicts lethal prostate cancer," *European Urology*, vol. 70, no. 4, pp. 549–552, 2016.

[89] X. Xu, J. Hou, J. Lv, Y. Huang, J. Pu, and L. Wang, "Overexpression of lncRNA GAS5 suppresses prostatic epithelial cell proliferation by regulating COX-2 in chronic nonbacterial prostatitis," *Cell Cycle*, vol. 18, no. 9, pp. 923–931, 2019.

[90] M. Koshiyama, N. Matsumura, and I. Konishi, "Subtypes of ovarian cancer and ovarian cancer screening," *Diagnostics*, vol. 7, no. 1, p. 12, 2017.

[91] J. M. Engreitz, A. Pandya-Jones, P. McDonel et al., "The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome," *Science*, vol. 341, no. 6147, 2013.

[92] Z. Zhu, L. Song, J. He, Y. Sun, X. Liu, and X. Zou, "Ectopic expressed long non-coding RNA H19 contributes to malignant cell behavior of ovarian cancer," *International Journal of Clinical and Experimental Pathology*, vol. 8, no. 9, pp. 10082–10091, 2015.

[93] I. J. Matouk, E. Raveh, R. Abu-lail et al., "Oncofetal H19 RNA promotes tumor metastasis," *Biochimica et Biophysica Acta*, vol. 1843, no. 7, pp. 1414–1426, 2014.

[94] Y. Dong, G. Liang, B. Yuan, C. Yang, R. Gao, and X. Zhou, "MALAT1 promotes the proliferation and metastasis of osteosarcoma cells by activating the PI3K/Akt pathway," *Tumour Biology*, vol. 36, no. 3, pp. 1477–1486, 2015.

[95] S. P. Liu, J. X. Yang, D. Y. Cao, and K. Shen, "Identification of differentially expressed long non-coding RNAs in human ovarian cancer cells with different metastatic potentials," *Cancer Biology & Medicine*, vol. 10, no. 3, pp. 138–141, 2013.

[96] S. Liu, X. Jiang, W. Li, D. Cao, K. Shen, and J. Yang, "Inhibition of the long non-coding RNA MALAT1 suppresses tumorigenicity and induces apoptosis in the human ovarian cancer SKOV3 cell line," *Oncology Letters*, vol. 11, no. 6, pp. 3686–3692, 2016.

[97] M. Gong, C. Luo, H. Meng et al., "Upregulated LINC00565 accelerates ovarian cancer progression by targeting GAS6," *Oncotargets and Therapy*, vol. Volume 12, pp. 10011–10022, 2019.

[98] K. Huang, W. S. Fan, X. Y. Fu, Y. L. Li, and Y. G. Meng, "Long noncoding RNA DARS-AS1 acts as an oncogene by targeting miR-532-3p in ovarian cancer," *European Review for Medical and Pharmacological Sciences*, vol. 24, no. 13, 2020.

[99] X. Zhao, Z. Cheng, and J. Wang, "Long noncoding RNA FEZF1-AS1 promotes proliferation and inhibits apoptosis in ovarian cancer by activation of JAK-STAT3 pathway," *Medical Science Monitor*, vol. 24, pp. 8088–8095, 2018.

[100] Y. Zhang and F. Ruan, "LncRNA LEF1-AS1 promotes ovarian cancer development through interacting with miR-1285-3p," *Cancer Management and Research*, vol. 12, pp. 687–694, 2020.

[101] B. L. Li and X. P. Wan, "The role of lncRNAs in the development of endometrial carcinoma (Review)," *Oncology Letters*, vol. 16, no. 3, pp. 3424–3429, 2018.

[102] C. Lee, S. J. Kim, J. Y. Na et al., "Alterations in promoter usage and expression levels of insulin-like growth factor-II and H19 genes in cervical and endometrial cancer," *Cancer Research and Treatment*, vol. 35, no. 4, pp. 314–322, 2003.

[103] J. Yu, L. Jiang, Y. Gao et al., "LncRNA CCAT1 negatively regulates miR-181a-5p to promote endometrial carcinoma cell proliferation and migration," *Experimental and Therapeutic Medicine*, vol. 17, no. 5, pp. 4259–4266, 2019.

[104] Z. Cui, X. An, J. Li, Q. Liu, and W. Liu, "LncRNA MIR22HG negatively regulates miR-141-3p to enhance DAPK1 expression and inhibits endometrial carcinoma cells proliferation," *Biomedicine & Pharmacotherapy*, vol. 104, pp. 223–228, 2018.

[105] K.-X. Sun, D. D. Wu, S. Chen, Y. Zhao, and Z. H. Zong, "LncRNA MEG3 inhibit endometrial carcinoma tumorigenesis and progression through PI3K pathway," *Apoptosis*, vol. 22, no. 12, pp. 1543–1552, 2017.

[106] D. Cui, J. Ma, Y. Liu et al., "Analysis of long non-coding RNA expression profiles using RNA sequencing in ovarian endometriosis," *Gene*, vol. 673, pp. 140–148, 2018.

[107] J. Liu, Y. Wang, P. Chen et al., "AC002454.1 and CDK6 synergistically promote endometrial cell migration and invasion in endometriosis," *Reproduction*, vol. 157, no. 6, pp. 535–543, 2019.

[108] J. Yu, L. H. Chen, B. Zhang, and Q. M. Zheng, "The modulation of endometriosis by lncRNA MALAT1 via NF-κB/iNOS," *European Review for Medical and Pharmacological Sciences*, vol. 23, no. 10, pp. 4073–4080, 2019.

[109] D. Lin, Q. Huang, R. Wu et al., "Long non-coding RNA AFAP1-AS1 promoting epithelial-mesenchymal transition of endometriosis is correlated with transcription factor ZEB1," *American Journal of Reproductive Immunology*, vol. 81, no. 1, 2019.

[110] C. Zhang, W. Wu, H. Zhu et al., "Knockdown of long noncoding RNA CCDC144NL-AS1 attenuates migration and invasion phenotypes in endometrial stromal cells from endometriosis†," *Biology of Reproduction*, vol. 100, no. 4, pp. 939–949, 2019.

[111] J. Dong, M. Su, W. Chang, K. Zhang, S. Wu, and T. Xu, "Long non-coding RNAs on the stage of cervical cancer," *Oncology Reports*, vol. 38, no. 4, pp. 1923–1931, 2017.

[112] L. Yang et al., "High MALAT1 expression predicts a poor prognosis of cervical cancer and promotes cancer cell growth and invasion," *European Review for Medical and Pharmacological Sciences*, vol. 19, no. 17, pp. 3187–3193, 2015.

[113] H. J. Kim, D. W. Lee, G. W. Yim et al., "Long non-coding RNA HOTAIR is associated with human cervical cancer progression," *International Journal of Oncology*, vol. 46, no. 2, pp. 521–530, 2015.

[114] L. Zhang, Y. Li, and L. Sona, "Long non-coding RNA RP11-480I12.5 promotes cervical carcinoma progression by regulating the Wnt/β-catenin signaling pathway," *Oncology Letters*, vol. 19, no. 1, pp. 469–475, 2020.

[115] Q. Wang, S. P. Yan, D. X. Chu et al., "Silencing of long noncoding RNA RP1-93H18.6 acts as a tumor suppressor in cervical cancer through the blockade of the PI3K/Akt axis," *Molecular Therapy-Nucleic Acids*, vol. 19, pp. 304–317, 2020.

[116] J. Liang, S. Zhang, W. Wang et al., "Long non-coding RNA DSCAM-AS1 contributes to the tumorigenesis of cervical cancer by targeting miR-877-5p/ATXN7L3 axis," *Bioscience Reports*, vol. 40, no. 1, 2020.

[117] X. Wang, J. Zhang, and Y. Wang, "Long noncoding RNA GAS5-AS1 suppresses growth and metastasis of cervical cancer by increasing GAS5 stability," *American Journal of Translational Research*, vol. 11, no. 8, pp. 4909–4921, 2019.

[118] Q.-Q. Chang, C. Y. Chen, Z. Chen, and S. Chang, "LncRNA PVT1 promotes proliferation and invasion through enhancing Smad3 expression by sponging miR-140-5p in cervical cancer," *Radiology and Oncology*, vol. 53, no. 4, pp. 443–452, 2019.

[119] Y. Li, C. Chen, Y. Ma et al., "Multi-system reproductive metabolic disorder: significance for the pathogenesis and therapy of polycystic ovary syndrome (PCOS)," *Life Sciences*, vol. 228, pp. 167–175, 2019.

[120] L. Qin, C. C. Huang, X. M. Yan, Y. Wang, Z. Y. Li, and X. C. Wei, "Long non-coding RNA H19 is associated with polycystic ovary syndrome in Chinese women: a preliminary study," *Endocrine Journal*, vol. 66, no. 7, pp. 587–595, 2019.

[121] E. Zeira, R. Abramovitch, K. Meir et al., "The knockdown of H19lncRNA reveals its regulatory role in pluripotency and tumorigenesis of human embryonic carcinoma cells," *Oncotarget*, vol. 6, no. 33, pp. 34691–34703, 2015.

[122] A. Sharma, J. J. W. Seow, C. A. Dutertre et al., "Onco-fetal reprogramming of endothelial cells drives immunosuppressive macrophages in hepatocellular carcinoma," *Cell*, vol. 183, no. 2, pp. 377–394.e21, 2020.

[123] T. Rajagopal, S. Talluri, R. L. Akshaya, and N. R. Dunna, "HOTAIR LncRNA: a novel oncogenic propellant in human cancer," *Clinica Chimica Acta*, vol. 503, pp. 1–18, 2020.

[124] A. Bardhan, A. Banerjee, K. Basu, D. K. Pal, and A. Ghosh, "PRNCR1: a long non-coding RNA with a pivotal oncogenic role in cancer," *Human Genetics*, vol. 141, no. 1, pp. 15–29, 2022.

[125] C. Klec, F. Prinz, and M. Pichler, "Involvement of the long noncodingRNA NEAT1 in carcinogenesis," *Molecular Oncology*, vol. 13, no. 1, pp. 46–60, 2019.

[126] T. Xiong, J. Li, F. Chen, and F. Zhang, "PCAT-1: a novel oncogenic long non-coding RNA in human cancers," *International Journal of Biological Sciences*, vol. 15, no. 4, pp. 847–856, 2019.

[127] B. Goyal, S. R. M. Yadav, N. Awasthee, S. Gupta, A. B. Kunnumakkara, and S. C. Gupta, "Diagnostic, prognostic, and therapeutic significance of long non-coding RNA MALAT1 in cancer," *Biochimica Et Biophysica Acta. Reviews on Cancer*, vol. 1875, no. 2, 2021.

[128] X. Bai, S. Zhang, J. Qiao et al., "Long non-coding RNA SChLAP1 regulates the proliferation of triple negative breast cancer cells via the miR-524-5p/HMGA2 axis," *Molecular Medicine Reports*, vol. 23, no. 6, 2021.

[129] J. Zhang, Z. Shi, Y. Nan, and M. Li, "Inhibiting malignant phenotypes of the bladder cancer cells by silencing long noncoding RNA SChLAP1," *International Urology and Nephrology*, vol. 48, no. 5, pp. 711–716, 2016.

[130] X. Yang, Z. Xie, X. Lei, and R. Gan, "Long non-coding RNA GAS5 in human cancer (Review)," *Oncology Letters*, vol. 20, no. 3, pp. 2587–2594, 2020.

[131] K. C. Huang, P. H. Rao, C. C. Lau et al., "Relationship of XIST expression and responses of ovarian cancer to chemotherapy," *Molecular Cancer Therapeutics*, vol. 1, no. 10, pp. 769–776, 2002.

[132] S. Yin, J. Dou, G. Yang, and F. Chen, "Long non-coding RNA XIST expression as a prognostic factor in human cancers: a meta-analysis," *The International Journal of Biological Markers*, vol. 34, no. 4, pp. 327–333, 2019.

[133] B. Alipoor, S. N. Parvar, Z. Sabati, H. Ghaedi, and H. Ghasemi, "An updated review of the H19 lncRNA in human cancer: molecular mechanism and diagnostic and therapeutic importance," *Molecular Biology Reports*, vol. 47, no. 8, pp. 6357–6374, 2020.

[134] J. Hu, G. Ni, L. Mao et al., " LINC00565 promotes proliferation and inhibits apoptosis of gastric cancer by targeting miR-665/AKT3 axis ," *Oncotargets and Therapy*, vol. -Volume 12, pp. 7865–7875, 2019.

[135] X. Shao, T. Zhao, L. Xi et al., "LINC00565 promotes the progression of colorectal cancer by upregulating EZH2," *Oncology Letters*, vol. 21, no. 1, p. 53, 2020.

[136] Y. Feng, G. Wei, L. Zhang et al., "LncRNA DARS-AS1 aggravates the growth and metastasis of hepatocellular carcinoma via regulating the miR-3200-5p-cytoskeleton associated protein 2 (CKAP2) axis," *Bioengineered*, vol. 12, no. 1, pp. 8217–8232, 2021.

[137] W. Zheng, X. Tian, L. Cai et al., "LncRNA DARS-AS1 regulates microRNA-129 to promote malignant progression of thyroid cancer," *European Review for Medical and Pharmacological Sciences*, vol. 23, no. 23, pp. 10443–10452, 2019.

[138] Y. Zhou, S. Xu, H. Xia et al., "Long noncoding RNA FEZF1-AS1 in human cancers," *Clinica Chimica Acta*, vol. 497, pp. 20–26, 2019.

[139] S. Zeng, C. Zhou, D. H. Yang et al., "LEF1-AS1 is implicated in the malignant development of glioblastoma via sponging miR-543 to upregulate EN2," *Brain Research*, vol. 1736, p. 146781, 2020.

[140] Y. Cheng, J. Wu, B. Qin, B. C. Zou, Y. H. Wang, and Y. Li, "CREB1-induced lncRNA LEF1-AS1 contributes to colorectal cancer progression via the miR-489/DIAPH1 axis," *Biochemical and Biophysical Research Communications*, vol. 526, no. 3, pp. 678–684, 2020.

[141] H. He and M. Qin, "Long non-coding RNA LEF1-AS1 is involved in the progression of retinoblastoma through regulating the Wnt/β-catenin pathway," *Clinical and Experimental Pharmacology & Physiology*, vol. 47, no. 5, pp. 886–891, 2020.

[142] Z. Liu, Q. Chen, and S. S. Hann, "The functions and oncogenic roles of CCAT1 in human cancer," *Biomedicine & Pharmacotherapy*, vol. 115, 2019.

[143] L. Zhang, C. Li, and X. Su, "Emerging impact of the long noncoding RNA MIR22HG on proliferation and apoptosis in multiple human cancers," *Journal of Experimental & Clinical Cancer Research*, vol. 39, no. 1, 2020.

[144] A. Al-Rugeebah, M. Alanazi, and N. R. Parine, "MEG3: an oncogenic long non-coding RNA in different cancers," *Pathology Oncology Research*, vol. 25, no. 3, pp. 859–874, 2019.

[145] Y. Wu, L. Zhang, S. He et al., "Identification of immune-related LncRNA for predicting prognosis and immunotherapeutic response in bladder cancer," *Aging*, vol. 12, no. 22, pp. 23306–23325, 2020.

[146] F. Zhang, J. Li, H. Xiao, Y. Zou, Y. Liu, and W. Huang, "AFAP1-AS1: a novel oncogenic long non-coding RNA in human cancers," *Cell Proliferation*, vol. 51, no. 1, 2018.

[147] H. Fan, Y. Ge, X. Ma et al., "Long non-coding RNA CCDC144NL-AS1 sponges miR-143-3p and regulates MAP3K7 by acting as a competing endogenous RNA in gastric cancer," *Cell Death & Disease*, vol. 11, no. 7, 2020.

[148] J. He, J. Guan, S. Liao et al., "Long noncoding RNA CCDC144NL-AS1 promotes the oncogenicity of osteosarcoma by acting as a molecular sponge for microRNA-490-3p and thereby increasing HMGA2 expression," *Oncotargets and Therapy*, vol. 14, pp. 1–13, 2021.

[149] Y. Zhang, H. Zhang, and S. Wu, "LncRNA-CCDC144NL-AS1 promotes the development of hepatocellular carcinoma by inducing WDR5 expression via sponging miR-940," *Journal of Hepatocellular Carcinoma*, vol. 8, pp. 333–348, 2021.

[150] L. Zhang, B. Chi, J. Chai et al., "LncRNA CCDC144NL-AS1 serves as a prognosis biomarker for non-small cell lung cancer and promotes cellular function by targeting miR-490-3p," *Molecular Biotechnology*, vol. 63, no. 10, pp. 933–940, 2021.

[151] J.-J. Qiu, Y. Y. Lin, X. Y. Tang, Y. Ding, X. F. Yi, and K. Q. Hua, "Extracellular vesicle-mediated transfer of the lncRNA-TC0101441 promotes endometriosis migration/invasion," *Experimental Cell Research*, vol. 388, no. 1, 2020.

[152] W. Wang and J. Wu, "Identification of long noncoding RNA TC0101441 as a novel biomarker for diagnosis and prognosis of gastric cancer," *International Journal of Clinical and Experimental Pathology*, vol. 14, no. 3, pp. 363–368, 2021.

[153] H. Huang, Z. Zhu, and Y. Song, "Downregulation of lncrna uca1 as a diagnostic and prognostic biomarker for ovarian endometriosis," *Revista Da Associacao Medica Brasileira*, vol. 65, no. 3, pp. 336–341, 1992.

[154] F. Yao, Q. Wang, and Q. Wu, "The prognostic value and mechanisms of lncRNA UCA1 in human cancer," *Cancer Management and Research*, vol. 11, pp. 7685–7696, 2019.

[155] S. Ghazal, B. McKinnon, J. Zhou et al., "H19 lncRNAalters stromal cell growth viaIGFsignaling in the endometrium of women with endometriosis," *EMBO Molecular Medicine*, vol. 7, no. 8, pp. 996–1003, 2015.

[156] Z. Xu, L. Zhang, Q. Yu, Y. Zhang, L. Yan, and Z. J. Chen, "The estrogen-regulated lncRNA H19/miR-216a-5p axis alters stromal cell invasion and migration via ACTA2 in endometriosis," *Molecular Human Reproduction*, vol. 25, no. 9, pp. 550–561, 2019.

[157] J.-J. Qiu, X. J. Lin, T. T. Zheng, X. Y. Tang, Y. Zhang, and K. Q. Hua, "The exosomal long noncoding RNA aHIF is upregulated in serum from patients with endometriosis and promotes angiogenesis in endometriosis," *Reproductive Sciences*, vol. 26, no. 12, pp. 1590–1602, 2019.

[158] K. Liao, X. Ma, B. Chen et al., "Upregulated AHIF-mediated radioresistance in glioblastoma," *Biochemical and Biophysical Research Communications*, vol. 509, no. 2, pp. 617–623, 2019.

[159] X. Tang, S. Liu, Y. Liu et al., "Circulating serum exosomal aHIF is a novel prognostic predictor for epithelial ovarian cancer," *Oncotargets and Therapy*, vol. 12, pp. 7699–7711, 2019.

[160] X. Gao, Y. Lai, Z. Zhang et al., "Long non-coding RNA RP11-480I12.5 promotes the proliferation, migration, and invasion of breast cancer cells through the miR-490-3p-AURKA-Wnt/β-catenin axis," *Frontiers in Oncology*, vol. 10, 2020.

[161] W. Lou, B. Ding, G. Zhong, J. Yao, W. Fan, and P. Fu, "RP11-480I12.5-004 promotes growth and tumorigenesis of breast cancer by relieving miR-29c-3p-mediated AKT3 and CDK6

degradation," *Molecular Therapy - Nucleic Acids*, vol. 21, pp. 916–931, 2020.

[162] S. Ghafouri-Fard, T. Khoshbakht, M. Taheri, and K. Ebrahimzadeh, "A review on the carcinogenic roles of DSCAM-AS1," *Frontiers in Cell and Development Biology*, vol. 9, 2021.

[163] Y. Wu, H. Lyu, H. Liu, X. Shi, Y. Song, and B. Liu, "Downregulation of the long noncoding RNA GAS5-AS1 contributes to tumor metastasis in non-small cell lung cancer," *Scientific Reports*, vol. 6, no. 1, 2016.

[164] Y. Wang, W. Jing, W. Ma, C. Liang, H. Chai, and J. Tu, "Down-regulation of long non-coding RNA GAS5-AS1 and its prognostic and diagnostic significance in hepatocellular carcinoma," *Cancer Biomarkers*, vol. 22, no. 2, pp. 227–236, 2018.

[165] D. Lu, P. Luo, Q. Wang, Y. Ye, and B. Wang, "lncRNA PVT1 in cancer: a review and meta-analysis," *Clinica Chimica Acta*, vol. 474, pp. 1–7, 2017.

[166] G. Liu, S. Liu, G. Xing, and F. Wang, "lncRNA PVT1/microRNA-17-5p/PTEN axis regulates secretion of E2 and P4, proliferation, and apoptosis of ovarian granulosa cells in PCOS," *Molecular Therapy-Nucleic Acids*, vol. 20, pp. 205–216, 2020.

[167] Q. Han, W. Zhang, J. Meng, L. Ma, and A. Li, "LncRNA-LET inhibits cell viability, migration and EMT while induces apoptosis by up-regulation of TIMP2 in human granulosa-like tumor cell line KGN," *Biomedicine & Pharmacotherapy*, vol. 100, pp. 250–256, 2018.

[168] F. Yang, X. S. Huo, S. X. Yuan et al., "Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis," *Molecular Cell*, vol. 49, no. 6, pp. 1083–1096, 2013.

[169] F. Hou, J. Li, J. Peng, Z. Teng, J. Feng, and W. Xia, "LncRNA TMPO-AS1 suppresses the maturation of miR-335-5p to participate in polycystic ovary syndrome," *Journal of Ovarian Research*, vol. 14, no. 1, 2021.

[170] Q. Zheng, J. Jia, Z. Zhou, Q. Chu, W. Lian, and Z. Chen, "The emerging role of thymopoietin-antisense RNA 1 as long noncoding RNA in the pathogenesis of human cancers," *DNA and Cell Biology*, vol. 40, no. 7, pp. 848–857, 2021.

[171] J. Zhen, J. Li, X. Li et al., "Downregulating lncRNA NEAT1 induces proliferation and represses apoptosis of ovarian granulosa cells in polycystic ovary syndrome via microRNA-381/IGF1 axis," *Journal of Biomedical Science*, vol. 28, no. 1, 2021.

[172] H. Gao, J. Jiang, Y. Shi, J. Chen, L. Zhao, and C. Wang, "The LINC00477/miR-128 axis promotes the progression of polycystic ovary syndrome by regulating ovarian granulosa cell proliferation and apoptosis," *Reproductive Biology and Endocrinology*, vol. 19, no. 1, p. 29, 2021.

[173] H. Zhao, Y. He, H. Li et al., "The opposite role of alternatively spliced isoforms of LINC00477 in gastric cancer," *Cancer Management and Research*, vol. 11, pp. 4569–4576, 2019.

Hindawi
Stem Cells International
Volume 2021, Article ID 1624669, 17 pages
https://doi.org/10.1155/2021/1624669

Review Article

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

Yixuan Low,[1] **Dennis Eng Kiat Tan**,[1] **Zhenhua Hu**,[1] **Shawn Ying Xuan Tan**,[1] **and Wee-Wei Tee**[1,2]

[1]*Chromatin Dynamics and Disease Epigenetics Group, Institute of Molecular and Cell Biology, Agency for Science, Technology, And Research (A*STAR), Singapore 138673*
[2]*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

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

Academic Editor: Xinyi Lu

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

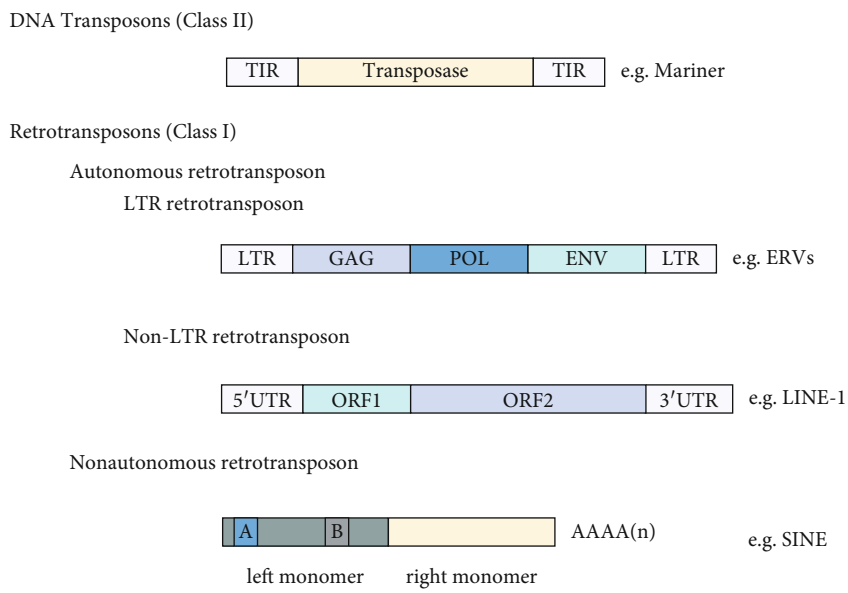


Figure 1: Schematic representations of key TE classes present in mammalian genomes. TEs can be broadly classified into DNA transposons (class II) and retrotransposons (class I). For a specific breakdown of each order, please refer to the review by [6]. Abbreviations: GAG: capsid protein; ENV: envelope 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

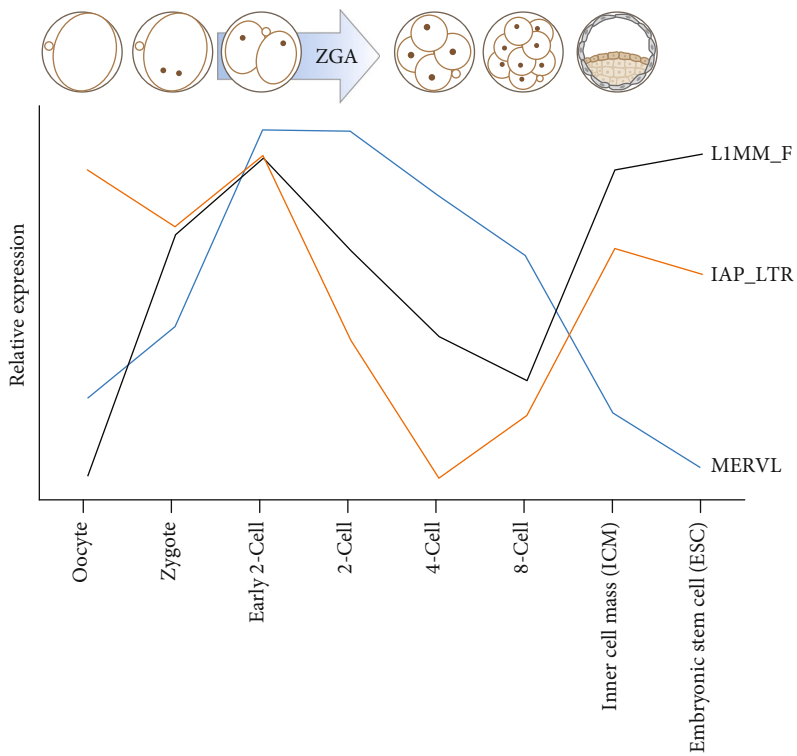


Figure 2: Graphical representation of the dynamics of LINE1, MuERVL, and IAP transcript levels during mouse preimplantation development and in mESCs. The relative expression levels of the TEs plotted are based on the analysis of published RNAseq datasets [45].

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 (MuERVL) 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, MuERVL 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 MuERVL 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 MuERVL 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 MuERVL) 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 MuERVL 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 MuERVL elements in mice contributes to the failure of the chimeric transcript expression and a decrease in the GAG protein content [92]. Moreover, even a modest reduction of MuERVL transcripts was sufficient to cause developmental delays, thereby implicating MuERVL in the regulation of early developmental programs [92]. However, it remains unclear whether the resultant phenotype is a direct consequence of the loss of MuERVL transcripts or the depletion of chimeric transcripts. The identification of the exact contributor to developmental abnormalities in MuERVL-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 MuERVL 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, MuERVL elements are also involved in shaping the 3D genome during development, as evidenced by Hi-C analysis of MT2_mm and canonical MuERVL, 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 MuERVL 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 MuERVL 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 MuERVL-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 MuERVL 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, MuERVL-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 MuERVL, indicating that SETDB1-mediated H3K9me3 is likely not responsible for the silencing of MuERVL elements in mESCs. Instead, MuERVL elements are enriched for G9a-dependent H3K9me2 [47], and catalytically active G9a is required for silencing MuERVL 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 MuERVL [103]. Intriguingly, H3K9me3 is less enriched on MuERVL elements in 2-cell embryos, suggesting that SETDB1 may have an indirect role in repressing MuERVL elements [104]. Post-ZGA, MuERVL- 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 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_mm initiated chimeric transcripts are both robustly expressed in 2C-like mESCs [45, 114, 116]. This suggests that the MT2_mm expression in 2C-like mESCs could potentiate its own repression by promoting the expression

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
Transcription factors		
DUX	[63]	[56, 63, 64, 164]
p53	-	[64]
DPPA2/DPPA4	-	[77, 78]
NELFA	-	[65]
GATA2	[80]	[80]
ZSCAN4	-	[57, 59, 77, 186]
Posttranscriptional regulators		
miR-344	-	[71]
miR-34a	-	[80]
Posttranslational modifiers		
SUMO2	-	[161]
PIAS4	[79]	[79]
Chromatin-associated regulators		
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]

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 $H = A, C$, 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 α-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^6-adenosine methyltransferase (METTL3) and YT521-B homology domain C1 (YTHDC1) m^6A mRNA reader modulate 2C-like transitions in mESCs [72]. Mechanistically, YTHDC1 binds to m^6A-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. Vandewege, and D. A. Ray, "Mammalian transposable elements and their impacts on genome evolution," *Chromosom Resolution*, 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. Worrad, 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 crocodilian 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 MERVL/HERVL 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^6A 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 opposingly 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-34aexpands 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. Füchtbauer 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_ $^{-/-}$ 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 HP1γ/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., "Zscan 4c activates endogenous retrovirus MERVL 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.

Hindawi
Stem Cells International
Volume 2021, Article ID 3829286, 13 pages
https://doi.org/10.1155/2021/3829286

Review Article

Applications of *piggyBac* Transposons for Genome Manipulation in Stem Cells

Yi Sun,[1,2] **Guang Liu**,[1,2] **and Yue Huang**[1,2]

[1]*State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China*
[2]*Department of Medical Genetics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100005, China*

Correspondence should be addressed to Guang Liu; liuguang@ibms.pumc.edu.cn and Yue Huang; huangyue@pumc.edu.cn

Received 21 May 2021; Accepted 16 August 2021; Published 16 September 2021

Academic Editor: Xinyi Lu

Transposons are mobile genetic elements in the genome. The *piggyBac* (PB) transposon system is increasingly being used for stem cell research due to its high transposition efficiency and seamless excision capacity. Over the past few decades, forward genetic screens based on PB transposons have been successfully established to identify genes associated with drug resistance and stem cell-related characteristics. Moreover, PB transposon is regarded as a promising gene therapy vector and has been used in some clinically relevant stem cells. Here, we review the recent progress on the basic biology of PB, highlight its applications in current stem cell research, and discuss its advantages and challenges.

1. Introduction

In 1981, Evans and Martin isolated and established separately undifferentiated embryonic stem cell (ESC) lines from the inner cell mass (ICM) of mouse blastocysts [1, 2]. Subsequently, human ESCs (hESCs) were successfully isolated by Thomson et al. [3] in 1998, and hESCs provide unparalleled tools for studying human embryonic development and regenerative medicine [4]. Additionally, mouse-induced pluripotent stem cells (miPSCs) [5] and human iPSCs (hiPSCs) were generated in 2006 and 2007, respectively [6, 7]. Two key features of ESCs and iPSCs are self-renewal, the ability to proliferate indefinitely and pluripotency and the ability to differentiate into various tissue cell types under appropriate culture conditions. As major types of pluripotent stem cells (PSCs), ESCs and iPSCs provide powerful tools to study the gene function. In particular, hiPSCs hold great promise for generating patient-specific human PSCs (hPSCs) for disease modeling and drug discoveries [8]. In addition to PSCs, other types of stem cells derived from postnatal animal tissues are widely used, such as mesenchymal stem cells (MSCs) [9], hematopoietic stem cells (HSCs) [10], and spermatogonial stem cells (SSCs) [11]. Over the past few decades, stem cell biology and mammalian functional genetics studies have developed closely together, yielding remarkable potentials for the application of regenerative medicine to drug discovery, disease modeling, and the development of novel therapeutic strategies [12].

DNA transposons are mobile genetic elements that can move throughout the genome via a "cut-and-paste" mechanism called transposition, and they are usually inactive in rodents and human cells in nature [13]. Over the past two decades, a series of active recombinant transposons have been generated and used as novel tools for functional genomics research in mice and other vertebrates [14]. Among them, *Sleeping Beauty* (SB) and *piggyBac* (PB) are the most commonly used eukaryotic DNA transposons [15, 16]. SB, a Tc1-like transposable element isolated from the salmonid fish genome, was the first transposon used in mouse and human cells [17, 18]. Although SB can function effectively in mouse somatic cells, it is not highly active in ESCs [19]. PB, however, which is derived from the cabbage looper moth *Trichoplusia ni*, shows high transposition efficiency in different mammalian cell lines, including ESCs, in addition to

other organisms [20–22]. Subsequent studies have demonstrated that the translocation activity of PB is significantly higher than SB in mammalian cell lines [23]. Moreover, unlike SB, which always leaves a 2–5 bp footprint mutation after mobilization and has a strong tendency for "local hopping (reinsertion close to the original donor site)," PB exhibits the unique and valuable feature of seamless excision after transposition [24–26]. With the help of the PB system, scores of transgenic animals have been successfully generated, including mice [20, 27], rats [28, 29], pigs [30, 31], and goats [32]. Besides, PB has been used as a nonviral vector for insertional mutagenesis [33], genetic screens [34–38], iPSCs engineering [39–41], gene therapies [15, 42–45], and novel CAR-T cell therapeutic strategies [46–49]. In this review, we will look back to the advancements of PB transposon in stem cells and regenerative medicine, and discuss its wide applications, so as to provide a reference for future research.

2. Characteristics of the PB Transposon

2.1. Integration Site Preference. The PB element was originally discovered in insect cells as a repetitive element while propagating baculovirus in the TN-386 cell line as shown by Fraser et al. [50] and isolated by Cary *et al.* in 1989 [51]. The inserted mobile DNA was carried by the virus in the form of a "piggyBack," hence the name *piggyBac*; "*Bac*" stands for it being a baculovirus-related discovery. In 2005, Ding et al. found that PB elements can actively transpose in a variety of human and mouse cell lines, as well as in mouse germline cells [20]. The original PB element is a 2,475 bp fragment within an open reading frame (ORF) that encodes a functional transposase of 594 amino acids, flanked by 311 bp 5′ end and 235 bp 3′ end sequences, each containing asymmetric inverted terminal repeats (ITRs) carrying transposase binding sites (Figure 1(a)). The 35 bp 5′ end ITR (5′PBITR) and 63 bp 3′ end ITR (3′PBITR) were shown to be sufficient for activity both in vivo and in vitro [52]. Importantly, the PB element can be divided into two functional components, ITRs and the PB transposase (PBase), to form a binary transposition system, and have been split into a helper plasmid and a donor plasmid (Figure 1(b)). The PBase, which can be transiently expressed by the helper plasmid, excises any DNA sequence of interest flanked by the ITRs in circular donor plasmid via binding to the ITRs (i.e., cut) and reintegrates the sequence into the TTAA site in the genome (i.e., paste) (Figure 1(c)) [53, 54]. The insertion site can be detected using Splinkerette PCR combined with DNA sequencing [34, 55]. The further advantage is seamless excision that the reexpression of PBase can remove the transposon completely to obtain transposon-free cells [56].

Several studies have shown that the distribution of PB transposons has no correlation with gene density or expression level, but rather depends on the distribution pattern of TTAA sites [57] and was negatively influenced by genomic methylation [22, 58]. Theoretically, there is an average of one TTAA site every 256 bp (four to the power of four) in the genome, but the protein-coding regions have a higher GC content compared to other positions, leading to a lower frequency of TTAA sites [59]. In addition, only about 1% of PB insertion sites are located in the 5′ region within 1000 bp upstream of the transcription start site (TSS), which is much lower than the proportion for retrovirus systems [22, 58, 60]. Compared to lentivirus systems [21, 60], PB preferably integrates into genomic safe harbors (GSHs), which are defined based on five criteria for its relative location to ultraconserved regions, noncoding RNAs, and coding genes, especially cancer-related genes [61, 62].

2.2. Mutagenic Cassettes. As discussed above, transposons acted as DNA delivery vehicles for genetic modifications. Several PB-based vectors that have been used for insertional mutagenesis contain two main features. These are (1) mutagenic gene trap cassettes to mediate target gene expression (loss or gain-of-function, LOF, or GOF) and (2) reporter cassettes, whose expression is dependent or independent of the correct splicing between exons of the trapped gene and mutagenic gene trap cassettes [63, 64]. Based on the strategy used for mutating genes, gene trapping can be mainly divided into promoter trapping and polyadenylation (polyA) trapping [65]. In promoter trapping, mutagenic cassettes usually include a splice acceptor (SA) followed by reporter genes and polyA signals in one or both orientations. After integrated into an intron of the expressed gene, the SA-report-polyA element can disrupt the expression of the trapped gene by splicing into upstream exons, which results in a gene trap fusion transcript, and the expression of reporter gene is driven by the endogenous promoter of the trapped gene. As the expression of such a reporter cassette depends on an endogenous promoter, they can only drive transcriptional activation in a tissue of interest [63]. Thus, a reporter driven by an exogenous promoter can be separately used and is independent of the splicing fused transcript, which has allowed more than 90% of mutational coverage of all mouse genes with unbiased distribution throughout the genome [66]. The reporter cassettes used are usually fluorescent proteins (e.g., green fluorescent protein, GFP; red fluorescent protein, and RFP), antibiotic resistance (puromycin, neomycin, hygromycin, etc.), or β-galactosidase.

In polyA trapping, transposon insertions utilize a unidirectional exogenous strong promoter followed with a splice donor (SD), but lacks a polyA signal (Figure 1(d)). If the orientation of the exogenous promoter-SD element is consistent with the direction of the transcription of the trapped gene, the element will be spliced into endogenous, downstream exons, hence initiating gene transcription regardless of transcriptional activity [63]. Some trap cassettes with strong viral enhancers/promoters may result in overexpression of truncated or full-length protein products of the trapped gene. Moreover, the promoter of the trapped gene may be transactivated by strong enhancer elements inside the transposon, leading to the overexpression of a full-length transcript [63]. It is worth noting that vector integrations always tend to occur in the last introns (3′-end most) of the trapped gene in poly-A trapping. By inserting an internal ribosome entry site (IRES) sequence between the

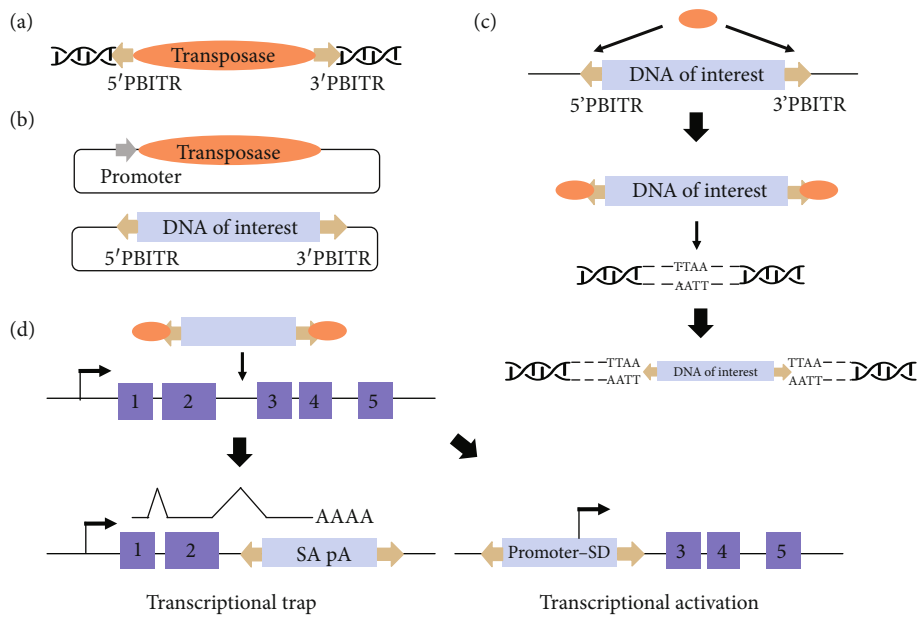


FIGURE 1: The *piggyBac* (*PB*) transposon System. (a) The original PB elements consist of the transposase gene (orange) and inverted terminal repeats (ITRs, yellow arrows). (b) A binary transposition system for gene delivery in plasmids. One is a transposase expression helper plasmid driven by a promoter (grey arrow), and the other is a donor plasmid that contains a DNA sequence of interest (light purple) and/or drug selection marker flanked by ITRs. (c) Vector-to-chromosome transposition. After cotransfection of the binary system in vitro or in vivo, the transposon carrying a transgene is excised from the donor plasmid and then integrated into the chromosome via transposase interaction with host genome sites containing TTAA segments. (d) Insertion of transposons can disrupt or promote gene expression. In the example shown in this figure, a transposon is integrated between exons 2 and 3 (numbered purple), which may lead to two possible outcomes: (1) the transposon hijacks transcription via the splicing receptor-polyadenylate signaling (SA-ployA) element, disrupting gene function and leading to the trap of transcript expression (exons 1–2), and (2) the transposon drives the expression of downstream gene sequences (exons 3–5) via the promoter-splicing donor (SD) element.

reporter cassette and the SD site to prevent nonsense-mediated mRNA decay (NMD) of chimeric transcripts, the bias in the vector integration site can be effectively removed [67]. These features enable the PB system to be a rapid, high-throughput, and traceable mutagenesis tool for constructing mutant libraries for LOF or GOF screening and identification of insertional genes for further validation.

2.3. Cargo Capacity. Genomic sequences contain protein-coding regions and important *cis*-acting regulatory elements (promoters, enhancers, repressors, etc.) that are essential for appropriate spatial–temporal gene expression. Therefore, the capacity to deliver large cargo is critical for achieving successful gene expression regulation. Retroviral and lentiviral vectors' cargo capacity is restricted to about 10 kb and also has immunogenic and tumorigenic potential [68]. Non-viral systems, such as SB transposon, are also limited to 5–6 kb in cargo size and have shown a reduced transposition efficiency when cargo size reaches 10 kb [69]. These characteristics limit the use of selectable markers, inducible cassettes, and large regulatory sequences. However, Li et al. showed that in mESCs, giant PB transposons could mobilize 100 kb DNA fragments to endogenous genomic sites with good cargo integrity, and transposons could be seamlessly excised after transposition [70]. Since the transposition efficiency decreases with increasing cargo size, 100 kb is unlikely to be the upper limit of PB cargo capacity. In general, PB can carry multiple genes during transposition, providing great advantages for multiplexed genetic manipulations, including insertional mutagenesis.

2.4. Transposase. Engineering the PBase is the key to enhancing PB transposition efficiency in mammalian cells. A mouse codon-optimized version of the PBase (mPBase) mediates a 20-fold increase in vector-to-chromosome transposition relative to the original native version [54] and also elevated the rates of chromosomal transposition from PB donor loci in mESCs [71]. The enhanced *PiggyBac* (ePiggyBac) system, which contains a human codon-optimized transposase and the T53C/C136T mutant 5′PBITR, could increase genome integration efficiency by 10-fold in hESCs [72]. Subsequently, a hyperactive PBase (hyPBase), with a total of 7 amino acids (aa) substitutions as shown by Yusa et al., can mediate more efficient transposition and outperformed the mPBase by 10-fold without compromising genomic integrity [73]. An in vivo study reported that the hyPBase had a 20-fold increase in the liver-directed expression compared to mPBase [74]. Moreover, it is well known that the PB transposon can be excised by the reexpression of transposase, but there is still the possibility of transposon jumping into new locations. To solve this problem, Li et al. generated an excision competent/integration defective (Exc^+/Int^-) PBase by amino acids mutation at a catalytic domain [75]. As the integration of the PB transposase vector into the host genome may lead to multiple transposition cycles, scientists have discovered that transfection of PB

transposase mRNA (a short half-life) instead of a plasmid can effectively reduce the potential genetic toxicity [76, 77]. An optimized PB transposon system will significantly expand its application in various fields.

2.5. Comparison/Combination with Other Nucleases. The PBase is a very efficient enzyme that actively integrates DNA fragments into the genome in a random manner [78]. Recently, engineered nucleases, including transcription activator-like effector nucleases (TALENs), zinc finger nucleases (ZFNs), and clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated 9 (Cas9), have been widely used for gene transfer and modification through generating double-strand DNA breaks (DSBs), which can be repaired by homologous directed recombination (HDR) [79]. However, all of these systems exhibit offtarget effects and nonenzymatic DNA insertion [80]. Thus, some works have sought to design PBase fused with these nucleases for integrating DNA into a unique user-defined chromosome site. Although the chimeric TALE-PBase [81] or ZFP-PBase [82] targeting of a unique genomic locus increased transposition efficiency, no targeted transposition was demonstrated [83]. The CRISPR/Cas9 system uses a short guide RNA (sgRNA) to guide the DNA endonuclease Cas9 to a specific target site and facilitates mutation insertion [84]. A specific Cas9 mutant lacking endonuclease activity (dCas9) fused with transcriptional repressor or activation domains has also been generated to promote transcriptional inhibition or activation when coexpressed with targeted sgRNAs [85, 86]. Lena et al. fused dCas9 to PBase and targeted it to specific genomic sites using dual sgRNAs [87]. Thus, the ease of design and application of dCas9-PBase, which can edit genes at precise genomic loci, improves future medical applications.

3. Functional Genomics Using PB-Mediated Genetic Approaches

Identifying genes that are important for specific biological phenotypes and diseases is a crucial goal of genetic analysis, and genetic screens have proven to be one of the most effective approaches [88]. The reverse genetic analyses are hypothesis-driven investigations of a phenotype driven by the disruption of predefined genes [89], while forward genetic screens are phenotype-derived approaches that generally involve high-throughput mutant libraries generation, specific phenotype selection, and mutations validation [90].

3.1. Loss-of-Function Screening. LOF genetic screens using mammalian cell lines are powerful tools for identifying genes required for many cellular processes. Since ESCs can differentiate into a variety of cell types, including germ cells, and have relatively stable genomes amenable to various genetic manipulations, these cells have become attractive models for analyzing developmental events or disease phenotypes in vitro [80]. With the excellent mutagenic ability of PB transposon and less bias towards certain genomic hot spots, a genome-wide mutant library could be rapidly and efficiently established in ESCs, allowing researchers to perform phenotype-based genetic screening in mammalian cells, similar to studies that have been done in yeast for the past 30 years [91]. In combination with high-throughput next-generation sequencing (NGS) technologies [91, 92], hundreds to thousands of genes trapped by PB transposons could be easily identified, enabling the study of the molecular mechanisms of practically any biological process studied (Table 1).

3.1.1. The Problem of Homozygosity Mutations. Since most of phenotypic changes in mammalian cells require both copies of an autosomal gene to be inactivated (except in some cases of haploinsufficiency) [93], the genome-wide LOF screen of recessive mutations is quite time-consuming and rather difficult using diploid cells [94]. This issue was partially solved by generating a Bloom's syndrome gene-(*Blm*-) deficient ESCs, which lead to a higher rate of mitotic recombination between sister chromatids. *Blm*-null ESCs harboring heterozygous mutation conversed to homozygous mutations through a loss-of-heterozygosity (LOH), which occurs at a rate of about 10^{-4} events/locus/cell/division. Thus, a mutant library needs to be expanded for at least 14 population doublings to promote homozygous mutant generation for further LOF screens, such as the resistance to 6-thioguanine (mismatch repair mutants) and retroviral infection [34, 95–97]. Huang et al. used a PB transposon vector, which carried two drug resistance genes but could express only one at a time and *Blm*-null ESCs to isolate homozygous mutant cell clones successfully. The two expressed drug-resistant genes could be switched by Cre recombinase, and this allowed selection for the increase in homozygous mutants that occur after LOH [94]. However, due to the low frequency of LOH in each generation per cell and homozygous cells accounting for only a very small fraction of *Blm*-null cells, it is not easy to achieve a sufficient number of homozygous mutants for genetic screening [91]. Recently, as an encouraging breakthrough in cell biology, haploid ESC (haESC) lines have been generated in several species, including medaka fish [98], mice [99–102], rats [103], monkeys [104], and humans [105–107]. As there is only one set of chromosomes in haploid cells, it becomes quite easy to generate loss-of-function mutations using haESCs, which hold great promise for both forward and reverse genetic screens [38, 91, 99, 108–116].

3.1.2. Stem Cell Characteristic-Related Screening. Due to the infinite self-renewal ability and haploid properties, haESCs have become powerful tools for generating a tremendous number of homozygous mutation pools [117]. The PB transposon system has also been successfully applied to haESCs to identify different mechanisms of stemness and differentiation. Although the mechanisms of self-renewal of PSCs have become clearer, less is known about how these robust pluripotency programs are modulated to enable fate transitions. PB-mediated large-scale libraries in haESCs for the genetic exploration of the exit-from-pluripotency have been reported, and researchers have identified the RNA binding protein Pum1 and the conserved small zinc finger protein Zfp706 as being required for exit from self-renewal state timely and efficiently [108]. In addition, the combination

Table 1: Genetic screening using PB transposon.

	Screening purpose	Cell type	Strain	Genes identified	Reference
Loss of function					
Drug resistance	6-thioguanine	*Blm*-deficient ESCs	Mouse	*Dnmt1*	Guo et al., 2004 (ref [96])
	6-thioguanine	haESCs	Mouse	*Msh2, Hprt*	Leeb et al., 2011 (ref [100])
	6-thioguanine	haESCs	Mouse	Mismatch repair genes *(Msh2, Msh6, Mlh1)*	Pettitt et al., 2013 (ref [124])
	Olaparib (PARP inhibitor)	haESCs	Mouse	*Parp1*	Pettitt et al., 2013 (ref [124])
	Talazoparib (PARP inhibitor)	haESCs	Mouse	*Ewsr1*	Pettitt et al., 2017 (ref [113])
	Doxorubicin	haESCs	Mouse	*Rmi2⊠Pdk4, and Acbd6*	Liu et al., 2017 (ref [91])
	Tetrodotoxin-like toxicant	haNPCs	Rhesus monkey	*B4GALT6*	Wang et al., 2018 (ref [38])
	6-thioguanine	haTSCs	Mouse	*Hprt*	Cui et al., 2019 (ref [155])
	Puromycin	haESCs	Mouse	*Lrp6*	Mao et al., 2020 (ref [126])
Stem cell-related characteristics	Exit from self-renewal	haESCs	Mouse	*Zfp706, Pum1*	Leeb et al., 2014 (ref [108])
	Exit-from-pluripotency	haESCs	Mouse	*Garnl3, Ifltd1, Sema5a, Cdk5rap2* and *Phf21a*	Liu et al., 2017 (ref [91])
	Spongiotrophoblast specification	haiTSCs	Mouse	*Htra1*	Peng et al., 2019 (ref [118])
	Reprogramming factors	haEpiSCs	Mouse	*Hs3st3b1*	Gao et al., 2021 (ref [122])
	Haploidy maintenance	haESCs	Mouse	*Etl4*	Zhang et al., 2020 (ref [123])
Gain of function					
Drug resistance	PluriSIn-1(SCD1 inhibitor)	ESC	Human	RAS pathway genes	Weissbein et al., [2019] (ref [127])
Stem cell-related characteristics	Ground state pluripotency	EpiSCs	Mouse	*Nr5a*	Guo et al., 2010 (ref [35])
	Cell differentiation	ESC	Human	*RHOA*	Gayle et al., 2015 (ref [156])
	Growth advantage	ESC	Human	RHO-ROCK pathway genes	Weissbein et al., 2019 (ref [127])
	Teratoma formation	ESC	Human	PI3K-AKT and HIPPO pathways genes	Weissbein et al., 2019 (ref [127])

of PB with newly established haploid stem cell lines from other cell types also plays a vital role in the study of lineage-specific functional genomics. Recently, Cui et al. generated mouse parthenogenetic haploid trophoblast stem cells (haTSCs), which can also serve as a powerful tool for forward genetic screens in placental biology and disorders [117]. In another study, Peng et al. obtained haploid-induced trophoblast stem cells (haiTSCs) from *p53*-deficient haESCs by overexpressing the *Cdx2* gene in vitro. PB-mediated high-throughput mutation in haiTSCs was performed and used to screen factors related to the trophoblast lineage, and then *Htra1* was validated as a blocker of spongiotrophoblast specification [118].

Mouse epiblast stem cells (EpiSCs) are derived from the postimplantation egg cylinder epiblast. Unlike ESCs in a naïve pluripotent state, EpiSCs are in a primed pluripotent state and have been widely used to explore the intricate mechanisms of reprogramming [119–121]. Recently, Gao et al. established haploid EpiSCs (haEpiSCs) from mouse postimplantation epiblast at embryonic day 6.5 (E6.5) by microinjecting *p53*-knockout haESCs into normal blastocysts. Through a massive PB-mediated mutagenesis protocol, researchers determined *Hs3st3b1* as a key modulator that may impede the reprogramming process, providing a valuable resource for reprogramming research [122].

Although haploid stem cells have many advantages in genetic screening, the haploid state is generally unstable in culture. As haESCs tend to become diploids spontaneously, it hampers their application in functional genomic researches [99, 100]. In a recent study, we used a genome-wide haESC homozygous mutant library based on PB transposon mutagenesis to screen the potential haploidy-maintenance factors and found that *Etl4*-deficiency reduced the rate of self-diploidization in haESCs. This gene was found to be linked to an energy metabolism transition, thus providing a novel strategy for maintaining haploid status during cell culture by regulating cell metabolism [123].

3.1.3. Drug Resistance Screening. Pettitt et al. used a PB transposon-based dual-directional gene trap vector and mouse haESCs to generate large-scale gene mutant libraries. The resistance to olaparib, a clinical poly (ADP-ribose) polymerase (PARP) inhibitor, was screened, and it was determined that the toxicity of olaparib in normal cells was mainly mediated by PARP1 [124]. At present, almost all reported genetic screens based on mixed mutant pools must rely on strong positive selections of resistant clones, and "negative selection"-based screens are not easy to conduct using these mixed pools due to the possible interference and interplay among different mutant cells, which can interfere with the readout of the quantitative deep sequencing of such a screen [125]. Therefore, we generated arrayed haploid mutant libraries with up to 85% homozygous mutant clones and then conducted a negative screen to discover mutations conferring sensitivity to the DNA-damaging drug doxorubicin, an anticancer drug frequently used in clinic [91]. Recently, Mao et al. developed an inducible self-inactivating PB system (named "One-Shot") that allows rapid construction of a mutant library in mouse haESCs and haploid neural stem cell-like cells (haNSCLCs) with single-copy mutation site per cell and puromycin-related resistance was chosen to evaluate this system [126]. Through PB transposons, high-throughput trap mutations can be effectively integrated into haploid neural progenitor cells (haNPCs), which can remain haploid and maintain the potential to differentiate into neurons and glia for long periods in vitro. The target genes of a tetrodotoxin-like toxicant A803467 (B4GALT6) were uncovered subsequently using such a strategy [38]. These studies have expanded the scope of genetic screens in mammalian cells.

3.2. Gain-of-Function Screening. In addition to LOH screens, forward genetic analysis using PB-based GOF mutagenesis enables researchers to more fully explore various biological processes functionally. Since the genetic changes acquired during the culture of hPSCs may influence their availability for research and future treatments, Weissbein et al. used a PB transposon vector that contained the cytomegalovirus (CMV) enhancer and promoter sequences followed by the SD from the rabbit beta-globin intron, to construct genome-wide libraries of hPSCs. After screening, they uncovered that the overexpression of the RAS pathway led to resistance to the hPSC-specific drug PluriSIn-1, and inactivation of the RHO-ROCK pathway resulted in a growth advantage in culture adaptation [127].

As discussed earlier, EpiSCs may be the barrier in somatic cell reprogramming. Therefore, Guo et al. performed a genome-wide PB insertional activation screen in EpiSCs to identify the factors that can overcome the impediment between EpiSCs and iPSCs [35]. The gene-trap activation vector contained a murine stem cell virus (MSCV) long terminal repeat (LTR) with an SD site from exon 1 of mouse *Foxf2*, which could promote full or truncated protein expression when integrated upstream or within a gene [128]. To date, GOF screening using transposons has been relatively rare, and this has usually been in combination with a LOF to form a bifunctional activating and inactivating transposon system. For example, transgenic mice with these bifunctional activating and inactivating transposons, which carry different promoter/enhancer elements and bidirectional SA with SV40 polyA signals, have been used for the discovery of oncogenes and tumor suppressor genes [27].

3.3. Comparison with Other Screening Systems. Other forward genetic screening methods in functional genomics research include cDNA libraries, RNA interference (RNAi) libraries, and libraries using the CRISPR/Cas9 system for GOF or LOF screens [129]. Compared with transposon-induced mutagenesis, these methods each have distinct advantages and disadvantages, and the combination of different methods can provide complementary techniques for uncovering functional genes (Table 2).

4. PB in Stem Cell-Based Preclinical Studies

Stem cells, as ideal targets for gene therapy, require effective tools for the transient or permanent transfer of genetic information into eukaryotic genomes. Through transposon-based genetic manipulation, therapeutic genes can be introduced with stable phenotypic correction, and stem cells edited can be expanded in vitro, followed by differentiation into particular cell lineages for specific therapeutic needs. Currently, there is widespread evidence that robust transposon-mediated gene transfer can be achieved in several clinically relevant stem cell types, such as hESCs, iPSCs, HSCs, MSCs, or myoblasts.

4.1. hPSCs. Over the past two decades, culture conditions have been a major focus for hPSC research [3, 130]. Recently, extended or expanded pluripotent stem cells (EPSCs) have been reported to have the additional ability to contribute to both embryonic and extraembryonic tissues [131–133]. It was pointed out that Gao et al. generated doxycycline-(Dox-) dependent porcine iPSCs via stable genomic integration of complementary DNA (Yamanaka factors *OCT4, MYC, SOX2*, and *KLF4* together with *LIN28, NANOG, LRH1*, and *RARG*) in porcine fetal fibroblasts (PFFs) using PB transposition. Under similar conditions, hESCs and hiPSCs can be transformed into EPSCs [132, 134]. The successful generation of EPSCs provides tools for embryogenesis and transformation research in regenerative medicine. In addition, hPSCs can also be used for the construction of disease models and therapeutic applications. COVID-19, caused

Table 2: Comparison of genome-wide screening libraries based on cDNA, RNAi, CRISPR/Cas9, and PB transposons.

	cDNA library	RNAi library	CRISPR/Cas9 library	*piggyBac* library
Work mode	Gain of function	Loss of function	Loss of function/gain of function	Loss of function/gain of function
Vehicle	cDNA	Sh/siRNA	sgRNA	*piggyBac* transposon plasmids
Targeting restrictions	Part of transcripts	Only targets mRNA	Protospacer adjacent motif (PAM) must be present	Only at TTAA site
Mutagenesis efficiency	≥2 standard deviations induced expression signals	≥70% gene knockdown	Knockout, knockdown, or overexpress with different kinds of libraries achievable	Activation levels of genes variable, inactivation achievable in haploidy
Genome coverage	Depend on library design	Depend on library design	Depend on library design	Genome-wide in principle, but influenced by integration site preference
Types of mutations	Overexpression	Knockdown	Chromosomal deletions and translocations	Gene activation and inactivation are due to transposon insertion
Reversibility	Potentially reversible	Potentially reversible	Knockdown or overexpressed libraries reversible, knockout libraries irreversible	Reversible
Limitations	Abundance of transcripts varies	High offtarget effects	Less offtarget effects	Biallelic gene inactivation rare in diploid cells; integrations for activation need to be upstream of the transcription start site
Cytotoxicity	Variable to high	Variable to high	Low	Low

by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been declared a global pandemic by the World Health Organization. In addition to respiratory failure, COVID-19 can cause clinical complications in other systems, including metabolism, the heart, the nervous system, and the gastrointestinal tract [135]. An hESC line WAe001-A-58 was generated by PB transposon vector, which carried the Tet-On gene expression system of the SARS-CoV-2 nucleocapsid (N) protein-coding sequence, from the hESC line WA01 (H1), providing an ideal platform for further elucidating the pathological role of the N protein [136].

4.2. iPSCs. Initially, Dr. Shinya Yamanaka and his colleagues expressed four genes (encoding transcription factors Oct4, Sox2, Klf4, and c-Myc) in somatic cells using retroviruses, and these somatic cells were reprogrammed into an embryonic-like state with similar developmental capabilities [5]. However, due to safety concerns, permanent insertion of the virus in the genome may limit the clinical applications of iPSCs. Using the PB transposon system, mouse and human iPSCs have been successfully generated, and reprogramming factors can be removed from these pluripotent cells without any traces via the reexpression of PBase [137, 138], thus minimizing potential concerns associated with insertional oncogenesis. To better control copy numbers in the genome, all four reprogramming factors can be introduced into one vector using approximately 20 amino acid long self-cleaving 2A peptides to separate these different genes [137, 138].

In recent years, PB transposons, combined with TALENs or the CRISPR/Cas9 system, have been used for the genome editing of iPSCs to correct gene defects [25, 26, 139, 140]. Genome editing relies on the introduction of double-strand breaks at target sites using "nucleases" to allow the occurrences of error-prone nonhomologous end-joining (NHEJ) or HDR near the nuclease cutting site, followed by the traceless removal of selectable gene fragments via PBase [24]. This strategy has recently been used to achieve the correction of mutations in the hemoglobin beta chain gene. By combining PB with TALENs or CRISPR/Cas9, the mutated β-globin gene in sickle cell disease- (SCD-) specific iPSCs or β-thalassemia patient-derived iPSCs was successfully seamlessly corrected without any detectible offtarget or adverse chromosomal alterations [25, 26]. Similarly, it has recently been shown in iPSCs derived from patients with Huntington's disease that the combination of PB transposon with the CRISPR/Cas9 system may support gene therapy in these genetic disorders induced by trinucleotide repeat expansion [141]. Corrected stem cells successfully differentiated into excitable, synaptically active forebrain neurons.

Genetic manipulation of iPSCs before transplantation may further threaten genomic stability, which can affect their differentiation, characterization, tumorigenicity, and uncontrolled cellular behavior [142]. Therefore, whole-genome sequencing is needed to detect such changes [143], and more preclinical trials in mice and other animal models will be necessary to further confirm the therapeutic potential of reprogrammed cells in vivo [144].

4.3. HSCs. HSCs are ideal tools for gene therapy in hematologic diseases due to their ability for self-renew and differentiation into different lymphohematopoietic lineages. The PB

transposon system has been used for stable gene transfer of CD34+ HSCs; although, comparative analysis has shown higher activity of SB100X, the most hyperactive version of the SB transposase currently [145, 146]. PB transposon-modified HSCs continue to express functional globin chain proteins, exhibiting a reduced sickle phenotype and an improvement in disease progression. Later, the hyPBase, which is more active than SB100X in other cell types, has been developed [147], but the comparison with SB100X in HSCs has not been performed yet.

4.4. Mesenchymal/Stromal Stem Cells. Human MSCs originate from human embryonic mesoderm and/or can be isolated from fetal and adult tissues, such as bone marrow (BM), umbilical cord (UC), adipose, etc. [9], and are a heterogeneous subset of nonhematopoietic multipotent stromal stem cells. MSCs can be differentiated into ectodermal (e.g., neuronal cells), mesodermal (e.g., osteocytes, chondrocytes and adipocytes), and endodermal lineages (e.g., hepatocytes). It has been reported that the PB system was applicable to gene integration in MSCs [148]. Yang et al. generated immortalized human UC-derived MSCs (iUC-MSCs) using the PB-based monkey virus 40 T antigen (SV40T) system. These cells positively expressed MSC markers and did not induce tumorigenesis in vivo with the retained potential for trilineage differentiation after BMP9 stimulation, which has laid a foundation for further study and applications in UC-MSCs [149]. Moreover, MSCs are considered excellent cancer therapeutic tools in view of their unique ability to target tumor cells. Interferon-gamma-(IFN-γ-) expressing adipose-derived MSCs (AD-MSCs) generated by PB-mediated gene transfer were engrafted into tumor stroma in a mouse model of melanoma and could inhibit tumor growth and angiogenesis, prolong the survival of mice and exhibit an important implication for future cancer treatment [150].

4.5. Myoblasts. Myoblasts are self-renewing adult muscle progenitor cells that can eventually differentiate into skeletal muscle fibers for the potential treatment of muscle disorders. Reports have shown that PB-mediated gene transfer can be used to deliver therapeutic genes into myoblasts effectively. Based on the use of the PB transposon system, the genes encoding either full-length human dystrophin or truncated microdystrophins could be successfully introduced into myoblasts and expressed in differentiated multinucleated myotubules [151], paving the way toward a PB-mediated gene therapy approach for Duchenne muscular dystrophy (DMD).

4.6. Safety Issues of piggyBac. DNA transposons allow nonviral stable gene transfer and potentially replace the need for viral vectors, but there are still a few safety issues to consider carefully.

(i) Insertional mutagenesis is one of the major concerns of any integration-based gene therapy. Since PB transposons exhibit a higher integration preference for transcriptional units, they may potentially lead to the activation of oncogenes or the disruption of tumor suppressor genes, thus promoting malignant transformation. It has been reported that no growth advantage was observed in PB-modified primary human cells during a 140-day experiment [152], and no observable tumor formation was found in livers of wild-type mice modified with PB for one year [153]. In addition, transposon integrations can be redirected to a demonstrated safe-harbor site, which can be achieved by transposase modified to carry site-specific DNA binding domains at its N- or C-termini [82]

(ii) Investigators have found that the probability of plasmid backbone DNA integration is relatively high in PB-modified human embryonic kidney (HEK-293) cells [152]. Although this problem can be nearly eliminated by flow cytometry to sort cells based on characteristic elements in the integrated backbone, it still requires close attention. Besides, it remains to be further explored if plasmid backbone integration exists in clinically relevant cells and animal models

(iii) Previous studies [51] have reported that the 5′ PBITR has potential promoter activity. To avoid possible influence, gene-trap cassettes could be placed opposite the 5′ PBITR for chromatin integration [52]

(iv) Despite the widespread assumption that nonviral vectors should not elicit any immune response, foreign DNA itself has the potential to activate the innate immune system [154]. Thus, some immune regulation may still be needed after stable transposition into the host genome

5. Conclusions and Perspectives

Transposon-based technologies hold great promise for the development of powerful genomic tools. There is no doubt that there will be more reports in the future using PB for gene delivery in stem cells and other fields of research. By combining transposon technology with accurate gene editing techniques, the continued development, refinement, and clinical transformation using PB may herald an exciting and promising new era of gene therapy.

Data Availability

No data is available.

Conflicts of Interest

The authors declare that they have no conflict of interest.

Authors' Contributions

Y. H. conceived and designed the study. Y. S. and G. L. were engaged in data collection, organization, and analyses. The manuscript was written and revised by Y. S., G. L., and Y. H.

Acknowledgments

We are grateful to Meili Zhang for helpful discussion and comments on the manuscript. This work was supported by the National Key Research and Development Program of China (2016YFA0100103 to Y.H.), CAMS Innovation Fund for Medical Sciences (2016-I2M-3-002 to Y.H. and G.L.), and National Natural Science Foundation of China (31970813 to Y.H.).

References

[1] M. J. Evans and M. H. Kaufman, "Establishment in culture of pluripotential cells from mouse embryos," *Nature*, vol. 292, no. 5819, pp. 154–156, 1981.

[2] G. R. Martin, "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 78, no. 12, pp. 7634–7638, 1981.

[3] J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro et al., "Embryonic stem cell lines derived from human blastocysts," *Science*, vol. 282, no. 5391, pp. 1145–1147, 1998.

[4] B. E. Reubinoff, M. F. Pera, C. Y. Fong, A. Trounson, and A. Bongso, "Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro," *Nature Biotechnology*, vol. 18, no. 4, pp. 399–404, 2000.

[5] K. Takahashi and S. Yamanaka, "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors," *Cell*, vol. 126, no. 4, pp. 663–676, 2006.

[6] J. Yu, M. A. Vodyanik, K. Smuga-Otto et al., "Induced pluripotent stem cell lines derived from human somatic cells," *Science*, vol. 318, no. 5858, pp. 1917–1920, 2007.

[7] K. Takahashi, K. Tanabe, M. Ohnuki et al., "Induction of pluripotent stem cells from adult human fibroblasts by defined factors," *Cell*, vol. 131, no. 5, pp. 861–872, 2007.

[8] R. G. Rowe and G. Q. Daley, "Induced pluripotent stem cells in disease modelling and drug discovery," *Nature Reviews Genetics*, vol. 20, no. 7, pp. 377–388, 2019.

[9] P. Bianco, ""Mesenchymal" stem cells," *Annual Review of Cell and Developmental Biology*, vol. 30, no. 1, pp. 677–704, 2014.

[10] E. Dzierzak and A. Bigas, "Blood development: hematopoietic stem cell dependence and independence," *Cell Stem Cell*, vol. 22, no. 5, pp. 639–651, 2018.

[11] A. de Los Angeles, F. Ferrari, R. Xi et al., "Hallmarks of pluripotency," *Nature*, vol. 525, no. 7570, pp. 469–478, 2015.

[12] D. Cyranoski, "How human embryonic stem cells sparked a revolution," *Nature*, vol. 555, no. 7697, pp. 428–430, 2018.

[13] R. K. Aziz, M. Breitbart, and R. A. Edwards, "Transposases are the most abundant, most ubiquitous genes in nature," *Nucleic Acids Research*, vol. 38, no. 13, pp. 4207–4217, 2010.

[14] G. Bourque, K. H. Burns, M. Gehring et al., "Ten things you should know about transposable elements," *Genome Biology*, vol. 19, no. 1, p. 199, 2018.

[15] J. Tipanee, Y. C. Chai, T. VandenDriessche, and M. K. Chuah, "Preclinical and clinical advances in transposon-based gene therapy," *Bioscience Reports*, vol. 37, article BSR20160614, 2017.

[16] J. Tipanee, T. VandenDriessche, and M. K. Chuah, "Transposons: moving forward from preclinical studies to clinical trials," *Human Gene Therapy*, vol. 28, no. 11, pp. 1087–1104, 2017.

[17] Z. Ivics, P. B. Hackett, R. H. Plasterk, and Z. Izsvak, "Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells," *Cell*, vol. 91, no. 4, pp. 501–510, 1997.

[18] L. S. Collier, C. M. Carlson, S. Ravimohan, A. J. Dupuy, and D. A. Largaespada, "Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse," *Nature*, vol. 436, no. 7048, pp. 272–276, 2005.

[19] G. Luo, Z. Ivics, Z. Izsvak, and A. Bradley, "Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 18, pp. 10769–10773, 1998.

[20] S. Ding, X. Wu, G. Li, M. Han, Y. Zhuang, and T. Xu, "Efficient Transposition of the piggyBac (PB) Transposon in Mammalian Cells and Mice," *Cell*, vol. 122, no. 3, pp. 473–483, 2005.

[21] M. H. Wilson, C. J. Coates, and A. L. George Jr., "PiggyBac Transposon-mediated Gene Transfer in Human Cells," *Molecular Therapy*, vol. 15, no. 1, pp. 139–145, 2007.

[22] W. Wang, C. Lin, D. Lu et al., "Chromosomal transposition of PiggyBac in mouse embryonic stem cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 27, pp. 9290–9295, 2008.

[23] S. C. Wu, Y. J. Meir, C. J. Coates et al., "piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 41, pp. 15008–15013, 2006.

[24] Q. Chen, W. Luo, R. A. Veach, A. B. Hickman, M. H. Wilson, and F. Dyda, "Structural basis of seamless excision and specific targeting by _piggyBac_ transposase," *Nature Communications*, vol. 11, no. 1, p. 3446, 2020.

[25] N. Sun and H. Zhao, "Seamless correction of the sickle cell disease mutation of the HBB gene in human induced pluripotent stem cells using TALENs," *Biotechnology and Bioengineering*, vol. 111, no. 5, pp. 1048–1053, 2014.

[26] F. Xie, L. Ye, J. C. Chang et al., "Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 andpiggyBac," *Genome Research*, vol. 24, no. 9, pp. 1526–1533, 2014.

[27] R. Rad, L. Rad, W. Wang et al., "PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice," *Science*, vol. 330, no. 6007, pp. 1104–1107, 2010.

[28] T. Li, L. Shuai, J. Mao et al., "Efficient Production of Fluorescent Transgenic Rats using the piggyBac Transposon," *Scientific Reports*, vol. 6, no. 1, article 33225, 2016.

[29] M. G. Jiang, T. Li, C. Feng et al., "Generation of Transgenic Rats through Induced Pluripotent Stem Cells," *The Journal of Biological Chemistry*, vol. 288, no. 38, pp. 27150–27158, 2013.

[30] Z. Wu, Z. Xu, X. Zou et al., "Pig transgenesis by piggyBac transposition in combination with somatic cell nuclear transfer," *Transgenic Research*, vol. 22, no. 6, pp. 1107–1118, 2013.

[31] Z. Li, F. Zeng, F. Meng et al., "Generation of transgenic pigs by cytoplasmic injection of piggyBac transposase-based pmGENIE-3 plasmids," *Biology of Reproduction*, vol. 90, no. 5, p. 93, 2014.

[32] D. P. Bai, M. M. Yang, L. Qu, and Y. L. Chen, "Generation of a transgenic cashmere goat using the piggyBac transposition system," *Theriogenology*, vol. 93, pp. 1–6, 2017.

[33] M. J. Friedrich, I. F. Bronner, P. Liu, A. Bradley, and R. Rad, "PiggyBac Transposon-Based Insertional Mutagenesis in Mice," in *Cancer Driver Genes*, pp. 171–183, 2019.

[34] W. Wang, A. Bradley, and Y. Huang, "A piggyBac transposon-based genome-wide library of insertionally mutated Blm-deficient murine ES cells," *Genome Research*, vol. 19, no. 4, pp. 667–673, 2009.

[35] G. Guo and A. Smith, "A genome-wide screen in EpiSCs identifies Nr5a nuclear receptors as potent inducers of ground state pluripotency," *Development*, vol. 137, no. 19, pp. 3185–3192, 2010.

[36] J. Bai, K. Li, W. Tang et al., "A high-throughput screen for genes essential for PRRSV infection using a piggyBac based system," *Virology*, vol. 531, pp. 19–30, 2019.

[37] J. Weber, J. de la Rosa, C. S. Grove et al., "_PiggyBac_ transposon tools for recessive screening identify B-cell lymphoma drivers in mice," *Nature Communications*, vol. 10, no. 1, p. 1415, 2019.

[38] H. Wang, W. Zhang, J. Yu et al., "Genetic screening and multipotency in rhesus monkey haploid neural progenitor cells," *Development*, vol. 145, article dev160531, 2018.

[39] J. M. Vanslambrouck, L. E. Woodard, N. Suhaimi et al., "Direct reprogramming to human nephron progenitor-like cells using inducible _piggyBac_ transposon expression of _SNAI2-EYA1-SIX1_," *Kidney International*, vol. 95, no. 5, pp. 1153–1166, 2019.

[40] M. A. Park, H. S. Jung, and I. Slukvin, "Genetic engineering of human pluripotent stem cells using PiggyBac transposon system," *Current Protocols in Stem Cell Biology*, vol. 47, no. 1, p. e63, 2018.

[41] M. G. Garone, V. de Turris, A. Soloperto et al., "Conversion of human induced pluripotent stem cells (iPSCs) into functional spinal and cranial motor neurons using PiggyBac vectors," *Journal of Visualized Experiments*, vol. 2019, no. 147, 2019.

[42] L. E. Woodard and M. H. Wilson, "piggyBac-ing models and new therapeutic strategies," *Trends in Biotechnology*, vol. 33, no. 9, pp. 525–533, 2015.

[43] M. Cheng, X. Jin, L. Mu et al., "Combination of the clustered regularly interspaced short palindromic repeats (CRISPR)-associated 9 technique with the piggybac transposon system for mouse in utero electroporation to study cortical development," *Journal of Neuroscience Research*, vol. 94, no. 9, pp. 814–824, 2016.

[44] B. E. Hew, R. Sato, D. Mauro, I. Stoytchev, and J. B. Owens, "RNA-guided piggyBac transposition in human cells," *Synthetic Biology*, vol. 4, no. 1, article ysz018, 2019.

[45] S. Liu, Q. Wang, X. Yu et al., "HIV-1 inhibition in cells with CXCR4 mutant genome created by CRISPR-Cas9 and _piggyBac_ recombinant technologies," *Scientific Reports*, vol. 8, no. 1, p. 8573, 2018.

[46] D. Morita, N. Nishio, S. Saito et al., "Enhanced Expression of Anti-CD19 Chimeric Antigen Receptor in _piggyBac_ Transposon-Engineered T Cells," *Molecular Therapy-Methods & Clinical Development*, vol. 8, pp. 131–140, 2018.

[47] P. Ptackova, J. Musil, M. Stach et al., "A new approach to CAR T-cell gene engineering and cultivation using piggyBac transposon in the presence of IL-4, IL-7 and IL-21," *Cytotherapy*, vol. 20, no. 4, pp. 507–520, 2018.

[48] Z. Zhang, D. Jiang, H. Yang et al., "Modified CAR T cells targeting membrane-proximal epitope of mesothelin enhances the antitumor function against large solid tumor," *Cell Death & Disease*, vol. 10, no. 7, p. 476, 2019.

[49] D. C. Bishop, L. E. Clancy, R. Simms et al., "Development of CAR T-cell lymphoma in two of ten patients effectively treated with piggyBac modified CD19 CAR T-cells," *Blood*, no. article 2021010813, 2021.

[50] M. J. Fraser, G. E. Smith, and M. D. Summers, "Acquisition of Host Cell DNA sequences by baculoviruses: relationship between host DNA insertions and FP mutants of Autographa californica and galleria mellonella nuclear polyhedrosis viruses," *Journal of Virology*, vol. 47, no. 2, pp. 287–300, 1983.

[51] L. C. Cary, M. Goebel, B. G. Corsaro, H.-G. Wang, E. Rosen, and M. J. Fraser, "Transposon mutagenesis of baculoviruses: Analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses," *Virology*, vol. 172, no. 1, pp. 156–169, 1989.

[52] V. Solodushko, V. Bitko, and B. Fouty, "Minimal piggyBac vectors for chromatin integration," *Gene Therapy*, vol. 21, no. 1, pp. 1–9, 2014.

[53] Z. Ivics, M. A. Li, L. Mates et al., "Transposon-mediated genome manipulation in vertebrates," *Nature Methods*, vol. 6, no. 6, pp. 415–422, 2009.

[54] J. Cadinanos and A. Bradley, "Generation of an inducible and optimized piggyBac transposon system," *Nucleic Acids Research*, vol. 35, no. 12, p. e87, 2007.

[55] H. Mikkers, J. Allen, P. Knipscheer et al., "High-throughput retroviral tagging to identify components of specific signaling pathways in cancer," *Nature Genetics*, vol. 32, no. 1, pp. 153–159, 2002.

[56] M. J. Fraser, L. Cary, K. Boonvisudhi, and H. G. Wang, "Assay for movement of Lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA," *Virology*, vol. 211, no. 2, pp. 397–407, 1995.

[57] B. Troyanovsky, V. Bitko, V. Pastukh, B. Fouty, and V. Solodushko, "The functionality of minimal PiggyBac transposons in mammalian cells," *Molecular Therapy - Nucleic Acids*, vol. 5, p. e369, 2016.

[58] D. Yang, R. Liao, Y. Zheng, L. Sun, and T. Xu, "Analysis of PBase binding profile indicates an insertion target selection mechanism dependent on TTAA, but not transcriptional activity," *International Journal of Biological Sciences*, vol. 12, no. 9, pp. 1074–1082, 2016.

[59] K. Yusa, "Seamless genome editing in human pluripotent stem cells using custom endonuclease-based gene targeting and the piggyBac transposon," *Nature Protocols*, vol. 8, no. 10, pp. 2061–2078, 2013.

[60] M. Hamada, N. Nishio, Y. Okuno et al., "Integration Mapping of piggyBac-Mediated CD19 Chimeric Antigen Receptor T Cells Analyzed by Novel Tagmentation-Assisted PCR," *eBioMedicine*, vol. 34, pp. 18–26, 2018.

[61] E. P. Papapetrou, G. Lee, N. Malani et al., "Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells," *Nature Biotechnology*, vol. 29, no. 1, pp. 73–78, 2011.

[62] E. P. Papapetrou and A. Schambach, "Gene insertion into genomic safe harbors for human gene therapy," *Molecular Therapy*, vol. 24, no. 4, pp. 678–684, 2016.

[63] K. Kawakami, D. A. Largaespada, and Z. Ivics, "Transposons as tools for functional genomics in vertebrate models," *Trends in Genetics*, vol. 33, no. 11, pp. 784–801, 2017.

[64] W. L. Stanford, J. B. Cohn, and S. P. Cordes, "Gene-trap mutagenesis: past, present and beyond," *Nature Reviews. Genetics*, vol. 2, no. 10, pp. 756–768, 2001.

[65] R. H. Friedel and P. Soriano, "Gene trap mutagenesis in the mouse," *Methods in Enzymology*, vol. 477, pp. 243–269, 2010.

[66] A. Gragerov, K. Horie, M. Pavlova et al., "Large-scale, saturating insertional mutagenesis of the mouse genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 36, pp. 14406–14411, 2007.

[67] T. Shigeoka, M. Kawaichi, and Y. Ishida, "Suppression of nonsense-mediated mRNA decay permits unbiased gene trapping in mouse embryonic stem cells," *Nucleic Acids Research*, vol. 33, no. 2, p. e20, 2005.

[68] M. Di Matteo, E. Belay, M. K. Chuah, and T. Vandendriessche, "Recent developments in transposon-mediated gene therapy," *Expert Opinion on Biological Therapy*, vol. 12, no. 7, pp. 841–858, 2012.

[69] H. Zayed, Z. Izsvak, O. Walisko, and Z. Ivics, "Development of Hyperactive Sleeping Beauty Transposon Vectors by Mutational Analysis," *Molecular Therapy*, vol. 9, no. 2, pp. 292–304, 2004.

[70] M. A. Li, D. J. Turner, Z. Ning et al., "Mobilization of giant piggyBac transposons in the mouse genome," *Nucleic Acids Research*, vol. 39, no. 22, p. e148, 2011.

[71] Q. Liang, J. Kong, J. Stalker, and A. Bradley, "Chromosomal mobilization and reintegration of sleeping beauty and PiggyBac transposons," *Genesis*, vol. 47, no. 6, pp. 404–408, 2009.

[72] A. Lacoste, F. Berenshteyn, and A. H. Brivanlou, "An efficient and reversible transposable system for gene delivery and lineage- specific differentiation in human embryonic stem cells," *Cell Stem Cell*, vol. 5, no. 3, pp. 332–342, 2009.

[73] K. Yusa, L. Zhou, M. A. Li, A. Bradley, and N. L. Craig, "A hyperactive piggyBac transposase for mammalian applications," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 4, pp. 1531–1536, 2011.

[74] M. di Matteo, E. Samara-Kuko, N. J. Ward et al., "Hyperactive PiggyBac Transposons for Sustained and Robust Liver-targeted Gene Therapy," *Molecular Therapy*, vol. 22, no. 9, pp. 1614–1624, 2014.

[75] X. Li, E. R. Burnight, A. L. Cooney et al., "piggyBac transposase tools for genome engineering," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 25, pp. E2279–E2287, 2013.

[76] S. Bire, D. Ley, S. Casteret, N. Mermod, Y. Bigot, and F. Rouleux-Bonnin, "Optimization of the piggyBac transposon using mRNA and insulators: toward a more reliable gene delivery system," *PLoS One*, vol. 8, no. 12, article e82559, 2013.

[77] S. Bire, N. Ishac, and F. Rouleux-Bonnin, "In vitro synthesis, delivery, and bioavailability of exogenous mRNA in gene transfer mediated by PiggyBac transposition," *Methods in Molecular Biology*, vol. 1428, pp. 187–217, 2016.

[78] R. Mitra, J. Fain-Thornton, and N. L. Craig, "piggyBac can bypass DNA synthesis during cut and paste transposition," *The EMBO Journal*, vol. 27, no. 7, pp. 1097–1109, 2008.

[79] T. Gaj, C. A. Gersbach, and C. F. Barbas 3rd., "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering," *Trends in Biotechnology*, vol. 31, no. 7, pp. 397–405, 2013.

[80] W. T. Hendriks, C. R. Warren, and C. A. Cowan, "Genome editing in human pluripotent stem cells: approaches, pitfalls, and solutions," *Cell Stem Cell*, vol. 18, no. 1, pp. 53–65, 2016.

[81] J. B. Owens, D. Mauro, I. Stoytchev et al., "Transcription activator like effector (TALE)-directed piggyBac transposition in human cells," *Nucleic Acids Research*, vol. 41, no. 19, pp. 9197–9207, 2013.

[82] C. Kettlun, D. L. Galvan, A. L. George Jr., A. Kaja, and M. H. Wilson, "Manipulating piggyBac Transposon Chromosomal Integration Site Selection in Human Cells," *Molecular Therapy*, vol. 19, no. 9, pp. 1636–1644, 2011.

[83] L. Ye, Z. You, Q. Qian et al., "TAL effectors mediate high-efficiency transposition of the piggyBac transposon in silkworm Bombyx mori L," *Scientific Reports*, vol. 5, no. 1, article 17172, 2015.

[84] L. Cong, F. A. Ran, D. Cox et al., "Multiplex genome engineering using CRISPR/Cas systems," *Science*, vol. 339, no. 6121, pp. 819–823, 2013.

[85] L. A. Gilbert, M. H. Larson, L. Morsut et al., "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes," *Cell*, vol. 154, no. 2, pp. 442–451, 2013.

[86] S. Konermann, M. D. Brigham, A. E. Trevino et al., "Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex," *Nature*, vol. 517, no. 7536, pp. 583–588, 2015.

[87] L. Goshayeshi, S. Yousefi Taemeh, N. Dehdilani, M. Nasiri, M. M. Ghahramani Seno, and H. Dehghani, "CRISPR/d-Cas9-mediated transposition with specificity and efficiency of site-directed genomic insertions," *The FASEB Journal*, vol. 35, no. 2, article e21359, 2021.

[88] M. A. Li, S. J. Pettitt, K. Yusa, and A. Bradley, "Genome-wide forward genetic screens in mouse ES cells," *Methods in Enzymology*, vol. 477, pp. 217–242, 2010.

[89] S. Hardy, V. Legagneux, Y. Audic, and L. Paillard, "Reverse genetics in eukaryotes," *Biology of the Cell*, vol. 102, no. 10, pp. 561–580, 2010.

[90] E. M. Moresco, X. Li, and B. Beutler, "Going forward with genetics: recent technological advances and forward genetics in mice," *The American Journal of Pathology*, vol. 182, no. 5, pp. 1462–1473, 2013.

[91] G. Liu, X. Wang, Y. Liu et al., "Arrayed mutant haploid embryonic stem cell libraries facilitate phenotype-driven genetic screens," *Nucleic Acids Research*, vol. 45, no. 22, p. e180, 2017.

[92] C. Mayor-Ruiz, O. Dominguez, and O. Fernandez-Capetillo, "Trap^{Seq}: An RNA Sequencing-Based Pipeline for the Identification of Gene-Trap Insertions in Mammalian Cells," *Journal of Molecular Biology*, vol. 429, no. 18, pp. 2780–2789, 2017.

[93] W. C. Skarnes, B. Rosen, A. P. West et al., "A conditional knockout resource for the genome-wide study of mouse gene function," *Nature*, vol. 474, no. 7351, pp. 337–342, 2011.

[94] Y. Huang, S. J. Pettitt, G. Guo et al., "Isolation of homozygous mutant mouse embryonic stem cells using a dual selection system," *Nucleic Acids Research*, vol. 40, no. 3, p. e21, 2012.

[95] G. Luo, I. M. Santoro, L. D. McDaniel et al., "Cancer predisposition caused by elevated mitotic recombination in bloom mice," *Nature Genetics*, vol. 26, no. 4, pp. 424–429, 2000.

[96] G. Guo, W. Wang, and A. Bradley, "Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells," *Nature*, vol. 429, no. 6994, pp. 891–895, 2004.

[97] K. Yusa, K. Horie, G. Kondoh et al., "Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene," *Nature*, vol. 429, no. 6994, pp. 896–899, 2004.

[98] M. Yi, N. Hong, and Y. Hong, "Generation of medaka fish haploid embryonic stem cells," *Science*, vol. 326, no. 5951, pp. 430–433, 2009.

[99] U. Elling, J. Taubenschmid, G. Wirnsberger et al., "Forward and reverse genetics through derivation of haploid mouse embryonic stem cells," *Cell Stem Cell*, vol. 9, no. 6, pp. 563–574, 2011.

[100] M. Leeb and A. Wutz, "Derivation of haploid embryonic stem cells from mouse embryos," *Nature*, vol. 479, no. 7371, pp. 131–134, 2011.

[101] W. Li, L. Shuai, H. Wan et al., "Androgenetic haploid embryonic stem cells produce live transgenic mice," *Nature*, vol. 490, no. 7420, pp. 407–411, 2012.

[102] H. Yang, L. Shi, B. A. Wang et al., "Generation of genetically modified mice by oocyte injection of androgenetic haploid embryonic stem cells," *Cell*, vol. 149, no. 3, pp. 605–617, 2012.

[103] W. Li, X. Li, T. Li et al., "Genetic modification and screening in rat using haploid embryonic stem cells," *Cell Stem Cell*, vol. 14, no. 3, pp. 404–414, 2014.

[104] H. Yang, Z. Liu, Y. Ma et al., "Generation of haploid embryonic stem cells from Macaca fascicularis monkey parthenotes," *Cell Research*, vol. 23, no. 10, pp. 1187–1200, 2013.

[105] I. Sagi, G. Chia, T. Golan-Lev et al., "Derivation and differentiation of haploid human embryonic stem cells," *Nature*, vol. 532, no. 7597, pp. 107–111, 2016.

[106] C. Zhong, M. Zhang, Q. Yin et al., "Generation of human haploid embryonic stem cells from parthenogenetic embryos obtained by microsurgical removal of male pronucleus," *Cell Research*, vol. 26, no. 6, pp. 743–746, 2016.

[107] X. M. Zhang, K. Wu, Y. Zheng et al., "In vitro expansion of human sperm through nuclear transfer," *Cell Research*, vol. 30, no. 4, pp. 356–359, 2020.

[108] M. Leeb, S. Dietmann, M. Paramor, H. Niwa, and A. Smith, "Genetic exploration of the exit from self-renewal using haploid embryonic stem cells," *Cell Stem Cell*, vol. 14, no. 3, pp. 385–393, 2014.

[109] C. Zhong, Q. Yin, Z. Xie et al., "CRISPR-Cas9-mediated genetic screening in mice with haploid embryonic stem cells carrying a guide RNA library," *Cell Stem Cell*, vol. 17, no. 2, pp. 221–232, 2015.

[110] A. Monfort, G. Di Minin, A. Postlmayr et al., "Identification of Spen as a Crucial Factor for Xist Function through Forward Genetic Screening in Haploid Embryonic Stem Cells," *Cell Reports*, vol. 12, no. 4, pp. 554–561, 2015.

[111] J. V. Forment, M. Herzog, J. Coates et al., "Genome-wide genetic screening with chemically mutagenized haploid embryonic stem cells," *Nature Chemical Biology*, vol. 13, no. 1, pp. 12–14, 2017.

[112] U. Elling, R. A. Wimmer, A. Leibbrandt et al., "A reversible haploid mouse embryonic stem cell biobank resource for functional genomics," *Nature*, vol. 550, no. 7674, pp. 114–118, 2017.

[113] S. J. Pettitt, D. B. Krastev, H. N. Pemberton et al., "Genome-wide barcoded transposon screen for cancer drug sensitivity in haploid mouse embryonic stem cells," *Scientific Data*, vol. 4, no. 1, article 170020, 2017.

[114] Q. Li, Y. Li, S. Yang et al., "CRISPR-Cas9-mediated base-editing screening in mice identifies DND1 amino acids that are critical for primordial germ cell development," *Nature Cell Biology*, vol. 20, no. 11, pp. 1315–1325, 2018.

[115] A. Yilmaz, M. Peretz, A. Aharony, I. Sagi, and N. Benvenisty, "Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells," *Nature Cell Biology*, vol. 20, no. 5, pp. 610–619, 2018.

[116] M. Bai, Y. Han, Y. Wu et al., "Targeted genetic screening in mice through haploid embryonic stem cells identifies critical genes in bone development," *PLoS Biology*, vol. 17, no. 7, article e3000350, 2019.

[117] T. Cui, Z. Li, Q. Zhou, and W. Li, "Current advances in haploid stem cells," *Protein & Cell*, vol. 11, no. 1, pp. 23–33, 2020.

[118] K. Peng, X. Li, C. Wu et al., "Derivation of Haploid Trophoblast Stem Cells via Conversion In Vitro," *iScience*, vol. 11, pp. 508–518, 2019.

[119] G. Guo, J. Yang, J. Nichols et al., "Klf4 reverts developmentally programmed restriction of ground state pluripotency," *Development*, vol. 136, no. 7, pp. 1063–1069, 2009.

[120] A. Gillich, S. Bao, N. Grabole et al., "Epiblast stem cell-based system reveals reprogramming synergy of germline factors," *Cell Stem Cell*, vol. 10, no. 4, pp. 425–439, 2012.

[121] S. Bao, F. Tang, X. Li et al., "Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells," *Nature*, vol. 461, no. 7268, pp. 1292–1295, 2009.

[122] Q. Gao, W. Zhang, Y. Zhao et al., "High-throughput screening in postimplantation haploid epiblast stem cells reveals Hs3st3b1 as a modulator for reprogramming," *Stem Cells Translational Medicine*, vol. 10, no. 5, pp. 743–755, 2021.

[123] G. Zhang, X. Li, Y. Sun, X. Wang, G. Liu, and Y. Huang, "A Genetic Screen Identifies Etl4-Deficiency Capable of Stabilizing the Haploidy in Embryonic Stem Cells," *Stem Cell Reports*, vol. 16, no. 1, pp. 29–38, 2021.

[124] S. J. Pettitt, F. L. Rehman, I. Bajrami et al., "A genetic screen using the PiggyBac transposon in haploid cells identifies Parp1 as a mediator of olaparib toxicity," *PLoS One*, vol. 8, no. 4, article e61520, 2013.

[125] J. E. Carette, C. P. Guimaraes, I. Wuethrich et al., "Global gene disruption in human cells to assign genes to phenotypes by deep sequencing," *Nature Biotechnology*, vol. 29, no. 6, pp. 542–546, 2011.

[126] J. Mao, K. Xu, J. Han, G. Feng, Y. Zhang, and W. Li, "Rapid construction of a whole-genome mutant library by combining haploid stem cells and inducible self-inactivating PiggyBac transposon," *Protein & Cell*, vol. 11, no. 6, pp. 452–457, 2020.

[127] U. Weissbein, M. Peretz, O. Plotnik et al., "Genome-wide screen for culture adaptation and tumorigenicity-related genes in human pluripotent stem cells," *iScience*, vol. 11, pp. 398–408, 2019.

[128] A. J. Dupuy, K. Akagi, D. A. Largaespada, N. G. Copeland, and N. A. Jenkins, "Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system," *Nature*, vol. 436, no. 7048, pp. 221–226, 2005.

[129] A. Wutz, "Haploid mouse embryonic stem cells: rapid genetic screening and germline transmission," *Annual Review of Cell and Developmental Biology*, vol. 30, no. 1, pp. 705–722, 2014.

[130] W. Liu, C. Deng, C. Godoy-Parejo, Y. Zhang, and G. Chen, "Developments in cell culture systems for human pluripotent stem cells," *World Journal of Stem Cells*, vol. 11, no. 11, pp. 968–981, 2019.

[131] 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.

[132] X. Gao, M. Nowak-Imialek, X. Chen et al., "Establishment of porcine and human expanded potential stem cells," *Nature Cell Biology*, vol. 21, no. 6, pp. 687–699, 2019.

[133] 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.

[134] J. Yang, D. J. Ryan, G. Lan, X. Zou, and P. Liu, "In vitro establishment of expanded-potential stem cells from mouse pre-implantation embryos or embryonic stem cells," *Nature Protocols*, vol. 14, no. 2, pp. 350–378, 2019.

[135] L. Yang, Y. Han, B. E. Nilsson-Payant et al., "A human pluripotent stem cell-based platform to study SARS-CoV-2 tropism and model virus infection in human cells and organoids," *Cell Stem Cell*, vol. 27, no. 1, pp. 125–136.e7, 2020.

[136] F. Wu, Y. Chen, K. You et al., "Generation of WAe001-A-58 human embryonic stem cell line with inducible expression of the SARS-CoV-2 nucleocapsid protein," *Stem Cell Research*, vol. 53, article 102197, 2021.

[137] K. Woltjen, I. P. Michael, P. Mohseni et al., "piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells," *Nature*, vol. 458, no. 7239, pp. 766–770, 2009.

[138] K. Yusa, R. Rad, J. Takeda, and A. Bradley, "Generation of transgene-free induced pluripotent mouse stem cells by the _piggyBac_ transposon," *Nature Methods*, vol. 6, no. 5, pp. 363–369, 2009.

[139] H. S. Kim, J. M. Bernitz, D. F. Lee, and I. R. Lemischka, "Genomic editing tools to model human diseases with isogenic pluripotent stem cells," *Stem Cells and Development*, vol. 23, no. 22, pp. 2673–2686, 2014.

[140] G. Wang, L. Yang, D. Grishin et al., "Efficient, footprint-free human iPSC genome editing by consolidation of Cas9/-CRISPR and piggyBac technologies," *Nature Protocols*, vol. 12, no. 1, pp. 88–103, 2017.

[141] X. Xu, Y. Tay, B. Sim et al., "Reversal of phenotypic abnormalities by CRISPR/Cas9-mediated gene correction in Huntington disease patient-derived induced pluripotent stem cells," *Stem Cell Reports*, vol. 8, no. 3, pp. 619–633, 2017.

[142] T. I. Cornu, C. Mussolino, and T. Cathomen, "Refining strategies to translate genome editing to the clinic," *Nature Medicine*, vol. 23, no. 4, pp. 415–423, 2017.

[143] J. Shendure, S. Balasubramanian, G. M. Church et al., "DNA sequencing at 40: past, present and future," *Nature*, vol. 550, no. 7676, pp. 345–353, 2017.

[144] E. Zeggini, A. L. Gloyn, A. C. Barton, and L. V. Wain, "Translational genomics and precision medicine: moving from the lab to the clinic," *Science*, vol. 365, no. 6460, pp. 1409–1413, 2019.

[145] L. Mátés, M. K. Chuah, E. Belay et al., "Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates," *Nature Genetics*, vol. 41, no. 6, pp. 753–761, 2009.

[146] I. Grabundzija, M. Irgang, L. Mátés et al., "Comparative analysis of transposable element vector systems in human cells," *Molecular Therapy*, vol. 18, no. 6, pp. 1200–1209, 2010.

[147] J. E. Doherty, L. E. Huye, K. Yusa, L. Zhou, N. L. Craig, and M. H. Wilson, "Hyperactive piggyBac gene transfer in human cells and in vivo," *Human Gene Therapy*, vol. 23, no. 3, pp. 311–320, 2012.

[148] S. Wen, H. Zhang, Y. Li et al., "Characterization of constitutive promoters for piggyBac transposon-mediated stable transgene expression in mesenchymal stem cells (MSCs)," *PLoS One*, vol. 9, no. 4, article e94397, 2014.

[149] Y. Shu, C. Yang, X. Ji et al., "Reversibly immortalized human umbilical cord-derived mesenchymal stem cells (UC-MSCs) are responsive to BMP9-induced osteogenic and adipogenic differentiation," *Journal of Cellular Biochemistry*, vol. 119, no. 11, pp. 8872–8886, 2018.

[150] V. Bahrambeigi, N. Ahmadi, S. Moisyadi, J. Urschitz, R. Salehi, and S. Haghjooy Javanmard, "PhiC31/PiggyBac modified stromal stem cells: effect of interferon γ and/or tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL) on murine melanoma," *Molecular Cancer*, vol. 13, no. 1, p. 255, 2014.

[151] M. Loperfido, S. Jarmin, S. Dastidar et al., "piggyBac transposons expressing full-length human dystrophin enable genetic correction of dystrophic mesoangioblasts," *Nucleic Acids Research*, vol. 44, no. 2, pp. 744–760, 2016.

[152] S. Saha, L. E. Woodard, E. M. Charron, R. C. Welch, C. M. Rooney, and M. H. Wilson, "Evaluating the potential for undesired genomic effects of the piggyBac transposon system in human cells," *Nucleic Acids Research*, vol. 43, no. 3, pp. 1770–1782, 2015.

[153] S. K. Saridey, L. Liu, J. E. Doherty et al., "PiggyBac Transposon-based Inducible Gene Expression In Vivo After Somatic Cell Gene Transfer," *Molecular Therapy*, vol. 17, no. 12, pp. 2115–2120, 2009.

[154] S. R. Yant, L. Meuse, W. Chiu, Z. Ivics, Z. Izsvak, and M. A. Kay, "Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system," *Nature Genetics*, vol. 25, no. 1, pp. 35–41, 2000.

[155] T. Cui, L. Jiang, T. Li et al., "Derivation of mouse haploid trophoblast stem cells," *Cell Reports*, vol. 26, no. 2, pp. 407–414.e5, 2019.

[156] S. Gayle, Y. Pan, S. Landrette, and T. Xu, "piggyBac Insertional Mutagenesis Screen Identifies a Role for Nuclear RHOA in Human ES Cell Differentiation," *Stem Cell Reports*, vol. 4, no. 5, pp. 926–938, 2015.