



## OPEN ACCESS

## EDITED BY

Markus Friedrich,  
Wayne State University, United States

## REVIEWED BY

Shawn L. Chavez,  
Oregon Health and Science University,  
United States  
Joel Berletch,  
University of Washington, United States

## \*CORRESPONDENCE

Fengping Xu,  
xufengping@genomics.cn  
Yonglun Luo,  
alun@biomed.au.dk  
Ge Lin,  
linggf@hotmail.com

<sup>†</sup>These authors have contributed equally to this work

## SPECIALTY SECTION

This article was submitted to Evolutionary Developmental Biology, a section of the journal Frontiers in Cell and Developmental Biology

RECEIVED 16 August 2022

ACCEPTED 17 October 2022

PUBLISHED 11 November 2022

## CITATION

Li C, Zhang Y, Leng L, Pan X, Zhao D, Li X, Huang J, Bolund L, Lin G, Luo Y and Xu F (2022), The single-cell expression profile of transposable elements and transcription factors in human early biparental and uniparental embryonic development. *Front. Cell Dev. Biol.* 10:1020490. doi: 10.3389/fcell.2022.1020490

## COPYRIGHT

© 2022 Li, Zhang, Leng, Pan, Zhao, Li, Huang, Bolund, Lin, Luo and Xu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# The single-cell expression profile of transposable elements and transcription factors in human early biparental and uniparental embryonic development

Conghui Li<sup>1,2,3,4,5,6</sup>, Yue Zhang<sup>7</sup>, Lizhi Leng<sup>8,9,10</sup>, Xiaoguang Pan<sup>3,4</sup>, Depeng Zhao<sup>11</sup>, Xuemei Li<sup>11</sup>, Jinrong Huang<sup>1,2,3,4,5,6</sup>, Lars Bolund<sup>2,3,4,5,12,13</sup>, Ge Lin<sup>8,9,10\*†</sup>, Yonglun Luo<sup>1,2,3,4,5,12,13\*†</sup> and Fengping Xu<sup>1,2,3,4,5,14\*†</sup>

<sup>1</sup>College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, <sup>2</sup>BGI-Shenzhen, Beishan Industrial Zone, Shenzhen, China, <sup>3</sup>Qingdao-Europe Advanced Institute for Life Sciences, BGI-Shenzhen, Qingdao, China, <sup>4</sup>Lars Bolund Institute of Regenerative Medicine, BGI-Qingdao, BGI-Shenzhen, Qingdao, China, <sup>5</sup>China National GeneBank, BGI-Shenzhen, Shenzhen, China, <sup>6</sup>Department of Biology, University of Copenhagen, Copenhagen, Denmark, <sup>7</sup>BGI-Qingdao, BGI-Shenzhen, Qingdao, Shandong, China, <sup>8</sup>Institute of Reproductive and Stem Cell Engineering, School of Basic Medical Science, Central South University, Changsha, China, <sup>9</sup>Key Laboratory of Reproductive and Stem Cells Engineering, Ministry of Health, Changsha, China, <sup>10</sup>Reproductive & Genetic Hospital of CITIC-Xiangya, Changsha, China, <sup>11</sup>Department of Reproductive Medicine, Affiliated Shenzhen Maternity and Child Healthcare Hospital, Southern Medical University, Shenzhen, China, <sup>12</sup>Department of Biomedicine, Aarhus University, Aarhus, Denmark, <sup>13</sup>Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Denmark, <sup>14</sup>BGI Cell, BGI-Shenzhen, Shenzhen, China

Transposable elements (TEs) and transcription factors (TFs) are involved in the precise regulation of gene expression during the preimplantation stage. Activation of TEs is a key event for mammalian embryonic genome activation and preimplantation early embryonic development. TFs are involved in the regulation of drastic changes in gene expression patterns, but an inventory of the interplay between TEs and TFs during normal/abnormal human embryonic development is still lacking. Here we used single-cell RNA sequencing data generated from biparental and uniparental embryos to perform an integrative analysis of TE and TF expression. Our results showed that endogenous retroviruses (ERVs) are mainly expressed during the minor embryonic genome activation (EGA) process of early embryos, while Alu is gradually expressed in the middle and later stages. Some important ERVs (e.g., LTR5\_Hs, MLT2A1) and Alu TEs are expressed at significantly lower levels in androgenic embryos. Integrative analysis revealed that the expression of the transcription factors *CTCF* and *POU5F1* is correlated with the differential expression of ERV TEs. Comparative coexpression network analysis further showed distinct expression levels of important TFs (e.g., *LEUTX* and *ZSCAN5A*) in dizygotic embryos vs. parthenogenetic and androgenic embryos. This systematic investigation of TE and TF expression in human early embryonic development by single-cell RNA sequencing provides valuable insights into mammalian embryonic development.

## KEYWORDS

early embryonic development, transposable elements, transcriptional factors, androgenetic embryo, parthenogenetic embryo

## Introduction

The transcriptional activity in mammalian early embryos leads to obvious dynamic changes in gene expression, particularly during the transition from zygote to morula (Niakan et al., 2012). The early stages of embryogenesis and precise activation of the zygotic genome are key to successful development (Tadros and Lipshitz, 2009; Lee et al., 2014).

Previous studies have shown that transposable elements (TEs) and transcription factors (TFs) are involved in the precise regulation of gene expression during the preimplantation stage (Godini and Fallahi, 2018; Rodriguez-Terrones and Torres-Padilla, 2018). TEs constitute more than 40% of the human genome (de Koning APJGu et al., 2011), and are divided into two types: DNA transposons transposed by cut-and-paste and retrotransposons mediated by RNA (Finnegan, 1989). Among all TEs, retrotransposons are predominant, including long terminal repeat (LTR) retrotransposons [including endogenous retroviruses (ERVs)] and non-LTR retrotransposons [long interspersed elements (LINEs) and short interspersed elements (SINEs)] (Wicker et al., 2007). The expression of retrotransposons is a vital event for genome reprogramming during early embryonic development (Bui et al., 2009). Some studies have also provided substantial evidence for the extensive role of TEs in regulating gene expression during early embryonic development (Kigami et al., 2003; Macfarlan et al., 2012; Percharde et al., 2018; Deng et al., 2019). Many TEs have been found, but the regulatory effects of most of these TEs have not been studied in depth.

In addition, the EGA process is directly regulated and orchestrated by different types of TFs, and many studies have provided important insights into the TFs present during early embryonic development (Lee et al., 2013; Leichsenring et al., 2013; Goolam et al., 2016; De Iaco et al., 2017; Hendrickson et al., 2017). Some important interplay between TEs and TFs is known to occur during early embryonic development. For instance, L1 RNA inhibits the expression of the transcription factor DUX by recruiting nuclear protein/Kap1, thereby indirectly inhibiting the transcription of ERVs (Percharde et al., 2018). Polak and Domany (2006) reported that Alu elements are present upstream of the transcription start site of a large number of genes, and the Alu sequence was found to contain multiple functional TF binding sites. These cross-sectional studies suggest an association between TEs and TFs. Therefore, it is necessary to systematically explore the cooperative regulation of TEs and TFs in early embryonic development at the transcriptional level.

Androgenetic (AG) and parthenogenetic (PG) embryos have two paternal or maternal genomes, respectively, and are valuable tools for studying the effect of parental genomes on

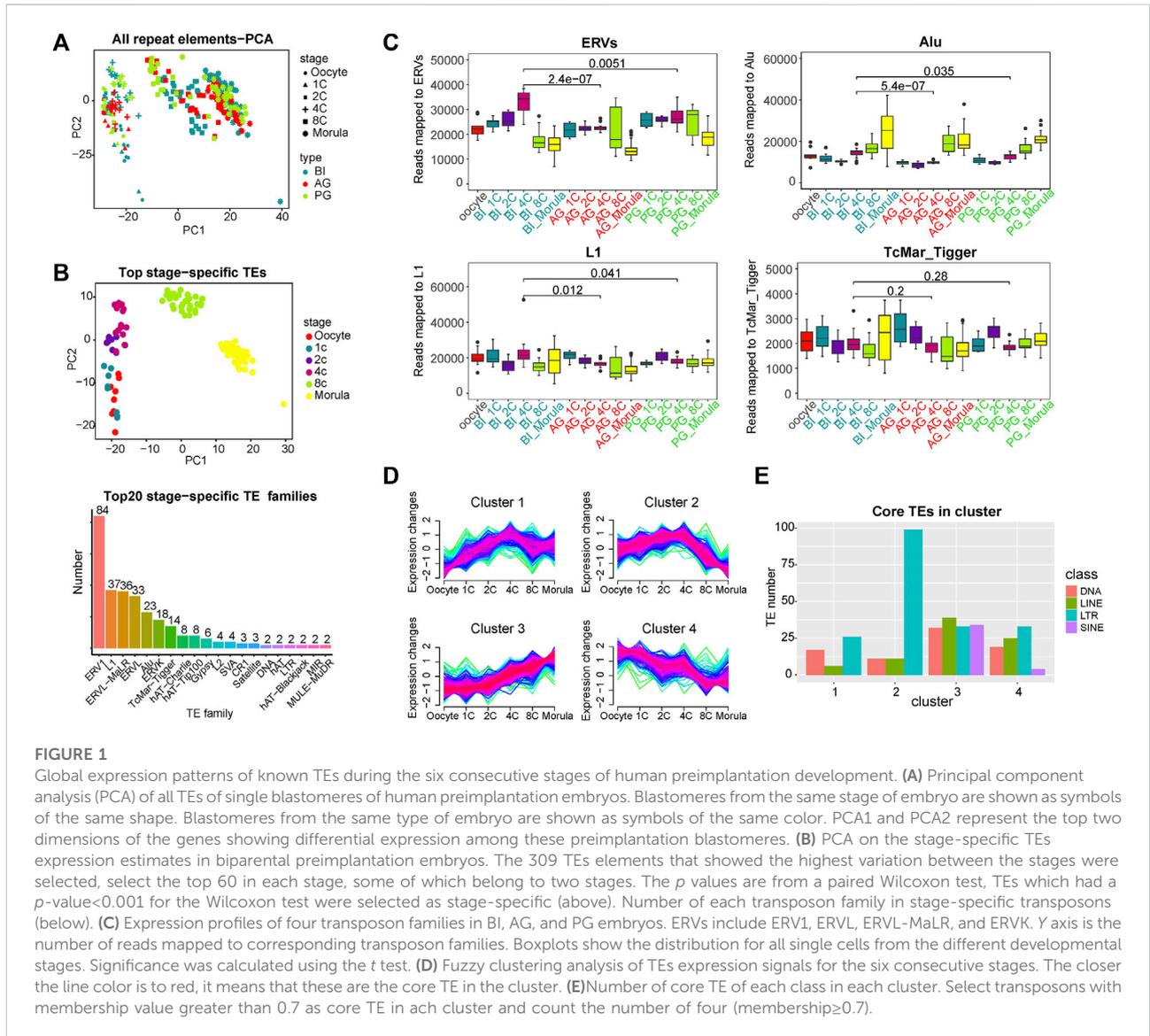
preimplantation embryo development (Lagutina et al., 2004). In the initial stage after fertilization, zygotes undergo maternal-to-zygotic transition (Schier, 2007), during which maternal factors are gradually consumed and degraded, and the maternal and paternal genomes are reactivated to produce new mRNA and proteins to support embryonic development. A number of previous studies have shown that the contribution of the paternal and maternal genomes to the mammalian embryo genome is not exactly the same, and the diploid genome from only one of the two parental sexes cannot support complete embryogenesis (McGrath and Solter, 1984; Park et al., 2011; Hu et al., 2015).

To date, a systematic understanding of the role of TEs and TFs in early embryos is still lacking, particularly regarding how these two categories of regulatory elements orchestrate the reprogramming of maternal and paternal genomes during preimplantation development. In this study, we used single-cell RNA-seq data to conduct a comprehensive study of TE expression profiles in biparental and uniparental embryonic cells for the first time, in addition to identifying the key expressed TFs. The results showed some potential connections between TEs and TFs that may play an important role in embryonic genome activation and further differentiation.

## Result

### Transposable elements show developmental stage specific expression profiles in both uniparental and biparental embryos

In our previous work, we reported that gene expression profiles showed phased differences during early embryonic development, and paternal and maternal genomes functioned differently during EGA and embryonic differentiation (Leng et al., 2019). This project provided us with a valuable single-cell RNA sequencing dataset to uncover the expression profiles of TEs in uniparental and biparental embryos. We quantified TE expression in all the samples, which contain transcriptome data of 285 single cells of the oocytes ( $n = 9$ ), 1-cell ( $n = 15$ ), 2-cell ( $n = 28$ ), 4-cell ( $n = 37$ ), 8-cell ( $n = 77$ ), and morula ( $n = 119$ ) stages in BI, AG, and PG embryos. To investigate whether the expression patterns of the TEs at each stage of the biparental and uniparental embryos were different, we performed a principal component analysis (PCA) based on the expression of 912 TEs. As shown in Figure 1A and Supplementary Table S1, the three types of embryos were clustered based on developmental stages rather



than embryonic types, which showed that the expression of TEs was stage specific. Several lines of evidence suggest that the human EGA process accelerates from the 4-cell stage (Braude et al., 1988; Dobson et al., 2004; Vassena et al., 2011; Xue et al., 2013; Yan et al., 2013), so this result indicates that the 4-cell stage may be a critical period in early embryonic development.

To identify the TEs with stage-specific expression in early embryos, we calculated the differential expression of all TEs at each stage by the Wilcoxon test (*p* value < 0.05, Supplementary Table S2). Specific TEs at each stage (the top 60 with the lowest *p* value) were used for PCA. In Figure 1B/above, the different stages are better separated. We counted the family types of all these stage-specific TEs, and the top 7 TE families with a number greater than 10 were ERV1 (84), L1 (37), ERVL-MaLR (36), ERVL (33), Alu(23), ERVK(18), and TcMar-Tigger (14)

(Figure 1B/below). The subsequent analysis of TEs was focused on the above families.

### Expression trend characteristics of transposable elements of different families

To more systematically analyse the transcription features of TEs from the four families at various stages of early embryonic development, a fraction of the transcriptome derived from these four families was calculated. In BI embryos, the ERV families were highly expressed from the 1- to 4-cell stage, and the expression decreased markedly from the 8-cell stage (Figure 1C). The expression level of the Alu family was low at

the early stage and gradually increased from the 4-cell stage. The differences in expression in the L1 and TcMar\_Tigger families in each stage were relatively mild. To further explore the expression characteristics of all the expressed TEs, we used a Mfuzz package to perform time-series analysis in BI embryos, and the TEs were divided into 4 clusters (Figure 1D). This showed that the expression trend of TEs was diversified in early embryonic development, and the expression trend of Cluster 1 and Cluster two showed a slight turning point at the 4-cell stage, suggesting that it may be related to EGA. Next, we investigated the class of core TEs in each cluster. The histogram in Figure 1E showed that LTR transposons (ERVs) had the highest percentage in Cluster 2, being upregulated from the 1- to 4-cell stage, maintaining a relatively high overall expression level, and downregulated from the 4-cell stage to morula stage. Previous studies by Xue and Yan have indicated that a minor EGA wave occurs during the 1- to 4-cell stage, and a major wave of EGA wave occurs at the 4-cell to morula stages (Xue et al., 2013; Yan et al., 2013). The expression trend of ERVs is highly consistent with the occurrence of minor EGA and major EGA, so we speculated that ERVs may be involved in the regulation of EGA. ERVs undergo a brief activation in the early stage of early embryonic development, which may contribute to the occurrence of minor EGA. Interestingly, SINE-class TEs seem to only appear in Cluster three and Cluster 4. The transcription of Cluster four shows a gradual decrease as the embryo develops, which mainly reflects the consumption and degradation of maternal materials, so the SINE-class TEs of the embryo themselves are gradually activated from the 4-cell stage in early embryonic development.

## Validation of transposable element expression in human and mouse embryos using public data

To validate the robustness of our findings, we analysed two independent single-cell RNA sequencing datasets (GSE36552 and GSE44183) of human embryos (Xue et al., 2013; Yan et al., 2013), both of which contained oocyte and single-cell samples at various developmental stages from zygote to morula. Indeed, using the same analysis strategy, we found that these two independent experimental datasets showed similar PCA results (Supplementary Figure S1A,B). The expression of TEs is also stage specific, and most importantly stage-specific TE families are also largely consistent.

We also took advantage of the scRNA-seq public data from zygote to morula stages in mice from Deng et al. (2014), and the average TE expression values in each stage were retrieved from Supplementary Table S3 of the paper published by Steven Xijin Ge (Ge, 2017). The results of the clustering analysis showed that the expression of TE in mouse embryos transitioned slightly at the 2-cell stage (Supplementary Figure S1C), which was earlier

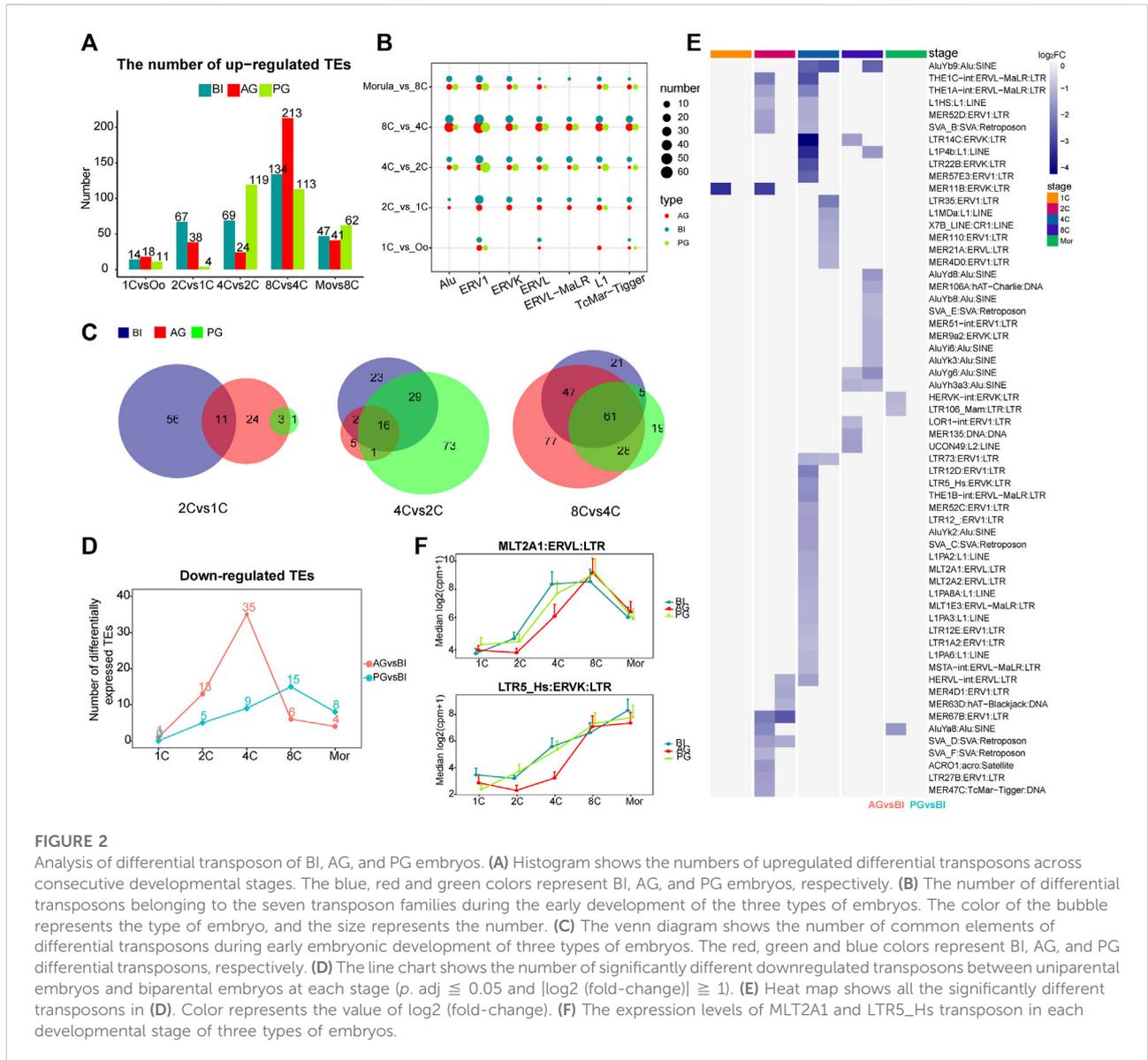
than the transition in human embryonic cells (4-cell stage). Evsikov et al. (2004) found that the LTRs sequences were abundant in the transcripts of 2-cell mouse embryos, and the major EGA in mice occurs during the 2-cell stage (Ram and Schultz, 1993). Various studies have fully confirmed that the EGA process in humans accelerates from the 4- to 8-cell stage, which is later than that in mice (Braude et al., 1988; Dobson et al., 2004; Vassena et al., 2011; Xue et al., 2013; Yan et al., 2013). The conservation of the expression trend of LTRs in humans and mice further indicates that ERVs may be related to the main EGA process in early embryos. In addition, these results also indicated that in both humans and mice, SINE-class TEs are gradually activated during early embryonic development, of course, Alu family which belongs to SINE-class is well known as a primate-specific family only considered in human here.

## Differences in transposable elements between biparental and uniparental embryos

EGA is one of the most important events that occur during preimplantation and embryo development in humans, and the embryonic genome should experience transcriptional quiescence to achieve large-scale transcriptional activity. To explore the activity of the TEs in the uniparental embryos during this process, we analysed the differentially expressed TEs during embryonic development, which was determined at a cutoff of FDR-corrected  $p$  value  $\leq 0.01$  and  $|\log_2(\text{fold change})| \geq 1$ .

We counted the number of upregulated TEs at each stage (Figure 2A). The statistical data of upregulated genes in the previous study (Supplementary Figure S2A) indicated that the expression dynamics of TEs are basically consistent with those of genes. Overall, the level of TE activation in uniparental embryos was lower than that in biparental embryos at the 1- to 2-cell stage. Compared to BI embryos, PG embryos showed compensatory upregulation at the 4-cell stage to a certain extent (69, 24, and 119 for BI, AG, and PG embryos, respectively), while this occurred at the 8-cell stage in AG embryos (134, 213, and 113 for BI, AG, and PG embryos, respectively), which may indicate that the TE activation in AG embryos in the minor EGA process is insufficient, and the TE activation in PG embryos is delayed. On the other hand, a large number of TEs are upregulated at the 8-cell stage indicating that TEs undergo a rapid and transient activation in the early embryonic development stage.

Next, we counted the number of the seven stage-specific TE families among the differentially expressed TEs at each stage. The dot plot in Figure 2B showed a huge distinction in the number of stage specific upregulated TEs in the three types of embryos from the 1- to 8-cell stage, with ERV1 having the largest difference, followed by Alu. Studies have shown that in embryonic stem cells, the ERV1 family contributes the greatest number of OCT4-



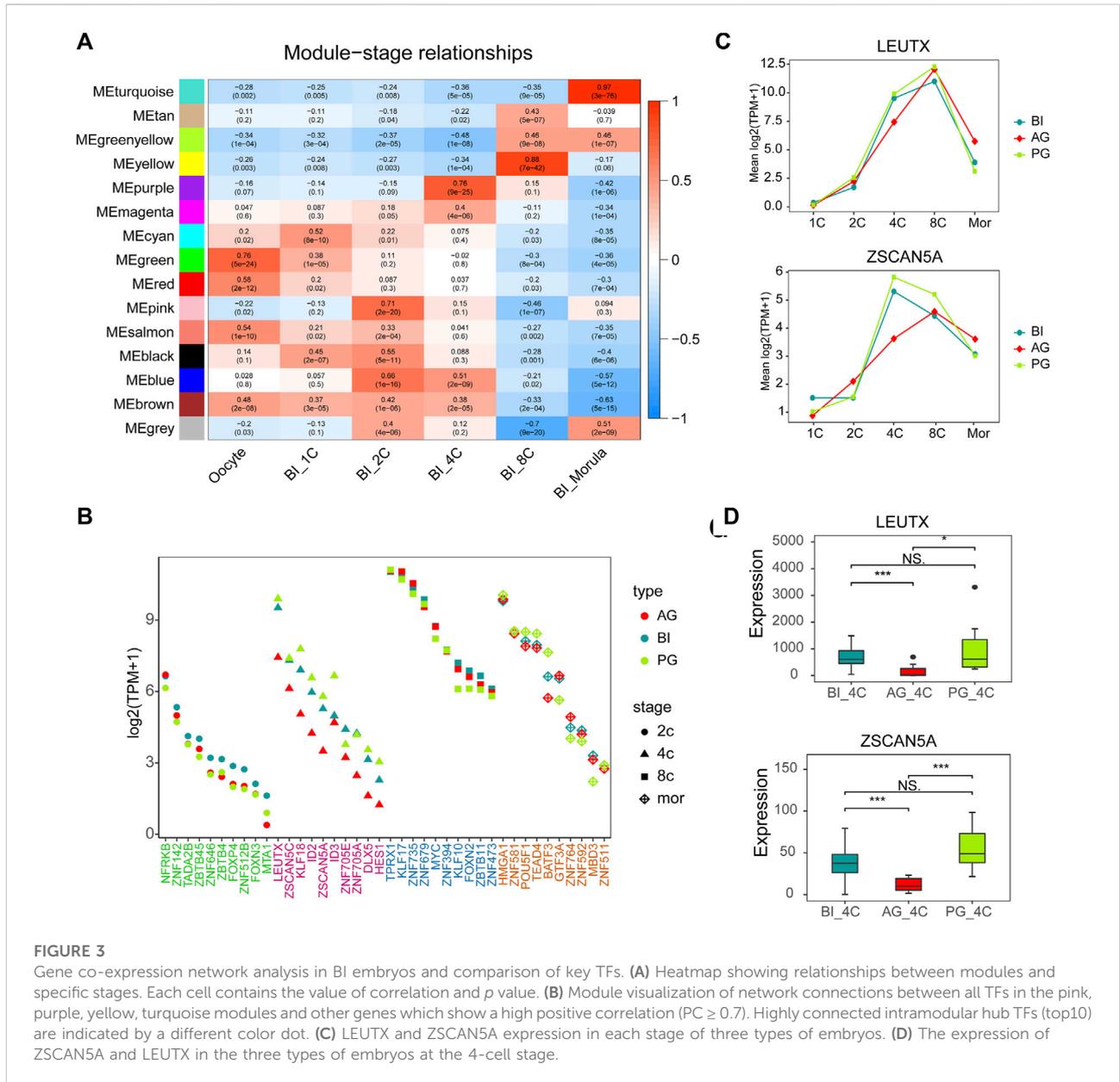
**FIGURE 2**

Analysis of differential transposon of BI, AG, and PG embryos. **(A)** Histogram shows the numbers of upregulated differential transposons across consecutive developmental stages. The blue, red and green colors represent BI, AG, and PG embryos, respectively. **(B)** The number of differential transposons belonging to the seven transposon families during the early development of the three types of embryos. The color of the bubble represents the type of embryo, and the size represents the number. **(C)** The venn diagram shows the number of common elements of differential transposons during early embryonic development of three types of embryos. The red, green and blue colors represent BI, AG, and PG differential transposons, respectively. **(D)** The line chart shows the number of significantly different downregulated transposons between uniparental embryos and biparental embryos at each stage ( $p_{adj} \leq 0.05$  and  $|\log_2(\text{fold-change})| \geq 1$ ). **(E)** Heat map shows all the significantly different transposons in **(D)**. Color represents the value of  $\log_2(\text{fold-change})$ . **(F)** The expression levels of MLT2A1 and LTR5\_Hs transposon in each developmental stage of three types of embryos.

and NANOG-binding sites (Kunarsko et al., 2010). Then, we compared the intersecting upregulated TEs among the three types of embryos from the 1- to 8-cell stages. The AG embryos shared more TEs with BI embryos at the 1- to 2-cell and the 4- to 8-cell stages, with PG embryos at the 2- to 4-cell stages (Figure 2C). This characteristic may indicate that in the early embryo development process, the activation of TEs also makes a paternal or maternal biased contribution to the specific stage.

Comparison of the transcription of the 4 TE families in the three types of embryos, as apparent in Figure 1C, showed that the transcription trends of the ERVs were markedly different. The significant differences started at the 4-cell stage. The transcription level in AG/PG embryos was lower than that in

BI embryos. There was a significant upregulation of ERVs at the 2- to 4-cell stages in BI embryos, but this tendency appeared in AG and PG embryos at the 8-cell stage, which can be as attributed to the upregulation of ERVs during the development of uniparental embryos later than that of normal embryos. In addition, compared with the BI embryos, the Alu family TEs of uniparental embryos were also downregulated to a certain extent at the 4-cell stage, and this downregulation was more pronounced in AG embryos. Moreover, according to our calculation, within the proximal promoter region (−2000 bp, 500 bp) of all upregulated genes, the most abundant type of TE was the Alu element (Supplementary Table S3). These findings, while preliminary, suggest that the lower expression of ERVs and Alu at the 4-cell stage may affect the minor EGA



**FIGURE 3**

Gene co-expression network analysis in BI embryos and comparison of key TFs. (A) Heatmap showing relationships between modules and specific stages. Each cell contains the value of correlation and *p* value. (B) Module visualization of network connections between all TFs in the pink, purple, yellow, turquoise modules and other genes which show a high positive correlation (PC ≥ 0.7). Highly connected intramodular hub TFs (top10) are indicated by a different color dot. (C) LEUTX and ZSCAN5A expression in each stage of three types of embryos. (D) The expression of ZSCAN5A and LEUTX in the three types of embryos at the 4-cell stage.

process in uniparental embryos, resulting in a higher probability of embryo EGA initiation failure.

Next, we compared the differentially expressed TEs between uniparental embryos and biparental embryos at each stage. At the 2- and 4-cell stages, more TEs were significantly downregulated in AG embryos (13, and 35 for 2-, and 4-cell embryos, respectively), especially at the 4-cell stage (Figure 2D). All the differentially expressed TEs are shown in the heatmap (Figure 2E). As expected, the most downregulated TEs in AG embryos belonged to the ERV families. In addition, we found that at the 4-cell stage, most of the TEs downregulated in AG embryos belonged to the upregulated TEs at the 2- to 4-cell stage of BI embryos (Supplementary Figure S2B), including LTR5\_Hs MLT2A1,

MLT2A2, LTR12D, LIHS, and AluYb9. Grow et al. (2015) reported that LTR5\_Hs is bound by POU5F1, which is a master regulator of pluripotency, in the late stages of early embryonic development. The TEs MLT2A1 and MLT2A2 are specific to the 4- to 8-cell stages. The regulatory sequences contained in MLT2A1 and MLT2A2 are known to be potential targets of TFs such as DUX4, and OTX2 (Hendrickson et al., 2017; Liu et al., 2019), DUX4 activates HERVL by binding to the MLT2A1 element, which is a prominent event in ERV activation during EGA. The expression levels of LTR5\_Hs and MLT2A1 were shown in Figure 2F. The downregulated expression of these two TEs in AG embryos may strongly affect the process of early embryo development.

TABLE 1 Transcription factor with the highest connectivity within the significant module related to the stage.

Rank	Pink (2)	Purple (4)	Yellow (8)	Turquoise (morula)
1	FOXP4	ZSCAN5C	ZNF394	POU5F1
2	NFRKB	KLF18	ZNF473	ZNF581
3	TADA2B	DLX5	KLF17	GTF3A
4	ZNF646	ZNF705A	ZNF679	HMGA1
5	ZNF142	ZSCAN5A	TPRX1	BATF3
6	MTA1	ID3	MYC	ZNF592
7	ZBTB4	HES1	ZNF735	ZNF511
8	FOXP3	ID2	FOXP2	TEAD4
9	ZNF512B	ZNF705E	KLF10	ZNF764
10	ZBTB45	LEUTX	ZBTB11	MBD3

In summary, the transcription of the ERV superfamily and the Alu family differed greatly at the 4-cell stage among the three types of embryos. The expression level in uniparental embryos was lower than that in BI embryos, and the difference between AG embryos and BI embryos was greater.

## The expression of key transcription factors at the 4-cell stage differs greatly among the three types of embryos

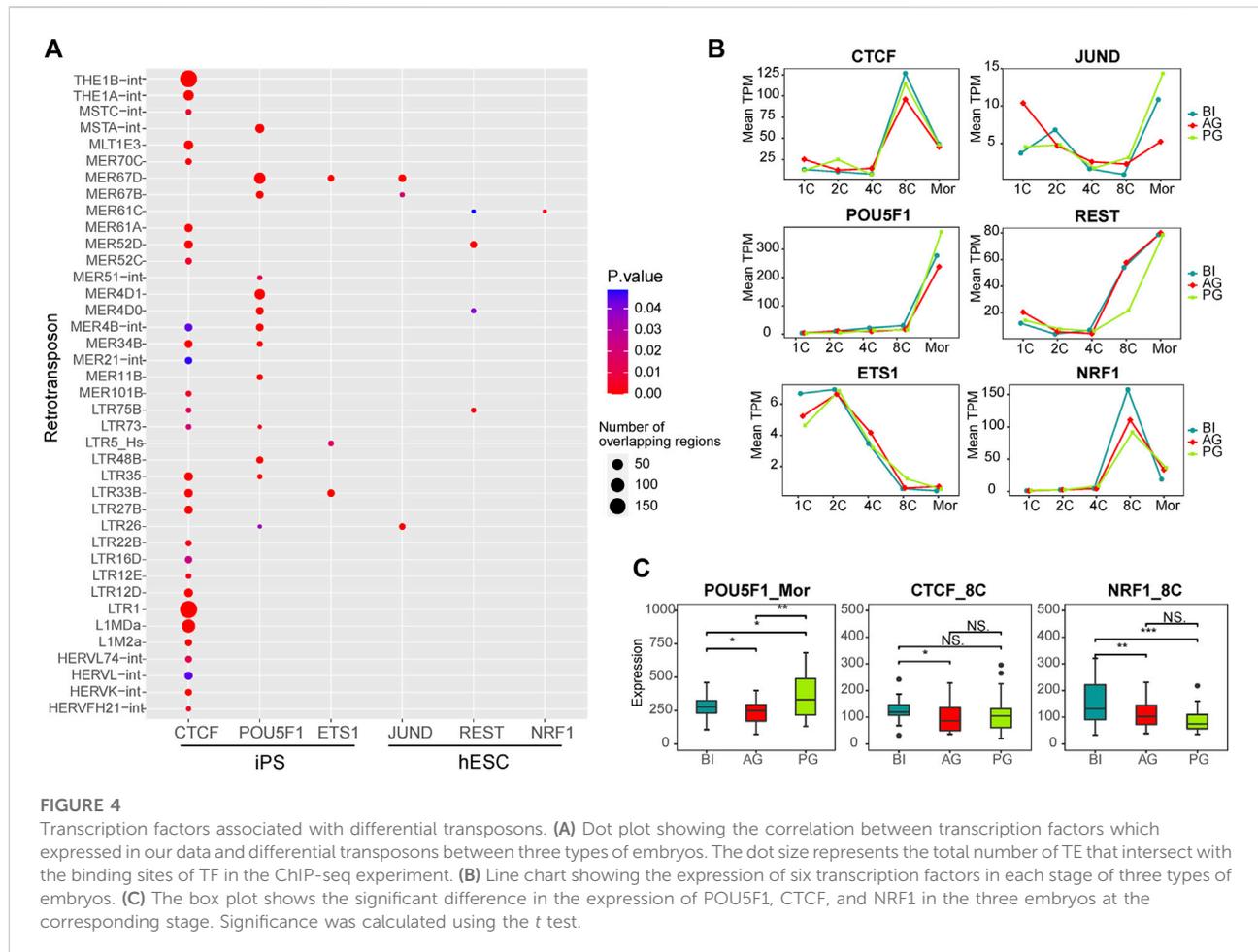
To obtain the interaction pattern of genes throughout early embryonic development, we constructed a coexpression network of all the genes to observe their regulation and try to identify important TF genes. We performed weighted gene correlation network analysis (WGCNA) on 14,066 genes after filtering out genes with low expression levels in BI embryos, WGCNA is an unbiased and unsupervised analysis method that can identify different coexpression modules corresponding to correlated transcripts. Notably, 5 out of 15 modules showed high stage-specific expression (correlation  $\geq 0.7$ ), these modules contain genes that tend to be overexpressed in a single stage of development (Figure 3A). Among them, the pink, purple, yellow, and turquoise modules were highly correlated with the 2-cell to morula stages (Supplementary Figures S3A,B), and the correlation was highly significant based on the  $p$  value ( $p < 10^{-20}$ ). Then we selected all TFs in these four modules and other genes that showed a high positive correlation with these TFs (weight  $> 0.1$ ) to draw network diagrams (Supplementary Figure S4). According to the results of GO function enrichment analysis (Supplementary Figure S5), genes in the pink module were related to histone modification and methylation, genes in the purple modules were related to the regulation of gliogenesis differentiation, genes in the yellow modules were related to RNA splicing and chromatin assembly, and genes in the turquoise modules were related to protein transport in the endoplasmic reticulum.

The top 10 TFs with the highest connectivity in each module are shown in Table 1. We compared the expression levels of all

40 TFs among the three types of embryos (Figure 3B). Interesting, the differences at the 2-cell and 4-cell stages were relatively regular. The expression of these TFs in AG and PG embryos was slightly lower than that in BI embryos at the 2-cell stage. All the 10 TF genes were greatly downregulated in AG embryos and eight transcription factors were upregulated in PG embryos compared with BI embryos. This suggests that the genomes of the two parental embryos have opposite effects on minor EGA at the 4-cell stage. Importantly, the number of significantly upregulated genes at the 1–8 cell stages (Supplementary Figure S2A) was positively correlated with the expression levels of TFs in the corresponding period, indicating that the degree of minor EGA is regulated by these key TFs. The difference in TF expression may contribute to insufficient reprogramming and premature differentiation of AG embryos. Further analysis of the expression characteristics and regulatory network of the key transcription factors ZSCAN5A and LEUTX, which have been examined in related reports, revealed that the expression of ZSCAN5A and LEUTX in AG embryos was significantly lower than that in PG and BI embryos at the 4-cell stage (Figures 3C,D). We also checked the expression of related genes coexpressed with ZSCAN5A at the 4-cell stage of the three types of embryos in the coexpression network and found that the expression of the coexpressed genes was consistent with ZSCAN5A (Supplementary Figure S6). This further illustrates the importance of key TFs at the 4-cell stage for the transcriptional activity of the minor EGA process.

## POU5F1 and CTCF may affect the development of uniparental embryos through differentially expressed transposable elements

Retrotransposons can act as enhancers to provide genes with cis-regulatory sequences. They influence the transcriptional activity of nearby genes by providing binding sites for TFs. To



further explore the correlation between TEs and TFs in early embryonic development, according to a report by Karakulah (2018), we downloaded data regarding the correlation between TFs and TEs in iPS cells and hESCs from the RTFADB website. Among them, the CTCF, POU5F1, and ETS1 transcription factors in iPS cells and JUND, REST, and NRF1 transcription factors in hESCs cells were not expressed at low levels in early embryonic cells. Therefore, we focused on the correlation between these six TFs and the differentially expressed TEs between the three types of embryos (Figure 4A; Supplementary Table S4). CTCF was associated with the most TEs, followed by POU5F1 (OCT4). Among them, the binding site of CTCF showed more overlap with LTR1, THE1B-int, L1MDa, THE1A-int, and MLT1E3 (180, 175, 90, 42, and 30, respectively). The binding site of POU5F1 showed 56 intersections with MER67D and 44 intersections with MER4D1, and these TEs included almost all of the ERV family.

Our WGCNA showed that POU5F1 is a key TF in the coexpression network in the morula stage, and CTCF and NRF1 are key TFs in the 8-cell stage coexpression network (Supplementary Figure S4). ChIP-Seq data from a previous

study showed that TEs account for 25% of the key TF binding sites of human embryonic stem cells, such as CTCF and POU5F1 (Kunarsø et al., 2010), POU5F1 is a known key regulator of embryonic stem cells and is related to pluripotency (Medvedev et al., 2008; Niakan and Eggan, 2013). CTCF is an important factor in the regulatory network (Hee Jung et al., 2019), which regulates chromatin structure in a variety of ways, including by isolating epigenetic transmission and participating in chromatin ring formation (Satou et al., 2015). The binding sites of CTCF are enriched in ERV elements (Ito et al., 2017), and their combination facilitates the organization of higher-order chromatin structures, thereby promoting the overall reconstruction of chromatin structure during preimplantation embryo development (Schmidt et al., 2012; Ke et al., 2017). Data from several sources have identified that NRF1 is related to early embryonic development, and defects in NRF1 are related to the decreased expression of various genes encoding centromeres and affecting mitosis, and may play a role in maintaining genome integrity (Oh et al., 2012; Williams et al., 2013; Zhang and Xiang, 2016; Sant et al., 2017). Therefore, we investigated their expression in the three embryos, POU5F1 was different at the

morula embryonic stage, while CTCF and NRF1 were different at the 8-cell stage (Figures 4B,C).

Taken together, the results suggest that the transcription factors POU5F1, and CTCF cooperate with the differentially expressed TEs of the ERV families such as LTR1, THE1B-int, THE1A-int, MLT1E3, and MER67D, as shown in Figure 4A, to regulate the major EGA process of early embryos. Their differential expression may affect the normal development of the three types of embryos.

## Discussion

Through the analysis of the single-cell transcriptome at each developmental stage of the early embryos, our research showed that TEs are transcribed in a stage-specific manner during the early development of BI, AG, and PG embryos. Among them, the expression of the ERV and the L1 family exhibit highly stage specific. ERVs are one of the most abundant transposable elements, accounting for 8% and 10% of the mouse and human genomes, respectively (Lander et al., 2001; Mouse Genome Sequencing Consortium Waterston et al., 2002). The expression of ERVs contributes to cellular plasticity and the activation of the embryonic genome, which is related to the establishment of pluripotency and totipotency (Lu et al., 2014). Studies have shown that knockout of L1 can lead to the failure of EGA and the maintenance of a 2-cell state in mouse preimplantation embryos (Percharde et al., 2018). In addition, other studies have shown that the activation of L1 after fertilization regulates the accessibility of chromatin in early mouse embryos (Jachowicz et al., 2017). Therefore, ERVs and L1 play important roles in the EGA process, and their expression may be a hallmark of cellular identity and cell potency that characterize the cell state in early human embryos.

The difference in TE expression among the three types of embryos was mainly reflected at the 4-cell stage. According to our data, ERVs are gradually expressed in the early stages of normal biparental embryos and have higher expression in 1- to 4- cell stage embryos, which suggests that the expression of ERVs is related to the minor EGA process. However, the transcription level of the TEs of uniparental embryos is lower than that of biparental embryos, especially AG embryos, such as LTR5\_Hs, MLT2A1, and other known important TEs of ERVs and the Alu family. It is known that most of the empty LTR elements in the human genome still maintain transcription and regulatory functions, affecting the expression of neighbouring genes (Rebollo et al., 2012). In mouse preimplantation embryos, ERVs account for a large part of the transcriptome of the minor EGA, and many important minor EGA-specific genes are regulated by LTR elements in ERVs (Peaston et al., 2004). The activation of MERVL in mice results in extremely high transcription as early as 8 h after fertilization (minor EGA) (Kigami et al., 2003). Downregulation of MERVL can cause developmental arrest at the 2-cell stage (Huang et al., 2017).

Although the transcription profiles of ERVs differ among species, the transcriptional activation of ERVs is a conserved event in early mammalian embryos (Rowe and Trono, 2011). Based on this, we speculated that the low expression of ERVs may affect the activation of the zygotic genome of uniparental embryos and the totipotency of embryo development.

According to our results, the expression of SINEs gradually increases with developmental stage. Ge reported that SINEs are associated with large-scale genome activation in early embryonic development in mice and humans, and the degree of activation is related to the position distribution and dosage of the SINE elements in the gene promoter (Ge, 2017). It has previously been observed that the expression and accessibility of the SINE families increase starting at the 8-cell stage in bovines, suggesting that these elements may act as promoters or enhancers (Halstead et al., 2020). As one of the most important and abundant SINE TEs that are still active in the human genome, Alu showed an indispensable regulatory function at the sequence level. Alu shares high sequence identity with the binding motifs of many important TFs and was proven to be bound by key TFs such as LEUTX, PITX2, and OTX2 (Polak and Domany, 2006; Töhönen et al., 2015; Jouhilahti et al., 2016). In addition, studies have shown that the Alu element is also an important CpG site provider. The GC-rich Alu sequence can introduce high GC content and methylation flexibility into the remote chromatin contact area, and regulate tissue-specific genes (Gu et al., 2016), showing its spatial role. Therefore, the changes in the expression of Alu and key TFs such as LEUTX, affect the normal development of early embryos.

The expression of key TFs such as ZSCAN5A and LEUTX at the 4-cell stage was lower in AG embryos, Sun et al. (2016) proved that knocking down ZSCAN5A can cause abnormalities in spindle assembly or attachment during mitosis, which can cause metaphase arrest and aneuploidy, leading to abnormal cell division (Sun et al., 2016). Based on the cleavage conditions of the three types of embryos, in the normally dividing embryos, the second cell division cycle of AG embryos is shorter and PG embryos are longer than BI embryos, showing the opposite trend (Leng et al., 2019). An implication of this is the possibility that the downregulation of ZSCAN5A at the 4-cell stage leads to a shorter second cleavage cycle in AG embryos, which leads to insufficient material reserves in the minor EGA process. In addition, studies have shown that LEUTX may be the main regulator of EGA, as 25% of the genes upregulated in 8-cell embryos were experimentally verified as LEUTX target genes. The expression of LEUTX at the 4-cell stage is critical to early embryonic development (Jouhilahti et al., 2016). Another study reported that the *de novo* motif overlapping with the Alu elements is similar to known consensus sequences of binding sites for PRD-like homeodomain containing TFs (Töhönen et al., 2015), such as LEUTX. Meanwhile, the expression level of Alu family TEs in AG and PG embryos was lower than that in BI embryos at the 4-cell stage, especially in AG embryos. Therefore, we speculate that the

downregulation of LEUTX and Alu retrotransposon elements in AG embryos plays an important role in transcriptional regulation. Further experiments and multiomics analysis are needed to prove this hypothesis. The molecular profiles of PG embryos, including the TF and TE profiles, are more similar to those of BI embryos, indicating that the maternal genome plays a more important role in regulating transcriptional activity during EGA. The longer duration of PG embryonic development than that of BI embryonic development may be caused by the lack of sperm in PG embryos, and the formation of mitotic centres may take longer (Bui et al., 2011; Escribá et al., 2016; Leng et al., 2019).

According to our statistics, the distribution of TEs in the promoter regions of genes in the regulatory networks at different stages has no obvious bias, indicating that the distribution of TEs may regulate the activation of the genome during the development of early embryos with epigenetic modifications such as methylation and chromatin accessibility. To further determine the relationship between the expression of a TE at a specific location and the expression of adjacent genes, the epigenetic regulation and chromatin state of the specific elements need to be determined. In addition, the current experimental sequencing methods used in TE research usually produce shorter reads (<150 bp). Given the highly repetitive sequence characteristics of TEs, it is necessary to align the sequences to specific positions with high reliability. In terms of experimental methods, we need longer sequencing read lengths, which can not only reduce the difficulty of alignment but also improve accuracy. In summary, the use of long-read sequencing methods, combined with the appearance of the chromatin open state and methylation, can provide a better understanding of the relationship between the expression of TE at a specific location and the expression of adjacent genes.

In summary, our systematic studies of the transcription trend and dynamic changes of TEs during early embryonic development revealed the influence of different TEs on the EGA process and provided a comprehensive regulatory framework for human early embryos. The differences in the expression of TEs and TFs among the three embryos were compared, and some important factors that may cause early embryo abnormal development were found, which will be helpful for analysing the molecular mechanism of early embryo development and may have an impact on developmental biology. It may also provide directions and ideas for follow-up research to improve the success rate of *in vitro* fertilization.

## Methods

### Sample information and data collection

To study the expression of TEs and TFs in the human early embryo we took advantage of downloaded single-cell RNA-seq public data from Gene Expression Omnibus (GEO) with accession number GSE133856 using the fastq-dump program

of SRAtools suite, which was generated by a previous project in our laboratory. This dataset contains 296 single cells from oocytes and embryos of the 1-,2-,4-,8-, morula-cell stage in BI, AG, and PG embryos. BI embryos acquired by Intracytoplasmic sperm injection (ICSI), AG embryos acquired by inducing two sperm into an enucleated oocyte, and PG embryos acquired by inhibiting second polar body extrusion. Embryos were cultured at 37.5°C in 6% CO<sub>2</sub>, 5% O<sub>2</sub>, and 89% N<sub>2</sub>. The culture medium was changed on day 3. To isolate individual embryo cells, the embryos were exposed to acidic Tyrode's solution for 3–5 s and then washed thoroughly in phosphate-buffered saline (PBS) containing 0.5% bovine serum albumin (BSA) to remove the zona pellucida. Zona-free embryos were incubated for 10 min (for the 2-, 4-, and 8-cell stages) or 15 min (for the morula stage) in Accutase medium, and then disaggregated by careful pipetting. Single blastomeres were placed into individual tubes containing 4 µl of lysis buffer or 0.5 µl of PBS for immediate preparation for RNA or DNA libraries, respectively. RNA-seq libraries were prepared using the SMARTSeq2 protocol, single-cell libraries were constructed using a Nextera XT DNA Library Preparation Kit (Illumina, Cat#FC-131-1096), and all libraries were sequenced on an Illumina HiSeq2500 or HiSeq X Ten instrument, according to the manufacturer's instructions (Leng et al., 2019).

### Data pre-processing

The initial quality check was carried out using FastQC (FastQC, 2020). Trimmomatic was used for trimming of sequences (Bolger et al., 2014), Human genome sequence (GRCh38) and annotation were downloaded from ENSEMBL, We used the STAR (Dobin et al., 2013) and HTSeq (Anders et al., 2015) programs to map and quantify gene expression. Because repetitive elements are distributed across many chromosome positions, to estimate the expression of TEs, we re-mapped reads using STAR to allow more multiple mapped reads using the following parameters: STAR–outFilterMultimapNmax 100 –winAnchorMultimapNmax 100 –outSAMmultNmax 100 –outSAMtype BAM SortedByCoordinate–outFilterMismatchNmax 3 (Ge, 2017). We used Tetranscripts (Jin et al., 2015), specifically designed to estimate both genes and TEs abundance, to calculate the expression level of repeated sequences by using an additional index of TEs based on UCSC repeatMasker files. The parameters we used are: Tetranscripts–format BAM–SortedByCoordinate –outFilterMismatchNmax 3mode multi–GTF genes. gtf–TE GRCh38\_rmsk\_TE.gtf –i 10 –stranded no. Because Repetitive elements have multiple duplicates of varying lengths, our general strategy was not to use length correction in RNA-seq. The raw counts of TE were normalized on the total number of mapped reads and multiplied by 1,000,000 obtaining expression values indicated as counts per

million (CPM) and the counts of gene were normalized into transcripts per million (TPM).

Here we collected 285 cells in total, including nine oocyte cells, 114 BI cells, 89 AG cells, and 73 PG cells ([Supplementary Table S5](#)). To select TEs and genes with a reproducible expression among the replicates of the same cell type, we selected TE and mRNAs with an expression value  $\geq 1$  (CPM or TPM) in at least 70% of replicates of at least 1 cell type. Finally, we identified 912 expressed TEs and 14,066 expressed genes for subsequent analysis. We also retrieved TE expression values in mice from the [Supplementary Table S3](#) of the paper published by Steven Xijin Ge ([Ge, 2017](#)) and collected two other human datasets GSE36552 and GSE44183 (the scRNA-seq data published by [Yan et al. \(2013\)](#) and [Xue et al. \(2013\)](#), respectively) for verifying our conclusion ([Supplementary Table S5](#)).

## Stage specific analysis

First, Principal component analysis (PCA) was performed to cluster biparental and uniparental embryos using all expressed TEs. The Wilcoxon test is a non-parametric statistical test used to compare two paired groups. The goal of the test is to determine whether two or more pairs are statistically different. Stage-specificity of individual elements was estimated using the Wilcoxon test on the CPM normalized data, every developmental stage was tested against all other stages. TEs with  $p$  value  $< 0.001$  by the Wilcoxon test were selected as stage-specific. The PCA plot was generated using R on the top 60 TEs that showed the strongest variance between the different developmental stages, with a total of 309 elements.

## Fuzzy clustering analysis and differential analysis of transposable element

We calculated the mean value of the expression of TEs for the biological replicates then made a log transformation, which was used as an input for Fuzzy analysis. The R package Mfuzz ([Kumar and Mfuzz, 2007](#)) was used for clustering analysis. Before clustering, we removed TEs with normalized values  $< 0$  at all stages. Finally, we divided all 912 TEs into 4 clusters and defined the core elements of a cluster with a membership value greater than 0.7. DESeq2 ([Love et al., 2014](#)) package was used to detect differential expression TEs in six consecutive developmental stages from oocyte to morula.

## Gene co-expression network analysis

All 14,066 filtered genes in BI embryo samples were applied to construct the co-expression network using the WGCNA package ([Langfelder and Horvath, 2008](#)), and then we selected TFs and

their mainly related genes in the modules whose correlation with the stages is greater than 0.7 for visualization. All known human TFs are downloaded from AnimalTFDB3.0. And the visualization of the core TF regulatory network was performed by Cytoscape software ([Shannon et al., 2003](#)). Moreover, the top 10 hub TFs in each module significantly related to the corresponding stage were performed Gene Ontology annotation using the ClusterProfiler package ([Ashburner et al., 2000](#); [Yu et al., 2012](#)).

## Associations between transposable elements and transcription factors

The associations between TEs and TFs in human is downloaded from RTFADB ([Karakulah, 2018](#)), which is a public repository of the overrepresented retrotransposon species including LTR retrotransposons, LINEs, and SINEs in the binding sites of the human and mouse TFs. By using ChIP-seq binding profiles of more than 3,000 transcription factors collected from human and mouse samples, RTFADB can search for more than 1500 retrotransposons on the binding sites of a total of 596 transcription factors. Here, we selected the data of the human embryonic stem cell and iPS cell that were closest to the early embryo type.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

## Author contributions

YL, GL, and FX conceived and designed this study. CL, YZ, and XP analyzed the data. LL, DZ, and XL performed the experiments. JH, YL, and LB supervised the project. CL designed and wrote the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was also supported by the Science, Technology, and Innovation Commission of Shenzhen Municipality under grant NO. JCYJ20200109150410232.

## Acknowledgments

The authors acknowledge with thanks to the Reproductive and Genetic Hospital of CITIC-Xiangya, Changsha, China, for

sharing data and giving a lot of support, thanks to the entire staff from LB Institute of Regenerative Medicine, BGI-Qingdao, BGI-Shenzhen, for their valuable work, thanks to China National Genebank for providing data storage and analysis platform.

## Conflict of interest

CL, YZ, XP, JH, LB, YL, and FX is employed by the BGI-Shenzhen.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcell.2022.1020490/full#supplementary-material>

### SUPPLEMENTARY FIGURE S1

Related to Figure 1. Validation of TE expression in human and mouse embryos from public data. (A) PCA result of early embryo data by Yan et al. (above), and the number of each transposon family in stage-specific transposons (below). (B) PCA results of early embryo data by Xue et al.

## References

- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi:10.1093/bioinformatics/btu638
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi:10.1038/75556
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170
- Braude, P., Bolton, V., and Moore, S. (1988). Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* 332, 459–461. doi:10.1038/332459a0
- Bui, H.-T., Wakayama, S., Mizutani, E., Park, K.-K., Kim, J.-H., Van Thuan, N., et al. (2011). Essential role of paternal chromatin in the regulation of transcriptional activity during mouse preimplantation development. *Reproduction* 141, 67–77. doi:10.1530/REP-10-0109
- Bui, L. C., Evsikov, A. V., Khan, D. R., Archilla, C., Peynot, N., Hénaut, A., et al. (2009). Retrotransposon expression as a defining event of genome reprogramming in fertilized and cloned bovine embryos. *Reproduction* 138, 289–299. doi:10.1530/REP-09-0042
- De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* 49, 941–945. doi:10.1038/ng.3858
- de Koning ApjGu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7, e1002384. doi:10.1371/journal.pgen.1002384
- Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 343, 193–196. doi:10.1126/science.1245316
- Deng, R., Han, C., Zhao, L., Zhang, Q., Yan, B., Cheng, R., et al. (2019). Identification and characterization of ERV transcripts in goat embryos. *Reproduction* 157, 115–126. doi:10.1530/REP-18-0336
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). Star: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635
- Dobson, A. T., Raja, R., Abeyta, M. J., Taylor, T., Shen, S., Haqq, C., et al. (2004). The unique transcriptome through day 3 of human preimplantation development. *Hum. Mol. Genet.* 13, 1461–1470. doi:10.1093/hmg/ddh157
- Escribá, M.-J., Escribá, L., Galiana, Y., Grau, N., Galán, A., and Pellicer, A. (2016). Kinetics of the early development of uniparental human haploid embryos. *Fertil. Steril.* 105, 1360–1368. doi:10.1016/j.fertnstert.2015.12.139
- Evsikov, A. V., de Vries, W. N., Peaston, A. E., Radford, E. E., Fancher, K. S., Chen, F. H., et al. (2004). Systems biology of the 2-cell mouse embryo. *Cytogenet. Genome Res.* 105, 240–250. doi:10.1159/000078195

(above), and the number of each transposon family in stage-specific transposons (below). (C) Fuzzy clustering analysis of TEs expression signals for the five consecutive stages in mouse. The closer the line color is to red, it means that these are the core TE in the cluster (left). Number of core TE of each class in each cluster. Select transposons with membership value greater than 0.7 as core TE in each cluster and count the number of four (membership  $\geq 0.7$ ) (right)

### SUPPLEMENTARY FIGURE S2

Related to Figure 2. (A) The number of upregulated genes at each stage of three types of embryos. (B) The number of transposons up-regulated from 2- to 4-cell stage in BI embryos overlapped with the transposons down-regulated in uniparental embryos than in BI embryos at the 4-cell stage.

### SUPPLEMENTARY FIGURE S3

The four modules most relevant to the 2-cell to the morula stage. (A) Scatterplot of gene significance (GS) for stages versus module membership (MM) in the four modules. GS and MM exhibit a very significant correlation, implying that hub genes of the four modules also tend to be highly correlated with the corresponding stage. (B) The expressions of eigengene in the four modules in all samples. The eigengenes of the four modules are all highly expressed in the corresponding stages.

### SUPPLEMENTARY FIGURE S4

The network diagrams of the four modules relevant to the 2-cell to the morula stage. The dot plot shows the comparison of the expression of the top ten TFs among the four modules in the three types of embryos (a, b, c, and d for 2-, 4-, 8-, and morula, respectively). The color of the dot represents the type of embryo, and the shape represents the stage.

### SUPPLEMENTARY FIGURE S5

GO function enrichment analysis of genes in the four modules. Genes in pink module are related to histone modification and methylation, genes in purple modules are related to regulation of gliogenesis differentiation, genes in yellow modules are related to RNA splicing and chromatin assembly, genes in turquoise modules are related to protein transport in the endoplasmic reticulum.

### SUPPLEMENTARY FIGURE S6

The expression of ZSCAN5A target genes in the three types of embryos at the 4-cell stage. The target genes here refer to the related gene of ZSCAN5A in the co-expression network. Significance was calculated using the t test.

- FastQC (2020). A quality control tool for high throughput sequence data - *ScienceOpen*. [Internet]. [cited 2020 Oct 20]; Available at: <https://www.scienceopen.com/document?vid=de674375-ab83-4595-afa9-4c8aa9e4e736>.
- Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5, 103–107. doi:10.1016/0168-9525(89)90039-5
- Ge, S. X. (2017). Exploratory bioinformatics investigation reveals importance of “junk” DNA in early embryo development. *BMC Genomics* 18, 200. doi:10.1186/s12864-017-3566-0
- Godini, R., and Fallahi, H. (2018). Dynamics changes in the transcription factors during early human embryonic development: GODINI and FALLAHI. *J. Cell. Physiol.* 234, 6489–6502. doi:10.1002/jcp.27386
- Goolam, M., Scialdone, A., Graham, S. J. L., Macaulay, I. C., Jedrusik, A., Hupalowska, A., et al. (2016). Heterogeneity in Oct4 and Sox2 targets biases cell fate in 4-cell mouse embryos. *Cell* 165, 61–74. doi:10.1016/j.cell.2016.01.047
- Grow, E. J., Flynn, R. A., Chavez, S. L., Bayless, N. L., Wossidlo, M., Wesche, D. J., et al. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 522, 221–225. doi:10.1038/nature14308
- Gu, Z., Jin, K., Crabbe, M. J. C., Zhang, Y., Liu, X., Huang, Y., et al. (2016). Enrichment analysis of Alu elements with different spatial chromatin proximity in the human genome. *Protein Cell* 7, 250–266. doi:10.1007/s13238-015-0240-7
- Halstead, M. M., Ma, X., Schultz, R. M., and Ross, P. J. (2020). Chromatin remodeling in bovine embryos indicates species-specific regulation of genome activation. *bioRxiv*.
- Hee Jung, Y., Kremysky, I., Gold, H., Rowley, M., Punyawai, K., Buonanno, A., et al. (2019). Maintenance of CTCF- and transcription factor-mediated interactions from the gametes to the early mouse embryo. *Mol. Cell* 75, 154–171. doi:10.1016/j.molcel.2019.04.014
- Hendrickson, P. G., Dorais, J. A., Grow, E. J., Whiddon, J. L., Lim, J.-W., Wike, C. L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and HERV1/HERV2 retrotransposons. *Nat. Genet.* 49, 925–934. doi:10.1038/ng.3844
- Hu, M., Tuanmu, L.-C., Wei, H., Gao, F., Li, L., and Zhang, S. (2015). Development and imprinted gene expression in uniparental preimplantation mouse embryos *in vitro*. *Mol. Biol. Rep.* 42, 345–353. doi:10.1007/s11033-014-3774-5
- Huang, Y., Kim, J. K., Do, D. V., Lee, C., Penfold, C. A., Zylicz, J. J., et al. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. *eLife* 6, e22345. doi:10.7554/eLife.22345
- Ito, J., Sugimoto, R., Nakaoka, H., Yamada, S., Kimura, T., Hayano, T., et al. (2017). Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *PLoS Genet.* 13, e1006883. doi:10.1371/journal.pgen.1006883
- Jachowicz, J. W., Bing, X., Pontabry, J., Bošković, A., Rando, O. J., and Torres-Padilla, M.-E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* 49, 1502–1510. doi:10.1038/ng.3945
- Jin, Y., Tam, O. H., Paniagua, E., and Hammell, M. (2015). TET transcripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* 31, 3593–3599. doi:10.1093/bioinformatics/btv422
- Jouhilahti, E.-M., Madisson, E., Vesterlund, L., Töhönen, V., Krjutskov, K., Reyes, A. P., et al. (2016). The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. *Development* 143, 3459–3469. doi:10.1242/dev.134510
- Karakulah, G. (2018). RTFADB: A database of computationally predicted associations between retrotransposons and transcription factors in the human and mouse genomes. *Genomics* 110, 257–262. doi:10.1016/j.ygeno.2017.11.002
- Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., et al. (2017). 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. *Cell* 170, 367–381. e20. doi:10.1016/j.cell.2017.06.029
- Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* 68, 651–654. doi:10.1095/biolreprod.102.007906
- Kumar, L., and Mfuzz, E. Futschik M. (2007). Mfuzz: A software package for soft clustering of microarray data. *Bioinformatics* 2, 5–7. doi:10.6026/97320630002005
- Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., et al. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* 42, 631–634. doi:10.1038/ng.600
- Lagutina, I., Lazzari, G., Duchi, R., and Galli, C. (2004). Developmental potential of bovine androgenetic and parthenogenetic embryos: A comparative study. *Biol. Reprod.* 70, 400–405. doi:10.1095/biolreprod.103.021972
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi:10.1038/35057062
- Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinform.* 9, 559. doi:10.1186/1471-2105-9-559
- Lee, M. T., Bonneau, A. R., and Giraldez, A. J. (2014). Zygotic genome activation during the maternal-to-zygotic transition. *Annu. Rev. Cell Dev. Biol.* 30, 581–613. doi:10.1146/annurev-cellbio-100913-013027
- Lee, M. T., Bonneau, A. R., Takacs, C. M., Bazzini, A. A., DiVito, K. R., Fleming, E. S., et al. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* 503, 360–364. doi:10.1038/nature12632
- Leichsenring, M., Maes, J., Mössner, R., Driever, W., and Onichtchouk, D. (2013). Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* 341, 1005–1009. doi:10.1126/science.1242527
- Leng, L., Sun, J., Huang, J., Gong, F., Yang, L., Zhang, S., et al. (2019). Single-cell transcriptome analysis of uniparental embryos reveals parent-of-origin effects on human preimplantation development. *Cell Stem Cell* 25, 697–712. doi:10.1016/j.stem.2019.09.004
- Liu, L., Leng, L., Liu, C., Lu, C., Yuan, Y., Wu, L., et al. (2019). An integrated chromatin accessibility and transcriptome landscape of human pre-implantation embryos. *Nat. Commun.* 10, 364. doi:10.1038/s41467-018-08244-0
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi:10.1186/s13059-014-0550-8
- Lu, X., Sachs, F., Ramsay, L., Jacques, P.-É., Göke, J., Bourque, G., et al. (2014). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* 21, 423–425. doi:10.1038/nsmb.2799
- Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., et al. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57–63. doi:10.1038/nature11244
- McGrath, J., and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* 37, 179–183. doi:10.1016/0092-8674(84)90313-1
- Medvedev, S. P., Shevchenko, A. I., Mazurok, N. A., and Zakiian, S. M. (2008). OCT4 and NANOG are the key genes in the system of pluripotency maintenance in mammalian cells. *Russ. J. Genet.* 44, 1377–1393. doi:10.1134/s1022795408120016
- Mouse Genome Sequencing Consortium Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562. doi:10.1038/nature01262
- Niakian, K. K., and Eggan, K. (2013). Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Dev. Biol.* 375, 54–64. doi:10.1016/j.ydbio.2012.12.008
- Niakian, K. K., Han, J., Pedersen, R. A., Simon, C., and Pera, R. A. R. (2012). Human pre-implantation embryo development. *Development* 139, 829–841. doi:10.1242/dev.060426
- Oh, D. H., Rigas, D., Cho, A., and Chan, J. Y. (2012). Deficiency in the nuclear-related factor erythroid 2 transcription factor (Nrf1) leads to genetic instability. *FEBS J.* 279, 4121–4130. doi:10.1111/febs.12005
- Park, C.-H., Uh, K.-J., Mulligan, B. P., Jeung, E.-B., Hyun, S.-H., Shin, T., et al. (2011). Analysis of imprinted gene expression in normal fertilized and uniparental preimplantation porcine embryos. *PLOS ONE* 6, e22216. doi:10.1371/journal.pone.0022216
- Peaston, A. E., Evsikov, A. V., Graber, J. H., de Vries, W. N., Holbrook, A. E., Solter, D., et al. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* 7, 597–606. doi:10.1016/j.devcel.2004.09.004
- Percharde, M., Lin, C.-J., Yin, Y., Guan, J., Peixoto, G. A., Bulut-Karslioglu, A., et al. (2018). A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell* 174, 391–405. e19. doi:10.1016/j.cell.2018.05.043
- Polak, P., and Domany, E. (2006). Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* 7, 133. doi:10.1186/1471-2164-7-133
- Ram, P. T., and Schultz, R. M. (1993). Reporter gene expression in G2 of the 1-cell mouse embryo. *Dev. Biol.* 156, 552–556. doi:10.1006/dbio.1993.1101
- Rebollo, R., Romanish, M. T., and Mager, D. L. (2012). Transposable elements: An abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* 46, 21–42. doi:10.1146/annurev-genet-110711-155621
- Rodriguez-Terrones, D., and Torres-Padilla, M.-E. (2018). Nimble and ready to mingle: Transposon outbursts of early development. *Trends Genet.* 34, 806–820. doi:10.1016/j.tig.2018.06.006
- Rowe, H. M., and Trono, D. (2011). Dynamic control of endogenous retroviruses during development. *Virology* 411, 273–287. doi:10.1016/j.virol.2010.12.007
- Sant, K. E., Hansen, J. M., Williams, L. M., Tran, N. L., Goldstone, J. V., Stegeman, J. J., et al. (2017). The role of Nrf1 and Nrf2 in the regulation of glutathione and redox dynamics in the developing zebrafish embryo. *Redox Biol.* 13, 207–218. doi:10.1016/j.redox.2017.05.023

- Satou, Y., Paola, M., Ishihara, K., Fukuda, A., Nosaka, K., Watanabe, T., et al. (2015). HTLV-1 inserts an ectopic CTCF-binding site into the human genome. *Retrovirology* 12, P12. doi:10.1186/1742-4690-12-s1-p12
- Schier, A. F. (2007). The maternal-zygotic transition: Death and birth of RNAs. *Science* 316, 406–407. doi:10.1126/science.1140693
- Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Gonçalves, Â., Kutter, C., et al. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148, 335–348. doi:10.1016/j.cell.2011.11.058
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303
- Sun, Y., Zhang, H., Kazemian, M., Troy, J. M., Seward, C., Lu, X., et al. (2016). ZSCAN5B and primate-specific paralogs bind RNA polymerase III genes and extra-TFHIC (ETC) sites to modulate mitotic progression. *Oncotarget* 7, 72571–72592. doi:10.18632/oncotarget.12508
- Tadros, W., and Lipshitz, H. D. (2009). The maternal-to-zygotic transition: A play in two acts. *Development* 136, 3033–3042. doi:10.1242/dev.033183
- Töhönen, V., Katayama, S., Vesterlund, L., Jouhilahti, E.-M., Sheikhi, M., Madisson, E., et al. (2015). Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nat. Commun.* 6, 8207. doi:10.1038/ncomms9207
- Vassena, R., Boué, S., González-Roca, E., Aran, B., Auer, H., Veiga, A., et al. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development* 138, 3699–3709. doi:10.1242/dev.064741
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi:10.1038/nrg2165
- Williams, L. M., Timme-Laragy, A. R., Goldstone, J. V., McArthur, A. G., Stegeman, J. J., Smolowitz, R. M., et al. (2013). Developmental expression of the Nfe2-related factor (Nrf) transcription factor family in the zebrafish, *Danio rerio*. *PLoS One* 8, e79574. doi:10.1371/journal.pone.0079574
- Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C., Feng, Y., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500, 593–597. doi:10.1038/nature12364
- Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1131–1139. doi:10.1038/nsmb.2660
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118
- Zhang, Y., and Xiang, Y. (2016). Molecular and cellular basis for the unique functioning of Nrf1, an indispensable transcription factor for maintaining cell homeostasis and organ integrity. *Biochem. J.* 473, 961–1000. doi:10.1042/BJ20151182