Rise and SINE: Roles of transcription factors and retrotransposable elements in zygotic genome activation of mammalian embryos

Pavel Kravchenko and Kikuë Tachibana

Department of Totipotency, Max Planck Institute of Biochemistry, Martinsried/Munich, Germany

Correspondence: K.T. tachibana@biochem.mpg.de

Abstract

In sexually reproducing organisms, life begins with the fusion of transcriptionally silent gametes, the oocyte and sperm. Although initiation of transcription in the embryo, known as zygotic genome activation (ZGA), is universally required for development, the transcription factors regulating this process are poorly conserved. In this Perspective article, we discuss recent insights into the mechanisms of ZGA in totipotent mammalian embryos. ZGA is regulated by several transcription factors including orphan nuclear receptors (OrphNRs) such as NR5A2, DUX, TPRXs and OBOX family proteins. We performed a meta-analysis to compile a list of pan-ZGA genes and found that most of these genes are targets of the above transcription factors. Remarkably, more than a third of these ZGA genes appear to be regulated by both OrphNRs such as NR5A2 and OBOX proteins, whose motifs co-occur in *SINE B1* retrotransposable elements, which are enriched near ZGA genes. We propose that ZGA in mice is activated by recruitment of multiple transcription factors to *SINE B1* elements that function as enhancers and discuss a potential relevance of this mechanism to *Alu* retrotransposable elements in human ZGA.

[H1] Introduction

All sexually reproducing multicellular organisms face the challenge of building a complex organism from a single fertilized egg (zygote). This seemingly simple single cell has plasticity of the highest level known as totipotency. The strict definition of totipotency is the developmental potential to generate all cell types, including extraembryonic tissues, and thus a whole organism (toti is the Latin word for 'whole')[1]. Reprogramming of the terminally differentiated gametes to totipotency is thought to occur within hours after fertilization, and the factors mediating reprogramming are provided as RNA and/or protein by the oocyte. This insight arose from somatic cell nuclear transfer experiments

by John Gurdon, in which transplantation of a nucleus from a somatic cell into an oocyte whose nucleus had been destroyed resulted in a viable and fertile frog[2,3]. Therefore, maternally provided factors from the oocyte can reprogram chromatin from sperm or potentially from any differentiated cell, with varying efficiencies, into a totipotent state. In mice, zygotic reprogramming is efficient (in natural mating, 8–10 oocytes are ovulated and nearly all are fertilized[4]) and occurs within hours, in contrast to induced reprogramming of somatic cells to pluripotency **[G]** by expression of the "Yamanaka" transcription factors **[G]**, which generates induced pluripotent stem cells (iPSCs) and is much less efficient (~0.01% of cells (Ref.[5]))[6]. This comparison highlights the power of the still largely unidentified reprogramming factors provided by the oocyte. The precise mechanism of how gamete reprogramming to totipotency is achieved remains one of the most fascinating mysteries in biology.

Several key processes are required to achieve totipotency after fertilization. The zygote forms by fusion of two terminally differentiated and transcriptionally silent gametes — egg and sperm. Sperm primarily contributes a pronucleus **[G]** to the embryo. By contrast, the egg provides the cytoplasm, including organelles and cytoskeleton, as well as a pronucleus. However, despite their similarity, the egg and zygote are fundamentally different. In particular, the egg is haploid and highly differentiated, whereas the zygote is diploid and totipotent. Importantly, the maternal-to-zygotic transition (MZT) occurs soon after fertilization and represents the handover of gene-expression control to the zygotic genome. The MZT includes the initiation of transcription from the zygotic genome, termed zygotic genome activation (ZGA), and the degradation of maternal RNAs and proteins[7]. In mammals, the MZT also entails epigenetic reprogramming of DNA and histones[8,9]. Thus, the proteome and transcriptome of the early embryo are distinct from that of either sperm or egg. How ZGA is initiated and regulated remains a major question (Box 1).

A hallmark of early embryos is the transcription of hundreds of thousands of copies of transposable elements (TEs) in mammalian genomes. Originally discovered in maize as "jumping genes" by Barbara McClintock, TEs are typically transcriptionally silent in somatic cells due to DNA methylation[10–13]. DNA demethylation in primordial germ cells and early embryos provides a window of opportunity for TE transcription and transposition into new genomic locations[14,15]. TEs can be classified according to their mechanism of transposition. Class I TEs are retrotransposons, which mobilize as an RNA intermediate and include endogenous retrovirus L (ERVL) long interspersed

nuclear elements (LINEs), and short interspersed nuclear elements (SINEs) in rodents and the related *Alu* elements in humans. TEs are thought to function as and drive the evolution of *cis*-regulatory elements[16]. Several TEs have been implicated in early development. Timely expression of *LINE-1* elements prior to ZGA promotes development[17,18], and depletion of *SINE B1* and of mouse ERVLs (MERVLs) causes development arrest at the morula stage[19–22], implying that TE transcripts or transcription might be involved in ZGA. Notably, the distribution of SINEs, but not LINEs, is conserved across species[23], and a bioinformatics analysis revealed that *SINE B1* is enriched near mouse ZGA genes[24]. TE perturbation experiments, however, are challenging because TEs can form chimeric transcripts with downstream genes[25] and are also found in introns, which may inadvertently cause the depletion of protein-coding mRNAs. Therefore, despite their expression in the early embryo, the precise roles of TEs in ZGA remains to be elucidated.

To initiate transcription of the silenced embryonic chromatin during ZGA, transcription factors and RNA polymerase II (Pol II) must overcome tight packing of nucleosomes. Transcription factor binding sites (TFBSs) can be (partially) concealed by the wrapping of DNA around histone octamers. This barrier can be overcome by nucleosome remodellers, by placeholder nucleosomes that may be more easily displaced, and by pioneer transcription factors (pioneer TFs)[26,27]. This class of transcription factors can bind their TFBSs in closed (nucleosome-occupied) chromatin and generate accessible chromatin by diverse and still poorly understood mechanisms. Pioneer factors have established roles in triggering ZGA in fruit flies, zebrafish and frogs, but the identity of these factors largely differs between species. Evidence has only recently emerged that orphan nuclear receptors (OrphNRs) with pioneer factor activity (such as the conserved nuclear receptor subfamily 5 group A member 2 (NR5A2)[28,29]), the species-specific transcription factors tetrapeptide repeat homeobox (TPRX)[30], and oocyte-specific homeobox (OBOX) protein families[31], are important for mammalian ZGA.

In this Perspective article, we discuss the timings of ZGA and its regulation by pioneer transcription factors. We highlight the recent discoveries of TFs that activate mouse and human ZGA and the unexpected convergence of their binding sites at *SINE B1* elements. We provide a cross-study comparison revealing common transcriptional targets of NR5A2 and OBOX proteins, which bind to *SINE B1*, suggesting that the recruitment of both is functionally important for ZGA. We propose

models of ZGA by transcription-factor recruitment to *SINE B1* elements. We direct the reader to excellent reviews on ZGA across species[26,32], epigenetic reprogramming[8], maternal RNA degradation[7], transposable elements[16] and pioneer factors[33] for further information.

[H1] The timing of ZGA

ZGA is not a single event of transcriptional "awakening". From invertebrates to humans, it can be broadly described to occur in at least two waves, with a minor wave that precedes a major wave[34,35]. The "minor" ZGA marks the onset of transcription from the embryo genome at dozens to hundreds of genes, and has been suggested to occur in an enhancer-independent manner[36,37]. Minor ZGA is dispensable for progression through the first cell cycle, but it is required for development beyond the second cell cycle in mouse embryos[38]. The ensuing "major" ZGA entails the activation of thousands of genes including developmental and housekeeping genes required for embryo self-sustainability, and the transcription of TEs including *LINE-1*, MERVLs and *SINE B1* (Ref.[36,38]). Hereafter, we refer to genes that are upregulated during the major ZGA as ZGA genes.

In general, the timing of these waves — both chronologically and in relation to nuclear divisions — is species-specific[34]. In *Caenorhabditis elegans*, the first zygotic transcripts are detected at the 4-cell embryo stage[39,40]. In the sea squirt *Ciona savignyi*, the first coordinated transcription begins between the 16-cell to 32-cell embryo stages. In organisms such as fruit flies, zebrafish and frogs, ZGA occurs after many rapid nuclear or cell divisions[26]. ZGA occurs between the 4-cell and 8-cell stages in humans, opossums and bovines[41–43]; between 8-cell and 16-cell stages in goats and in bovine cells in vitro[44–46] (8-cell to morula in yak cells in vitro[47]); between the 16-cell and morula stages in sheeps[48]; and at the 4-cell stage in pigs[49], although studies report different timings for the latter[50] (Fig. 1). Interestingly, in human embryos, the presence of both maternal and paternal genomes appears to affect the timing of ZGA compared to parthenogenetic (maternal only) and androgenetic (paternal only) embryos, although the underlying mechanisms remain unclear[51]. The mouse could be considered a "sprinter" in comparison to other mammals, with minor ZGA occurring in the zygote and major ZGA taking place already in the 2-cell embryo[52].

Several models have been proposed for the regulation of ZGA timing: 1) a dilution (nuclear-to-cytoplasmic (N/C) ratio) model, 2) a clock (activator-accumulation) model, and 3) a chromatin state

model. The dilution model proposes that a maternally-provided repressor of ZGA is diluted compared to the content of DNA with every cell division (and increasing the N/C ratio), eventually triggering ZGA[53–57]. Evidence for this model can be found in *Xenopus laevis, Danio rerio* and *Drosophila melanogaster*, where dilution of histone pools contributes to ZGA timing[58–61]. Whether dilution models can accurately describe timing mechanisms in mammals is less clear.

The clock model proposes that a molecular process such as translation of positive regulators of ZGA determines the timing of ZGA[53]. For example, the key regulator of ZGA in fruit flies, Zelda, is translationally upregulated upon fertilization. However, premature translation does not result in an earlier timing of ZGA[62], suggesting that a combination of the clock model and other mechanisms is likely involved in ZGA timing. Whether key regulators of ZGA in mammals are regulated in this fashion is not known.

The chromatin state model suggests that ZGA genes are marked for transcription activation by histone modifications associated with transcriptionally permissive chromatin or chromatin opening[63–65]. This model is supported by the requirement of Zelda, which is a pioneer factor, for chromatin opening prior to ZGA in the fruit fly[66–68]. In zebrafish, histone modifications associated with gene expression are present both prior to and during ZGA[69], implying a pre-patterning of developmental gene expression. The three models are not mutually exclusive, and it is likely that multiple mechanisms contribute to ZGA in different organisms and developmental contexts.

The two-wave view of ZGA is an over-simplification of the process. For example, some microRNAs are transcribed even prior to minor ZGA in zebrafish[71,72]. In humans, although ZGA occurs at the 4-cell to 8-cell stages, some genes appear to be transcribed in zygotes, according to microarray data[73]. Whether these transcripts can be detected by other methods and whether they have specific functions remains to be determined. Nevertheless, it is conceivable that the waves of ZGA might be further subdivided as nascent RNA-seq data of more time points is produced in the future.

The timings of ZGA and DNA replication coincide, raising the possibility that they are mechanistically linked[52,74–76]. Perturbations of DNA replication have caused opposing effects on ZGA, which could be partly due to activation of cell cycle checkpoints[77–80]. Another possibility is that ZGA timing is regulated by signaling pathways such as retinoic acid signaling[81].

[H1] Pioneer transcription factors regulate ZGA

ZGA in the totipotent embryo is followed by early developmental programs, but how is this transcriptional "awakening" initiated? Groundbreaking work on ZGA was carried out in fruit flies, where the pioneer TF Zelda binds to regulatory elements of ZGA genes prior to ZGA and is required for the MZT[66,82]. For several years, it was thought that ZGA is activated predominantly by one pioneer factor, Zelda. This model has been revised in light of evidence that Zelda cooperates with GAF and CLAMP[83–85], which also have pioneer factor activities, in regulating the majority of ZGA genes in *D. melanogaster* embryos. GAF and CLAMP transcription factors are highly conserved inside the order of Diptera[86,87], with Zelda being conserved across the clade of Pancrustacea[88], raising the question of ZGA regulation in other species lacking these factors and suggesting that ZGA is regulated by species-specific factors.

Zelda's role in ZGA has become a paradigm in ZGA research and seminal work in *D. rerio* and *X. laevis* has shown that pioneer factors are also crucial in vertebrates. In these species, the pluripotency-related pioneer factors of the NANOG homeobox, OCT, and SOX families are required to initiate ZGA[89,90]. Nanog, Pou5f1 (also called Oct4) and SoxB1 were identified to function in a combinatorial manner in *D. rerio*[89,91,92], and Pou5f3 and Sox3 were also suggested to have a cooperative effect in *X. laevis*[90]. Therefore, an emerging theme is that many of the early zygotic genes are co-regulated by multiple transcription factors. As will become apparent below, we think that co-regulation of zygotic genes is also a conserved feature in mammals, where it occurs through the recruitment of multiple transcription factors to retrotransposable elements.

As the NANOG, OCT and SOX families are conserved in mammals and exhibit similar combinatorial binding in mouse and human embryonic stem cells[93], it was hypothesized that these pioneer factors also regulate mammalian ZGA. However, oocyte-specific deletion of the mouse *Oct4a* gene variant did not prevent early mouse development[94]. This finding was surprising, because OCT4A was thought to be essential for ZGA. Importantly, Western blot analysis of 400 conditional knockout oocytes showed no detectable residual OCT4 protein that might have masked an essential phenotype, indicating that OCT4A is dispensable for ZGA in mice, in sharp contrast to fish and frogs[95]. The essential pioneer factors that trigger mammalian ZGA are therefore presumably distinct from those that activate teleost and amphibian ZGA.

Unlike transcription factors in general, pioneer factors can recognize and bind their motifs in the context of a nucleosome and facilitate "unwrapping" of the DNA from nucleosomes by diverse mechanisms, including the recruitment of chromatin remodelers and secondary transcription factors, to generate accessible chromatin[96] (Fig. 2). This mode of function is presumably also shared by pioneer factors that regulate ZGA . Consistent with this, BRG1, the catalytic subunit of the ATP-dependent chromatin remodeling SWI/SNF complexes, is required for ZGA[97].

The ability to remodel chromatin can also be influenced by nucleosome composition. In zebrafish, so-called placeholder nucleosomes, which contain the histone variant H2A.Z and the histone modifications H3 Lys4 mono-methylation (H3K4me1) and H3K27 acetylation (H3K27ac), mark positions that will become accessible during ZGA[27,98]. In fruit flies, H2A.Z marks transcription start sites (TSSs) of ZGA genes and is required for ZGA[99]. Current data from mouse embryos suggests that H2A.Z is incorporated into chromatin at the time of ZGA[100]. In addition to chromatin remodelers, epigenetic modifiers such as CBP–p300 for H3K27 acetylation at promoters and enhancers are required to facilitate recruitment of Pol II to ZGA genes[101].

[H1] Transcription factors that regulate mammalian ZGA

An understanding of mammalian ZGA is lagging behind that of invertebrate and other vertebrate model organisms, because unbiased screens in mammalian embryos are technically extremely challenging. In addition to this technical limitation, the poor evolutionary conservation of pioneer factors has meant that mammalian ZGA regulators could not be discovered through gene homology. It is therefore remarkable that several transcription factors, which we discuss in this section, have been associated with regulation of mouse or human ZGA in recent years.

[H2] NFYA

Nuclear transcription factor Y subunit alpha (NFYA) was identified as a potential ZGA regulator based on searches of binding-motifs in DNase I hypersensitive sites in 2-cell mouse embryos[102]. DNase I hypersensitive mapping in embryos was made possible by the development of low-input DNase-seq. NFYA is the DNA binding subunit of a NFY complex that has pioneer factor activity and forms a histone-like structure upon binding to DNA[103,104]. Depletion of NFYA in oocytes and analysis

of transcription in 2-cell embryos showed that, depending on the sample, up to 15% of ZGA genes are downregulated in embryos lacking NFYA[102]. Despite this phenotype, these embryos can progress beyond the 2-cell stage. Both depletion and genetic deletion embryos (produced through heterozygous intercrosses **[G]**) arrest in the morula stage[102,105]. The late arrest could either be due to inefficient protein depletion, or indicate that the function of NFYA is essential in morula embryos, but not earlier for ZGA. At the time of NFYA's discovery, low-input methods had not been developed yet that would allow a characterization of transcription-factor–chromatin binding in embryos. Therefore, NFYA genomic binding sites and direct target genes during ZGA are presently unknown.

[H2] DUX and MERVL

Double homeobox (DUX) is a family of placental mammalian-specific double-homeodomain **[G]** proteins that had been studied in muscular diseases and later discovered to be expressed during the minor ZGA in mouse embryos[106]. Mouse and human *DUX* genes form long multi-copy arrays and deletion of all gene copies is challenging[107]. An interesting early observation was that DUX4 overexpression in human myoblasts activates not only germ line genes but also transcription of class I retrotransposons, including ERVL-MaLR (mammalian apparent LTR retrotransposons)[108] and MERVL (also called MuERV-L) in mouse embryonic stem cells (mESCs)[106,109,110], which are activated during ZGA[111–113]. DUX overexpression can convert mESCs into a 2-cell-like-state[114], which shares transcriptional similarities with early embryos including the activation of MERVL and expression of ZSCAN4, a commonly used marker of ZGA[115,116]. Similarly, overexpression of human DUX in iPSCs revealed enriched binding of DUX4 to TSSs of early ZGA genes and transposable elements, including human ERVL (HERVL)[106]. These results are consistent with a role for mouse and human DUX proteins in promoting ZGA-like transcription.

The cell-culture-obtained findings discussed above do not answer the question: is DUX required for ZGA *in vivo*? To test this, mouse zygotes were microinjected with CRISPR–Cas9 gRNAs targeting *Dux* and the *Dux*-null embryos with reduced Dux mRNA showed a decreased efficiency of blastocyst formation[115]. By contrast, whole-body *Dux*-null females are viable and largely fertile[117,118], reflecting that constitutive knockout embryos develop to term and grow into adult mice. These data suggest that ZGA can proceed in the absence of DUX. Similarly, *DUX4* depletion in human zygotes does not prevent development[119]. *DUX* is transcriptionally regulated by developmental pluripotency-associated

2 (DPPA2) and DPPA4 (Ref.[120]), and oocyte-specific deletion of these transcription factors is also compatible with pre-implantation development[121]. These genetic deletions of *DUX* and its transcriptional activators suggest that the DUX regulatory network is not essential for ZGA, even though DUX is expressed during this process. It is possible that DUX acts redundantly with other transcription factors.

[H2] NR5A2 and orphan nuclear receptors

NR5A2 and potentially other OrphNRs were recently implicated as regulators of mouse ZGA. NR5A2 (known as LRH-1 in mammalian hepatocytes[122,123] and as fushi tarazu in fruit flies[124]), is a conserved OrphNR that, as the family name suggests, has no known specific ligands. It is expressed in a variety of organs including in liver, pancreas, ovaries and testes. NR5A2 is highly expressed in primordial germ cells, where it has been difficult to deplete[125], suggesting that it could be a long-lived protein. NR5A2 is expressed in mESCs, where it is required together with the steroid hormone receptor ESRRB for maintaining pluripotency[126]. NR5A2 is also expressed in pre-implantation mouse embryos and its depletion had transcriptional effects at the morula stage[126]. Although the embryos express related OrphNRs, including NR2C2 and ESRRB, the functions of these proteins are not fully redundant as constitutive *Nr5a2* deletion produced through heterozygous intercrosses are embryonic lethal at E6.5–E7.5 or earlier[127–129].

De novo motif searches were used to explore the hypothesis that motifs of transcription factors regulating ZGA are enriched in the cis-regulatory regions of ZGA genes[28]. The searches uncovered a super motif (a combination of individual adjacent motifs in a specific order) comprising 6 motifs and having 90% similarity to *SINE B1*. The presence of *SINE B1* and *SINE B2* elements is correlated with the upregulation of nearby genes during ZGA[24]. Intriguingly, motif #1 in these *SINE B1* elements can be recognized by NR5A2 (Fig. 3). Perturbation experiments, including siRNA-mediated depletion, protein degradation and chemical inhibition by SR1848, indicate that NR5A2 is required for gene expression during major ZGA[28]. Chemical inhibition of NR5A2 resulted in a 2-cell arrest and downregulation of 72% of 2508 major ZGA genes, including OBOX3 and OBOX6. Notably, chemical inhibition of NR5A2 had a much stronger effect on ZGA suppression compared to NR5A2 depletion, possibly owing to various factors such as inefficient siRNA-mediated depletion of maternal NR5A2 protein, chemical inhibition of other, potentially-relevant OrphNRs, and/or non-specific inhibition of

unknown factors. Overall, the data suggest that the inhibitor SR1848 targets additional OrphNRs, such as NR5A1 (steroidogenic factor-like subfamily) and NR2C2 (retinoid X receptor-like subfamily), but not NR3B2 (ESRRB)[28]. Thus, given these data and the potential roles of closely related OrphNRs, hereafter we shall treat findings pertaining to NR5A2 as relevant to other OrphNRs.

Is NR5A2 a pioneer factor that directly regulates ZGA? Nr5a2 fulfills the hallmarks of a pioneer factor since it promotes chromatin accessibility in vivo, can bind to its motif on nucleosomes in vitro and partially unwraps nucleosomal DNA[28,29]. Low-input characterization of the chromatin binding profiles of NR5A2 revealed preferred binding to enhancer-like elements, 70% of which corresponds to annotated *SINE B1* that are upstream and near to ZGA genes[28]. *SINE B1* is enriched near ZGA genes[24] and NR5A2 binding to these elements provided the first experimental evidence of involvement of *SINE B1* in ZGA regulation[28]. For example, mouse NR5A2 binds upstream of the gene encoding the pluripotency regulator *Nanog*, which is transcribed during ZGA. NR5A2 also activates transcription of other OrphNRs such as NR2C2 and ESRRB, which led to a proposed model of "many from one" (*ex uno plura*), in which a positive feedback loop of OrphNRs activation leads to potent ZGA activation[28].

In line with this model, ESRRB and estrogen receptors are required for embryo development and may have an important role in the reprogramming of mESCs to the naïve state[130,131]. NR5A2 promotes chromatin accessibility at its binding sites in 2-cell embryos and the recombinant protein binds to its motif on nucleosomes, which are the defining characteristics of pioneer factors[28]. NR5A2 continues to bind to *SINE B1* during pre-implantation development[132] and its binding preference shifts to non-*SINE B1* sites in mESCs[133]. What determines the preferential binding of NR5A2 to *SINE B1* in totipotent embryos is not known.

Is NR5A2 essential for ZGA? Different approaches to perturbing NR5A2 have yielded conflicting results. Depletion of NR5A2 in mouse oocytes prior to fertilization (followed ~50 hours later by analysis of 2-cell embryos) resulted in a decrease in ZGA, albeit much less than under chemical inhibition[28]. A shorter duration of depletion or base editing of *Nr5a2* in 1-cell embryos followed by analysis (within 24 h) of 2-cell embryos showed little effect on ZGA and a decrease in NR5A2 protein abundance[132]. An oocyte-specific deletion of *Nr5a2* using Zp3-Cre had little effect on ZGA and progression beyond the 2-cell stage, suggesting that NR5A2 is not essential[127] and that the strong

suppression of ZGA by chemical inhibition[28,132] might potentially have occurred through inhibition of other OrphNRs. However, whether NR5A2 protein was fully depleted following Zp3-Cre expression in oocytes, as had been demonstrated with Western blotting of OCT4 (Ref.[94]), remains to be demonstrated. Alternatively, presence of a residual maternal NR5A2 protein may be sufficient to activate ZGA through the "many from one" mechanism. Whether NR5A2 alone or together with other OrphNRs regulates ZGA therefore awaits clarification with more acute and specific methods.

[H2] OBOX and TPRX proteins

OBOX transcription factors have been implicated in the MZT for more than two decades owing to their high expression levels in oocytes and early embryos in the mouse[134]. The OBOX family is derived from a PRD-homeodomain-containing protein known as cone-rod homeobox (CRX) and includes seven genes and numerous pseudogenes. Whereas *OBOX* genes are rodent-specific, other CRX-derived transcription factors known as TPRX1, TPRX2 and TPRXL (TPRXs) exist in humans. Both OBOX and TPRXL clusters occur at syntenic regions of the human and mouse genomes[135].

The human-specific TPRX1 was identified as a genome-activation-associated factor in human embryos[136,137] and is upregulated in human 8-cell-like cells[138]. TPRX1, TPRX2 and TPRXL were studied as potential regulators of human ZGA owing to their high expression before or during ZGA. TPRXL, and TPRX1 and TPRX2 (TPRX1/2), are expressed during the minor ZGA[30]. To test whether TPRXs are involved in ZGA in human embryos, discarded triploid (3 pronuclei) embryos were used in depletion experiments. A triple TPRX depletion caused a decrease of expression in 25% of ZGA genes and had variable effects on early development, including a 7–9 cell arrest in some experiments, which would be consistent with defective major ZGA[30]. The genomic binding sites of TPRXs in embryos remain unknown due to the paucity of, and challenges in working with human embryonic material. It therefore remains to be determined to what extent the observed transcriptional changes are direct effects of TPRXs binding.

Complementing this work, a leap in understanding the functions of OBOX transcription factors has recently been achieved in studies of mouse embryogenesis. OBOX1/3/5/7 are highly translated in oocytes and OBOX 3/4 are transcribed during ZGA in embryos[139]. The timely translation of these transcription factors provided a rationale for testing their functions in ZGA. However, it had initially

been challenging to find evidence of OBOX function in development, possibly due to functional redundancies between members of the OBOX family. For example, *Obox6*-null mice are viable and fertile[140]. To overcome potential redundancies, a genomic region of 1.2 Mb encompassing *Obox1/2/3/4/5/7* and pseudogenes was deleted in mice[139]. Embryos carrying this deletion arrested at the 2-cell to 4-cell stages, demonstrating that the 1.2 Mb genomic region encodes genes that are essential for development[139]. OBOX3 overexpression could rescue development of the *Obox1/2/3/4/5/7*-null embryos, suggesting that at least one OBOX protein is required. The contributions of individual family members remain to be determined. For simplicity, hereafter we generally refer to all OBOX proteins collectively as OBOX.

The *Obox1/2/3/4/5/7*-null embryos downregulated 32% of 65 minor ZGA genes and 48% of 1107 major ZGA genes, including NR5A2 (Ref.[139]). Furthermore, in the absence of OBOX, some genes that are normally expressed only at the 1-cell stage are transcribed in 2-cell embryos. Therefore, a complex picture emerges in which some ZGA genes require OBOX for expression whereas other ZGA genes may require OBOX for timely repression[139]. It is conceivable that the embryonic arrest occurs as a consequence of failure not only to activate but also to appropriately repress genes[31].

A key question is whether OBOX proteins are directly regulating ZGA. Similar to TPRXs, there are currently no data on the binding sites of endogenous OBOX proteins in mouse embryos. Chromatin binding profiles have been obtained for overexpressed FLAG-tagged OBOX, which could represent endogenous binding sites but could also reflect binding to accessible ectopic sites[28]. Interestingly, OBOX-FLAG binds to its consensus motif, which corresponds entirely to motif #5 of the super motif in *SINE B1*[28] (Fig. 3). Consistent with this finding, a bioinformatics analysis had implicated PRD-like transcription factors in ZGA-gene regulation through occurrence of their binding motifs in retrotransposable elements[24,141]. Whether OBOX, or their related TPRXs, are pioneer factors is not known.

[H2] Functional implications of transcription-factor binding to *SINE B1*

Taken together, these findings reveal that no single transcription factor has yet been shown to regulate both mouse and human ZGA. Perhaps this conclusion is not surprising considering the relatively sparse overlap between mouse and human ZGA genes[142–144]. Nevertheless, it is also conceivable that a direct comparison of genes between different species may not be sufficient and higher-order analyses

are required[145,146]. Although *SINE B1* are rodent-specific, they share a common ancestry with human *Alu* retrotransposable elements. *SINE B1* and *Alu* belong to a family of non-long terminal repeat *7SL*-derived retrotransposons that are derived from a *7SL* RNA gene. It is thought that an ancestral fossil *Alu* monomer arose and subsequently diverged from a *7SL* RNA sequence[148,149]. Initially, these TEs consisted of monomers, as remains the case for *SINE B1*. A fusion of two types of monomers formed a dimeric *Alu* element, as it is currently found in humans. Within species, retrotransposition resulted in species-specific target genes that evolved over time[150].[Notably, NR5A2 and TPRXs binding motifs are conserved in *Alu*[28] and in the Alu-containing retrotransposons *SINE-VNTR-Alu*[147], which raises the possibility that the mechanism of ZGA activation may be conserved at least between mouse and human.

Although TEs diverged also in sequence, with different subfamilies harbouring deletions, their mode of transcription could explain how ZGA-relevant regulatory aspects could have been preserved. 7SL-derived retrotransposons such as *Alu* are transcribed by Pol III, which is stabilized by transcription factor IIIC (TFIIIC), a factor that recognizes boxes A and B in the body of the transposons[151]. These boxes are highly conserved in eumetazoa[152,153]. Importantly, box A overlaps with the OBOX binding motif and box B overlaps with the NR5A2 binding motifs in *SINE B1* (Fig. 3 and Fig S1D4). This overlap suggests that the evolutionary pressures to transcribe TEs also maintained these TFBSs whilst retrotransposition resulted in species-specific expansion around ZGA genes and therefore divergence in transcription activation[151]. Notably, TFIIIC and the conserved boxes A and B are known to establish regulatory interactions with other transcription factors[154–156]. This suggests that eumetazoa-conserved nuclear receptors[28,157] and PRD-like transcription factors[141,158,159] have co-evolved with retrotransposons in multicellular organisms with functional implications for ZGA[153].

[H1] Models of ZGA by transcription-factor recruitment to *SINE B1* elements

Synthesizing these findings, we propose that *SINE B1* sequences function as enhancer-like[24,28] and potentially as promoters that provide robustness to the initiation of ZGA through multiple TFBSs. According to this hypothesis, *SINE B1* or the related human *Alu* elements consist of arrays of motifs that can each be bound by a member of different transcription factor families. One of these families would be the OrphNRs, including NR5A2, and another family would be OBOX. Members of other transcription factor families might bind to other motifs in *SINE B1* and *Alu* elements, but as it is unknown whether such families are required for ZGA, we shall not consider them here. Below, we

discuss how members of the OrphNR and OBOX families might cooperate in promoting ZGA (Fig. 4A). For simplicity we assume that one motif would only be bound by one member of these families.

The simplest model one could imagine is that binding of a single transcription factor to one of the motifs in *SINE B1* is sufficient to promote transcription of a ZGA gene. According to this scenario, occupancy of a single TFBS (e.g. motif 1 by NR5A2, or motif 5 by OBOX) is sufficient for transcription activation. If either motif or transcription factor is lost, then binding of another transcription factor binding to another motif in *SINE B1* would rescue ZGA and therefore the identity of the ZGA-activating transcription factor would not be important. However, this model is not consistent with the observation that OrphNR inhibition or *OBOX* deletion decreases ZGA. Therefore, we favour a second model, in which at least two different transcription factors bound to their TFBSs in one or more *SINE B1* element are required to activate a ZGA gene transcription. **[Au: Repetition.]**Similarly, a third model proposes that binding of two factors would have a quantitative rather than qualitative effect on transcription output, which would be more subtle to detect but could have an impact on development (Fig. 4A).

A more complex version of the second model is that multiple different TFBSs have to be occupied in a specific manner with respect to the nucleosome dyad symmetry to allow histone eviction[29,160] (Fig. 4B). This model is analogous cooperativity in DNA binding by two transcription factors, but adjusted to dyad coordinates instead of linear DNA[161–164]. In this model, the orientations and spacings of motifs relative to each other would be relevant for gene activation.

All of these models assume that each of the transcription factors has pioneer activity, i.e. are able to bind to motifs on nucleosomes. Alternatively, transcription factors might need to cooperatively bind TFBS on nucleosomes, or true pioneer TFs might bind to TFBS on nucleosomes and cause DNA distortions[163,165] that enable the subsequent binding of transcription factors. For example, the TFBS of OBOX and the second TFBS of NR5A2 are far from one another on a linear scale but nearby each other on the nucleosome dyad axis (Fig. 3, Fig. 4B). Thus, co-binding of OBOX and NR5A2 might be required for pioneer activity, or their sequential binding might facilitate subsequent nucleosome eviction. Currently, NR5A2 is the only transcription factor with demonstrated pioneer factor ability at *SINE B1*, and may therefore initiate nucleosome destabilization[28,29].

[H1] OrphNRs and OBOX proteins co-regulate a subset of ZGA genes

An important question is to what extent the current ZGA-implicated transcription factors can explain the genome-wide transcriptional gene activation of ZGA in mouse embryos. Do these factors regulate primarily distinct genes or do some of their targets overlap, which could be explained by motif co-occurrences in *SINE B1*? Perhaps somewhat surprisingly, the answers to these questions cannot be found in the literature, for several reasons. First, lists of ZGA genes have been independently defined by different laboratories with different log fold changes (logFC) and *P* value criteria. Secondly, experiments have been carried out in different mouse strains, which may affect some of the ZGA genes. Therefore, a direct cross-comparison of ZGA regulation by transcription factors from separate studies is not possible without further analysis.

In order to compare data across studies, we compiled a "pan-ZGA" gene list. In brief, we performed a meta-analysis of eight publicly available polyA^{+} RNA-seq datasets for four different mouse strains in parallel, to enable robust identification of ZGA genes across strains (Supplementary Table 1). We analyzed a total of 2355 ZGA genes and found 200 core genes that were upregulated at least 4-fold in 2-cell embryos compared to zygotes in all strain combinations. Using a leave-one-out strategy[166] (see formula in Supplementary Figure 1), which includes the core gene set and removes biases of ZGA genes specific to one strain, we defined a more inclusive list of 542 pan-ZGA genes that are upregulated across strains and utilized this list for further analysis (Supplementary Figure 1A and Supplementary Table 1).

NFYA, DUX, NR5A2 and OBOX have been reported to regulate mouse ZGA[28,31,102,106]. Remarkably, a total of 85% (459) of pan-ZGA genes are regulated by these transcription factors (or by respective family members), suggesting that the majority of early-embryogenesis transcriptional gene regulation is in the realm of the known (Supplementary Figure 1B,C). By examining the contributions of individual factors, it emerged that 78% of pan-ZGA genes are regulated by OrphNRs (including NR5A2), according to combined depletion and chemical inhibition experiments, whereas substantially fewer genes (11%) were affected by NR5A2 depletion alone. OBOX proteins regulate 40% and DUX 4.6% of pan-ZGA genes. As NFYA-regulated genes overlapped with only 2.2% of pan-ZGA genes, NFYA data were not included in subsequent analyses.

Notably, OrphNRs, OBOX and DUX factors do not regulate exclusively distinct genes. Interestingly, we found that approximately 34% of the pan-ZGA genes (186 out of 542) are regulated by both OrphNRs and OBOX (Supplementary Figure 1B,C). This overlap suggests a common regulation that could be explained by co-occurrence of the NR5A2 and OBOX TFBSs in *SINE B1* (motifs #1 and #5, respectively) that are enriched upstream of ZGA genes (Fig. 3 and Supplementary Figure 1D)[24,28]. Moreover, ZGA genes with more *SINE B1* copies show the greatest increases in chromatin accessibility and gene expression during ZGA (Supplementary Figure 1E,F). These analyses support the notion that transcription factor recruitment to *SINE B1* elements regulates ZGA.

We noticed that the maximum overlap of genes (n=1273) was observed for OrphNRs and OBOX on target genes that are not strictly categorized as pan-ZGA (Supplementary Figure 1C). It is conceivable that these are, in fact, ZGA genes, but potentially fall below the threshold (4-fold increase) of our analysis. Similarly, more than 4200 genes are regulated solely by OrphNRs according to chemical inhibition experiments combined with depletion data, and these are also currently not classified as pan-ZGA genes. In the future, it will be important to obtain a more definitive ZGA gene list; this endeavor should be facilitated by the development of single-cell nascent RNA-seq techniques.

Whereas OrphNR and OBOX transcription factors control overlapping sets of genes, DUX appears to largely regulate a distinct set of pan-ZGA genes. However, this analysis has not discerned between minor and major ZGA. In addition, the viability of the *Dux*-null embryos[115] suggests that redundant mechanisms might activate DUX target genes, which would lead to an underestimation of DUX-regulated genes.

[H1] Concluding remarks and future perspective

A molecular understanding of mammalian ZGA is beginning to emerge. The identification of DUX, OrphNRs (NR5A2), TPRXs and OBOX proteins as ZGA regulators is providing an unprecedented opportunity to assign ZGA genes into different regulatory categories. A comparison of the target genes of these transcription factors across studies revealed that they regulate up to 85% of ZGA genes. In addition, two subsets of ZGA genes emerge: those that are regulated by more than one transcription factor, and those that are predominantly regulated by one transcription factor. Regulation by multiple

transcription factors could arise from independent or from inter-dependent binding to co-occurring motifs within *SINE B1*. Notably, the pioneer TFs Zelda, GAF and CLAMP display inter-dependent binding at some regulatory regions that drive fruit fly ZGA[82–84], and Nanog, Pou5fl and SoxB1 show co-operativity in chromatin opening in a context-dependent manner during zebrafish ZGA[89,91,92]. Fewer ZGA genes are regulated by one transcription factor only, and these genes are likely more sensitive to the abundance of that transcription factor. How the remaining 15% of ZGA genes are regulated should be exciting to elucidate in the future[167].

An intriguing yet unexplained aspect of ZGA regulation is the apparent lack of conservation of pioneer transcription factors controlling ZGA across the eumetazoan taxon, the group of "true animals" that undergo gastrulation and possess specialized cells organized into tissues[64]. Given that ZGA is a prerequisite for life, how can its regulators have diverged during evolution? Alternatively, could 'ZGA' actually entail different processes of transcriptional gene activation? The early ZGA of mammals begs the question whether this is the same process as the ZGA of zebrafish and frog, both of which are regulated by conserved pluripotency-related transcription factors. In fact, in mice OCT4 and SOX2 become essential later in development, on embryonic days E4.5 (blastocyst stage) and E6.0 (implantation), respectively[168,169]. As null embryos of these genes are generated through heterozygous intercrosses and therefore their phenotypes emerge following RNA and protein run-down, it is possible that OCT4 and SOX2 function when they are first expressed in morula embryos. We propose that the ancestral ZGA occurred closer to this later stage. The early ZGA that is seen in mammals is a consequence of transposition of TEs that harbour TFBS of transcription factors that are provided by the oocyte. Within mammals, conservation of *SINE B1*, *Alu* and other TEs may bring to light the conservation of binding by transcription factors to these elements. We note that these evolutionary speculations cannot explain why ZGA is driven by species-specific pioneer factors in fruit flies. One possible reason is that Diptera (and nematode worms) adopted a system based on PIWI-interacting RNAs for transcriptional silencing of TEs during embryogenesis[170], whereas other species went through a process of zinc-finger-protein diversification in response to TE expansion. It will be interesting to learn which transcription factors regulate ZGA in non-model organisms.

What might be a driving force to trigger ZGA earlier during development? We speculate that this could be a competition between maternal and embryonic transcriptional programs. As a mother and

its oocyte share the same genome, it may be important for the mother to avoid premature transcription activation of the embryonic transcription program in the oocyte, yet a fertilized egg must activate its genome to develop into an embryo. In such a potential conflict, mother cells should suppress pioneer factor activity and the embryo should be able to effectively and rapidly overcome the suppression in order to undergo ZGA. As the maternal-to-embryonic conflict evolves, maternal inhibition becomes stronger as well as pioneer factor activation becomes more potent and eventually happens early on. Thus, evolutionary selection favouring embryos could result in early-onset ZGA by potent pioneer factors .

Multiple and possibly redundant pioneer factors triggering ZGA in embryos are an effective strategy to escape maternal control. Pioneer factors are efficient in finding their cognate TFBSs at low concentrations[171–173]. Different compositions of pioneer TFs provided by oocytes could result in some transcriptional variability in embryos as long as key genes are activated during ZGA. This could contribute to variability in ZGA gene lists between *Mus musculus* strains. Intra-species differences are less likely to be caused by rare genetic changes in the coding sequences of transcription factors and instead could result from minor changes (in DNA methylation[174], histone modifications[175,176] or chromatin accessibility[177]) in promoters or enhancers regulating pioneer factor expression. Epigenetic differences could result in the deposition of slightly different pioneer factor combinations in individual oocytes. Such a model could explain how, in NR5A2 chemical inhibition[28] and *OBOX* deletion experiments[31], a subpopulation of perturbed cells was able to progress to blastocyst and 4-cell stage, respectively, whereas the majority of embryos arrested or fragmented at earlier time points.

The recent identification of ZGA regulators in mouse and human embryos is paving the way for a mechanistic understanding of the transcriptional awakening at the beginning of life. The power of genetics will continue to provide crucial information on whether newly identified ZGA regulators are essential or on how they contribute to the process. Tackling the challenge of redundancy — whether a motif can be bound by a family of transcription factors or by different types of transcription factors — will require ever more sophisticated methodologies. Lastly, what determines the specificity of these transcription factors remains poorly understood and will need to be individually investigated for each transcription factor, retrotransposable element and species. It will be exciting to see whether general

principles of ZGA emerge from all the studied organisms or whether the regulators and mechanisms of ZGA have been reinvented many times over during evolution.

Note in added proof: During the revision of this article, additional transcription factors involved in aspects of ZGA were identified including Duxbl[178], Klf17[179], and YY1[180], which were not included in the analyses.

References

1. Condic, M. L. Totipotency: what it is and what it is not. *Stem Cells Dev.* **23**, 796–812 (2014).

2. Gurdon, J. B., Elsdale, T. R. & Fischberg, M. Sexually Mature Individuals of Xenopus laevis from the Transplantation of Single Somatic Nuclei. *Nature* **182**, 64–65 (1958).

3. Gurdon, J. B. & Uehlinger, V. 'Fertile' Intestine Nuclei. *Nature* **210**, 1240–1241 (1966).

4. Nakao, S. *et al.* Synchronization of the ovulation and copulation timings increased the number of in vivo fertilized oocytes in superovulated female mice. *PLoS One* **18**, e0281330 (2023).

5. Yu, J. & Thomson, J. A. Induced Pluripotent Stem Cells. in *Principles of Tissue Engineering* 581–594 (Elsevier, 2014).

6. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).

7. Sha, Q.-Q. *et al.* Dynamics and clinical relevance of maternal mRNA clearance during the oocyte-to-embryo transition in humans. *Nat. Commun.* **11**, 4917 (2020).

8. Eckersley-Maslin, M. A., Alda-Catalinas, C. & Reik, W. Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat. Rev. Mol. Cell Biol.* **19**, 436–450 (2018).

9. Ladstätter, S. & Tachibana, K. Genomic insights into chromatin reprogramming to totipotency in embryos. *J. Cell Biol.* **218**, 70–82 (2019).

10. Grow, E. J. *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225 (2015).

11. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).

12. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**,

204–220 (2013).

13. Rowe, H. M. & Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **411**, 273–287 (2011).
14. Chernyavskaya, Y. *et al.* Loss of DNA methylation in zebrafish embryos activates retrotransposons to trigger antiviral signaling. *Development* **144**, 2925–2939 (2017).
15. Gkountela, S. *et al.* DNA Demethylation Dynamics in the Human Prenatal Germline. *Cell* **161**, 1425–1436 (2015).
16. Fueyo, R., Judd, J., Feschotte, C. & Wysocka, J. Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* **23**, 481–497 (2022).
17. Jachowicz, J. W. *et al.* LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49**, 1502–1510 (2017).
18. Percharde, M. *et al.* A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **174**, 391–405.e19 (2018).
19. Sakashita, A. *et al.* Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nat. Genet.* **55**, 484–495 (2023).
20. Zhang, W. *et al.* Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. *Nucleic Acids Res.* **47**, 8485–8501 (2019).
21. Lu, J. Y. *et al.* Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res.* **31**, 613–630 (2021).
22. Yang, J., Cook, L. & Chen, Z. Systematic evaluation of retroviral LTRs as cis-regulatory elements in mouse embryos. *Cell Rep.* **43**, 113775 (2024).
23. Mouse Genome Sequencing Consortium *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
24. Ge, S. X. Exploratory bioinformatics investigation reveals importance of ‘junk’ DNA in early embryo development. *BMC Genomics* **18**, 200 (2017).
25. Modzelewski, A. J. *et al.* A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform

essential for development. *Cell* **184**, 5541–5558.e22 (2021).

26. Schulz, K. N. & Harrison, M. M. Mechanisms regulating zygotic genome activation. *Nat. Rev. Genet.* **20**, 221–234 (2019).
27. Patrick J. Murphy, Shan Fu Wu, Cody R. James, Candice L. Wike, Bradley R. Cairns. Placeholder Nucleosomes Underlie Germline-to-Embryo DNA Methylation Reprogramming. *Cell* **172**, 993–1006.e13 (2018).
28. Gassler, J. *et al.* Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **378**, 1305–1315 (2022).
29. Kobayashi, W. *et al.* Nucleosome-bound NR5A2 structure reveals pioneer factor mechanism by DNA minor groove anchor competition. *Nat. Struct. Mol. Biol.* **31**, 757–766 (2024).
30. Zou, Z. *et al.* Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. *Science* **378**, abo7923 (2022).
31. Ji, S. *et al.* OBOX regulates mouse zygotic genome activation and early development. *Nature* **620**, 1047–1053 (2023).
32. Wu, E. & Vastenhouw, N. L. From mother to embryo: A molecular perspective on zygotic genome activation. *Curr. Top. Dev. Biol.* **140**, 209–254 (2020).
33. Barral, A. & Zaret, K. S. Pioneer factors: roles and their regulation in development. *Trends Genet.* **40**, 134–148 (2024).
34. Vastenhouw, N. L., Cao, W. X. & Lipshitz, H. D. The maternal-to-zygotic transition revisited. *Development* **146**, (2019).
35. Pálfy, M., Joseph, S. R. & Vastenhouw, N. L. The timing of zygotic genome activation. *Curr. Opin. Genet. Dev.* **43**, 53–60 (2017).
36. Heyn et al. The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species. *Cell Rep.* **6**, 285–292 (2014).
37. Abe, K.-I. *et al.* The first murine zygotic transcription is promiscuous and uncoupled from splicing and 3’ processing. *EMBO J.* **34**, 1523–1537 (2015).

38. Abe, K.-I. *et al.* Minor zygotic gene activation is essential for mouse preimplantation development. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E6780–E6788 (2018).
39. Sulston, J. E. *et al.* The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol.* **100**, 64–119 (1983).
40. Robertson, S. & Lin, R. The Maternal-to-Zygotic Transition in C. elegans. *Curr. Top. Dev. Biol.* **113**, 1–42 (2015).
41. Braude, P., Bolton, V. & Moore, S. Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* **332**, 459–461 (1988).
42. Mahadevaiah, S. K., Sangrithi, M. N., Hirota, T. & Turner, J. M. A. A single-cell transcriptome atlas of marsupial embryogenesis and X inactivation. *Nature* **586**, 612–617 (2020).
43. Kues, W. A. *et al.* Genome-wide expression profiling reveals distinct clusters of transcriptional regulation during bovine preimplantation development in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 19768–19773 (2008).
44. Graf, A. *et al.* Fine mapping of genome activation in bovine embryos by RNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4139–4144 (2014).
45. Ling, Y.-H. *et al.* Identification of lncRNAs by RNA Sequencing Analysis During Pre-Implantation Developmental Transformation in the Goat. *Front. Genet.* **10**, 1040 (2019).
46. Li, Y. *et al.* Transcription profiles of oocytes during maturation and embryos during preimplantation development in vivo in the goat. *Reprod. Fertil. Dev.* **32**, 714–725 (2020).
47. Yu, T. *et al.* Single-cell RNA-seq and single-cell bisulfite-sequencing reveal insights into yak preimplantation embryogenesis. *J. Biol. Chem.* **300**, 105562 (2024).
48. Fang, D., Tao, W., Wang, J., Huang, F. & Gao, Q. Transcriptome analysis of sheep embryos in vivo based on single cell RNA-seq. *Research Square* (2022) doi:10.21203/rs.3.rs-1199760/v1.
49. Schoenbeck, R. A., Peters, M. S., Rickords, L. F., Stumpf, T. T. & Prather, R. S. Characterization of deoxyribonucleic acid synthesis and the transition from maternal to embryonic control in the 4-cell porcine embryo. *Biol. Reprod.* **47**, 1118–1125 (1992).

50. Kong, Q. *et al.* Lineage specification and pluripotency revealed by transcriptome analysis from oocyte to blastocyst in pig. *FASEB J.* **34**, 691–705 (2020).

51. Yuan, S. *et al.* Human zygotic genome activation is initiated from paternal genome. *Cell Discov* **9**, 13 (2023).

52. Aoki, F., Worrad, D. M. & Schultz, R. M. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev. Biol.* **181**, 296–307 (1997).

53. Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic Genome Activation in Vertebrates. *Dev. Cell* **42**, 316–332 (2017).

54. Lee, D. R., Lee, J. E., Yoon, H. S., Roh, S. I. & Kim, M. K. Compaction in preimplantation mouse embryos is regulated by a cytoplasmic regulatory factor that alters between 1- and 2-cell stages in a concentration-dependent manner. *J. Exp. Zool.* **290**, 61–71 (2001).

55. Jukam, D., Kapoor, R. R., Straight, A. F. & Skotheim, J. M. The DNA-to-cytoplasm ratio broadly activates zygotic gene expression in Xenopus. *Curr. Biol.* **31**, 4269–4281.e8 (2021).

56. Jevtić, P. & Levy, D. L. Nuclear size scaling during Xenopus early development contributes to midblastula transition timing. *Curr. Biol.* **25**, 45–52 (2015).

57. Newport, J. & Kirschner, M. A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675–686 (1982).

58. Chen, H., Einstein, L. C., Little, S. C. & Good, M. C. Spatiotemporal Patterning of Zygotic Genome Activation in a Model Vertebrate Embryo. *Dev. Cell* **49**, 852–866.e7 (2019).

59. Amodeo, A. A., Jukam, D., Straight, A. F. & Skotheim, J. M. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E1086–95 (2015).

60. Lu, X., Li, J. M., Elemento, O., Tavazoie, S. & Wieschaus, E. F. Coupling of zygotic transcription to mitotic control at the Drosophila mid-blastula transition. *Development* **136**, 2101–2110 (2009).

61. Kane, D. A. & Kimmel, C. B. The zebrafish midblastula transition. *Development* **119**, 447–456 (1993).

62. Larson, E. D. *et al.* Premature translation of the Drosophila zygotic genome activator Zelda is not

sufficient to precociously activate gene expression. *G3* **12**, (2022).

63. Hontelez, S. *et al.* Embryonic transcription is controlled by maternally defined chromatin state. *Nat. Commun.* **6**, 10148 (2015).
64. Jukam, D., Shariati, S. A. M. & Skotheim, J. M. Zygotic Genome Activation in Vertebrates. *Dev. Cell* **42**, 316–332 (2017).
65. Vallot, A. & Tachibana, K. The emergence of genome architecture and zygotic genome activation. *Curr. Opin. Cell Biol.* **64**, 50–57 (2020).
66. Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R. & Eisen, M. B. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* **7**, e1002266 (2011).
67. Schulz, K. N. *et al.* Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res.* **25**, 1715–1726 (2015).
68. Hug, C. B., Grimaldi, A. G., Kruse, K. & Vaquerizas, J. M. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**, 216–228.e19 (2017).
69. Lindeman, L. C. *et al.* Prepatterning of developmental gene expression by modified histones before zygotic genome activation. *Dev. Cell* **21**, 993–1004 (2011).
70. Meier, M. *et al.* Cohesin facilitates zygotic genome activation in zebrafish. *Development* **145**, (2018).
71. Hadzhiev, Y. *et al.* The miR-430 locus with extreme promoter density forms a transcription body during the minor wave of zygotic genome activation. *Dev. Cell* **58**, 155–170.e8 (2023).
72. Bhat, P. *et al.* SLAMseq resolves the kinetics of maternal and zygotic gene expression during early zebrafish embryogenesis. *Cell Rep.* **42**, (2023).
73. Asami, M. *et al.* Human embryonic genome activation initiates at the one-cell stage. *Cell Stem Cell* **29**, 209–216.e4 (2022).
74. Nakatani, T. *et al.* DNA replication fork speed underlies cell fate changes and promotes reprogramming. *Nat. Genet.* **54**, 318–327 (2022).
75. Cho, C.-Y., Kemp, J. P., Jr, Duronio, R. J. & O'Farrell, P. H. Coordinating transcription and replication

to mitigate their conflicts in early Drosophila embryos. *Cell Rep.* **41**, 111507 (2022).

76. Sonehara, H., Nagata, M. & Aoki, F. Roles of the first and second round of DNA replication in the regulation of zygotic gene activation in mice. *J. Reprod. Dev.* **54**, 381–384 (2008).
77. Zhang, M., Kothari, P., Mullins, M. & Lampson, M. A. Regulation of zygotic genome activation and DNA damage checkpoint acquisition at the mid-blastula transition. *Cell Cycle* **13**, 3828–3838 (2014).
78. Blythe, S. A. & Wieschaus, E. F. Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. *Cell* **160**, 1169–1181 (2015).
79. McCleland, M. L., Shermoen, A. W. & O'Farrell, P. H. DNA replication times the cell cycle and contributes to the mid-blastula transition in Drosophila embryos. *J. Cell Biol.* **187**, 7–14 (2009).
80. Davis, W., Jr & Schultz, R. M. Role of the first round of DNA replication in reprogramming gene expression in the preimplantation mouse embryo. *Mol. Reprod. Dev.* **47**, 430–434 (1997).
81. Iturbide, A. *et al.* Retinoic acid signaling is critical during the totipotency window in early mammalian development. *Nat. Struct. Mol. Biol.* **28**, 521–532 (2021).
82. Liang, H.-L. *et al.* The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. *Nature* **456**, 400–403 (2008).
83. Gaskill, M. M., Gibson, T. J., Larson, E. D. & Harrison, M. M. GAF is essential for zygotic genome activation and chromatin accessibility in the early embryo. *Elife* **10**, (2021).
84. Duan, J. *et al.* CLAMP and Zelda function together to promote zygotic genome activation. *Elife* **10**, (2021).
85. Colonnetta, M. M., Abrahante, J. E., Schedl, P., Gohl, D. M. & Deshpande, G. CLAMP regulates zygotic genome activation in Drosophila embryos. *Genetics* **219**, (2021).
86. Kuzu, G. *et al.* Expansion of GA Dinucleotide Repeats Increases the Density of CLAMP Binding Sites on the X-Chromosome to Promote Drosophila Dosage Compensation. *PLoS Genet.* **12**, e1006120 (2016).
87. Ventos-Alfonso, A., Ylla, G. & Belles, X. Zelda and the maternal-to-zygotic transition in cockroaches. *FEBS J.* **286**, 3206–3221 (2019).

88. Ribeiro, L. *et al.* Evolution and multiple roles of the Pancrustacea specific transcription factor zelda in insects. *PLoS Genet.* **13**, e1006868 (2017).

89. Veil, M., Yampolsky, L. Y., Grüning, B. & Onichtchouk, D. Pou5f3, SoxB1, and Nanog remodel chromatin on high nucleosome affinity regions at zygotic genome activation. *Genome Res.* **29**, 383–395 (2019).

90. Gentsch, G. E., Spruce, T., Owens, N. D. L. & Smith, J. C. Maternal pluripotency factors initiate extensive chromatin remodelling to predefine first response to inductive signals. *Nat. Commun.* **10**, 1–22 (2019).

91. Lee, M. T. *et al.* Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**, 360–364 (2013).

92. Miao, L. *et al.* The landscape of pioneer factor activity reveals the mechanisms of chromatin reprogramming and genome activation. *Mol. Cell* **82**, 986–1002.e9 (2022).

93. Göke, J. *et al.* Combinatorial binding in human and mouse embryonic stem cells identifies conserved enhancers active in early embryonic development. *PLoS Comput. Biol.* **7**, e1002304 (2011).

94. Wu, G. & Schöler, H. R. Role of Oct4 in the early embryo development. *Cell Regen* **3**, 7 (2014).

95. Wu, G. *et al.* Establishment of totipotency does not depend on Oct4A. *Nat. Cell Biol.* **15**, 1089–1097 (2013).

96. Frederick, M. A. *et al.* A pioneer factor locally opens compacted chromatin to enable targeted ATP-dependent nucleosome remodeling. *Nat. Struct. Mol. Biol.* **30**, 31–37 (2023).

97. Bultman, S. J. *et al.* Maternal BRG1 regulates zygotic genome activation in the mouse. *Genes Dev.* **20**, 1744–1754 (2006).

98. Hickey, G. J. *et al.* Establishment of developmental gene silencing by ordered polycomb complex recruitment in early zebrafish embryos. *Elife* **11**, (2022).

99. Ibarra-Morales, D. *et al.* Histone variant H2A.Z regulates zygotic genome activation. *Nat. Commun.* **12**, 7002 (2021).

100. Liu, X. *et al.* Hierarchical Accumulation of Histone Variant H2A.Z Regulates Transcriptional States and

Histone Modifications in Early Mammalian Embryos. *Adv. Sci.* **9**, e2200057 (2022).

101. Wang, M., Chen, Z. & Zhang, Y. CBP/p300 and HDAC activities regulate H3K27 acetylation dynamics and zygotic genome activation in mouse preimplantation embryos. *EMBO J.* **41**, e112012 (2022).
102. Lu, F. *et al.* Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development. *Cell* **165**, 1375–1388 (2016).
103. Nardini, M. *et al.* Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell* **152**, 132–143 (2013).
104. Oldfield, A. J. *et al.* Histone-Fold Domain Protein NF-Y Promotes Chromatin Accessibility for Cell Type-Specific Master Transcription Factors. *Mol. Cell* **55**, 708–722 (2014).
105. Bhattacharya, A. *et al.* The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation. *Cancer Res.* **63**, 8167–8172 (2003).
106. Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
107. Hewitt, J. E. *et al.* Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* **3**, 1287–1295 (1994).
108. Geng, L. N. *et al.* DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* **22**, 38–51 (2012).
109. Wang, J., Lu, X., Zhang, W. & Liu, G.-H. Endogenous retroviruses in development and health. *Trends Microbiol.* (2023) doi:10.1016/j.tim.2023.09.006.
110. Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).
111. de la Rosa, S., Del Mar Rigual, M., Vargiu, P., Ortega, S. & Djouder, N. Endogenous retroviruses shape pluripotency specification in mouse embryos. *Sci Adv* **10**, eadk9394 (2024).
112. Schoorlemmer, J., Pérez-Palacios, R., Climent, M., Guallar, D. & Muniesa, P. Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency.

Front. Oncol. **4**, 14 (2014).

113. Thompson, P. J., Macfarlan, T. S. & Lorincz, M. C. Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Mol. Cell* **62**, 766–776 (2016).
114. Yang, F. *et al.* DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces a Totipotent 2C-like State. *Cell Stem Cell* **26**, 234–250.e7 (2020).
115. De Iaco, A. *et al.* DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).
116. Eckersley-Maslin, M. A. *et al.* MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. *Cell Rep.* **17**, 179–192 (2016).
117. Chen, Z. & Zhang, Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat. Genet.* **51**, 947–951 (2019).
118. Bosnakovski, D., Gearhart, M. D., Ho Choi, S. & Kyba, M. Dux facilitates post-implantation development, but is not essential for zygotic genome activation†. *Biol. Reprod.* **104**, 83–93 (2020).
119. Vuoristo, S. *et al.* is a multifunctional factor priming human embryonic genome activation. *iScience* **25**, 104137 (2022).
120. Eckersley-Maslin, M. *et al.* Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev.* **33**, 194–208 (2019).
121. Chen, Z., Xie, Z. & Zhang, Y. DPPA2 and DPPA4 are dispensable for mouse zygotic genome activation and pre-implantation development. *Development* **148**, (2021).
122. Lambrecht, R. *et al.* Liver receptor homolog-1 (NR5A2) orchestrates hepatic inflammation and TNF-induced cell death. *Cell Rep.* **42**, 113513 (2023).
123. Becker-André, M., André, E. & DeLamarter, J. F. Identification of nuclear receptor mRNAs by RT-PCR amplification of conserved zinc-finger motif sequences. *Biochem. Biophys. Res. Commun.* **194**, 1371–1379 (1993).
124. Yu, Y. *et al.* The nuclear hormone receptor Ftz-F1 is a cofactor for the Drosophila homeodomain protein Ftz. *Nature* **385**, 552–555 (1997).

125. Hackett, J. A. *et al.* Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nat. Commun.* **9**, 1–12 (2018).

126. Wu, J. *et al.* The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 (2016).

127. Festuccia, N. *et al.* Nr5a2 is essential for morula development. *bioRxiv* 2023.01.16.524255 (2023) doi:10.1101/2023.01.16.524255.

128. Gu, P. *et al.* Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. *Mol. Cell. Biol.* **25**, 3492–3505 (2005).

129. Paré, J.-F. *et al.* The fetoprotein transcription factor (FTF) gene is essential to embryogenesis and cholesterol homeostasis and is regulated by a DR4 element. *J. Biol. Chem.* **279**, 21206–21216 (2004).

130. Bondesson, M., Hao, R., Lin, C.-Y., Williams, C. & Gustafsson, J.-Å. Estrogen receptor signaling during vertebrate development. *Biochim. Biophys. Acta* **1849**, 142 (2015).

131. Esrrb Unlocks Silenced Enhancers for Reprogramming to Naive Pluripotency. *Cell Stem Cell* **23**, 266–275.e6 (2018).

132. Lai, F. *et al.* NR5A2 connects zygotic genome activation to the first lineage segregation in totipotent embryos. *Cell Res.* **33**, 952–966 (2023).

133. Festuccia, N., Owens, N., Chervova, A., Dubois, A. & Navarro, P. The combined action of Esrrb and Nr5a2 is essential for murine naïve pluripotency. *Development* **148**, (2021).

134. Rajkovic, A., Yan, C., Yan, W., Klysik, M. & Matzuk, M. M. Obox, a family of homeobox genes preferentially expressed in germ cells. *Genomics* **79**, 711–717 (2002).

135. Wilming, L. G., Boychenko, V. & Harrow, J. L. Comprehensive comparative homeobox gene annotation in human and mouse. *Database* **2015**, (2015).

136. Madissoon, E. *et al.* Characterization and target genes of nine human PRD-like homeobox domain genes expressed exclusively in early embryos. *Sci. Rep.* **6**, 1–14 (2016).

137. Maeso, I. *et al.* Evolutionary origin and functional divergence of totipotent cell homeobox genes in eutherian mammals. *BMC Biol.* **14**, 1–14 (2016).

138. Taubenschmid-Stowers, J. *et al.* 8C-like cells capture the human zygotic genome activation program in vitro. *Cell Stem Cell* **29**, 449–459.e6 (2022).

139. Guo, Y. *et al.* *Obox4*secures zygotic genome activation upon loss of*Dux*. *bioRxiv* (2022) doi:10.1101/2022.07.04.498763.

140. Cheng, W.-C., Hsieh-Li, H. M., Yeh, Y.-J. & Li, H. Mice lacking the Obox6 homeobox gene undergo normal early embryonic development and are fertile. *Dev. Dyn.* **236**, 2636–2642 (2007).

141. Töhönen, V. *et al.* Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nat. Commun.* **6**, 8207 (2015).

142. Hu, K., Li, W., Ma, S., Fang, D. & Xu, J. The identification and classification of candidate genes during the zygotic genome activation in the mammals. *Zygote* 1–11 (2024).

143. Boroviak, T. *et al.* Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. *Development* **145**, (2018).

144. Zhai, Y. *et al.* Profiling the transcriptomic signatures and identifying the patterns of zygotic genome activation - a comparative analysis between early porcine embryos and their counterparts in other three mammalian species. *BMC Genomics* **23**, 772 (2022).

145. Wu, J. *et al.* Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 (2018).

146. Bentsen, M. *et al.* ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 (2020).

147. Pontis, J. *et al.* Primate-specific transposable elements shape transcriptional networks during human development. *Nat. Commun.* **13**, 7178 (2022).

148. Quentin, Y. Origin of the Alu family: a family of Alu-like monomers gave birth to the left and the right arms of the Alu elements. *Nucleic Acids Res.* **20**, 3397–3401 (1992).

149. Quentin, Y. Emergence of master sequences in families of retroposons derived from 7sl RNA. *Genetica* **93**, 203–215 (1994).

150. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu

sequences. *Nat. Genet.* **35**, 41–48 (2003).

151. Moqtaderi, Z. *et al.* Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. *Nat. Struct. Mol. Biol.* **17**, 635–640 (2010).
152. Nishihara, H., Plazzi, F., Passamonti, M. & Okada, N. MetaSINEs: Broad Distribution of a Novel SINE Superfamily in Animals. *Genome Biol. Evol.* **8**, 528–539 (2016).
153. Senft, A. D. & Macfarlan, T. S. Transposable elements shape the evolution of mammalian development. *Nat. Rev. Genet.* **22**, 691–711 (2021).
154. Male, G. *et al.* Architecture of TFIIIC and its role in RNA polymerase III pre-initiation complex assembly. *Nat. Commun.* **6**, 7387 (2015).
155. Ciesla, M., Skowronek, E. & Boguta, M. Function of TFIIIC, RNA polymerase III initiation factor, in activation and repression of tRNA gene transcription. *Nucleic Acids Res.* **46**, 9444–9455 (2018).
156. Sun, Y. *et al.* ZSCAN5B and primate-specific paralogs bind RNA polymerase III genes and extra-TFIIIC (ETC) sites to modulate mitotic progression. *Oncotarget* **7**, 72571–72592 (2016).
157. Miglioli, A. *et al.* Nuclear Receptors and Development of Marine Invertebrates. *Genes* **12**, (2021).
158. Katayama, S. *et al.* Phylogenetic and mutational analyses of human LEUTX, a homeobox gene implicated in embryogenesis. *Sci. Rep.* **8**, 17421 (2018).
159. Leidenroth, A. & Hewitt, J. E. A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC Evol. Biol.* **10**, 364 (2010).
160. Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22534–22539 (2010).
161. Ibarra, I. L. *et al.* Mechanistic insights into transcription factor cooperativity and its impact on protein-phenotype interactions. *Nat. Commun.* **11**, 124 (2020).
162. Echigoya, K. *et al.* Nucleosome binding by the pioneer transcription factor OCT4. *Sci. Rep.* **10**, 11832 (2020).
163. Dodonova, S. O. *et al. Nature* **580**, 669–672 (2020).
164. Mondal, A., Mishra, S. K. & Bhattacherjee, A. Nucleosome breathing facilitates cooperative binding of

pluripotency factors Sox2 and Oct4 to DNA. *Biophys. J.* **121**, 4526–4542 (2022).

165. Michael, A. K. *et al.* Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science* **368**, 1460–1465 (2020).

166. Lachenbruch, P. A. An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics* **23**, 639–645 (1967).

167. Kobayashi, W. & Tachibana, K. Awakening of the zygotic genome by pioneer transcription factors. *Curr. Opin. Struct. Biol.* **71**, 94–100 (2021).

168. Avilion, A. A. *et al.* Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126–140 (2003).

169. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998).

170. Wells, J. N. *et al.* Transposable elements drive the evolution of metazoan zinc finger genes. *Genome Res.* **33**, 1325–1339 (2023).

171. Hansen, J. L. & Cohen, B. A. A quantitative metric of pioneer activity reveals that HNF4A has stronger in vivo pioneer activity than FOXA1. *Genome Biol.* **23**, 221 (2022).

172. Shahein, A. *et al.* Systematic analysis of low-affinity transcription factor binding site clusters in vitro and in vivo establishes their functional relevance. *Nat. Commun.* **13**, 5273 (2022).

173. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).

174. Maurano, M. T. *et al.* Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Rep.* **12**, 1184–1195 (2015).

175. Sinha, K. K., Bilokapic, S., Du, Y., Malik, D. & Halic, M. Histone modifications regulate pioneer transcription factor cooperativity. *Nature* **619**, 378–384 (2023).

176. Wu, K. *et al.* Dynamics of histone acetylation during human early embryogenesis. *Cell Discov* **9**, 29 (2023).

177. Zhao, Y. *et al.* ‘Stripe’ transcription factors provide accessibility to co-binding partners in mammalian

genomes. *Mol. Cell* **82**, 3398–3411.e11 (2022).

178. Vega-Sendino, M. *et al.* The homeobox transcription factor DUXBL controls exit from totipotency. *Nat. Genet.* **56**, 697–709 (2024).

179. Hu, Y. *et al.* Maternal KLF17 controls zygotic genome activation by acting as a messenger for RNA Pol II recruitment in mouse embryos. *Dev. Cell* **59**, 613–626.e6 (2024).

180. Wang, T. *et al.* Single-cell multi-omics profiling of human preimplantation embryos identifies cytoskeletal defects during embryonic arrest. *Nat. Cell Biol.* **26**, 263–277 (2024).

181. Berger, M. F. *et al.* Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276 (2008).

182. Vorontsov, I. E. *et al.* HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic Acids Res.* **52**, D154–D163 (2024).

183. Rauluseviciute, I. *et al.* JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **52**, D174–D182 (2024).

184. Liu, B. *et al.* The landscape of RNA Pol II binding reveals a stepwise transition during ZGA. *Nature* **587**, 139–144 (2020).

185. Nie, X. *et al.* Maternal TDP-43 interacts with RNA Pol II and regulates zygotic genome activation. *Nat. Commun.* **14**, 1–12 (2023).

186. Li, T. D. *et al.* TDP-43 safeguards the embryo genome from L1 retrotransposition. *Sci Adv* **8**, eabq3806 (2022).

Acknowledgements

The authors would like to thank Georgii A. Bazykin for insightful discussions. We also thank Jan-Michael Peters and members of the Totipotency Department for critical reading of the manuscript. Work in the Tachibana laboratory is funded by the European Research Council grant ERC-CoG-818556 TotipotentZygotChrom (K.T.) and the Max Planck Society. K.T. is an honorary professor in the Department of Biology, Ludwig-Maximilians-University, Munich.

Author contributions

Both authors contributed equally to the writing of the manuscript. The analysis was performed by P.K.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Molecular Cell Biology thanks Antonio Giraldez and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Figure 1 | **Overview of mouse preimplantation development stages and onset of zygotic genome activation in selected mammalian species.** In mice, zygotic genome activation (ZGA) occurs as a minor wave in the zygote that is followed by a major wave in the 2-cell embryo[52]. The onset of transcription during ZGA occurs in distinct embryonic stages in other mammalian embryos: in the 4-cell embryo in pigs[49]; between the 4-cell and 8-cell stages in human, opossum and bovine *in vivo* developed[41–43]: between the 8-cell and 16-cell stages in goat and bovine *in vitro* development[44–46]; and between 16-cell to morula embryos in sheep[48]. The blue line conveys maternal mRNA degradation over the course of maternal-to-zygotic transition and the red line represents embryonic mRNA synthesis during ZGA. Part **a** adapted with permission from ref.[28], AAAS.

Figure 2 | **A model of transcriptional activation by pioneer transcription factors during zygotic genome activation.** A pioneer transcription factor (TF) binds to its transcription-factor binding site (TFBS) in closed chromatin and increases the accessibility of DNA by recruiting chromatin remodelers and other pioneer transcription factors. Accessible DNA can now be occupied by transcription factors, allowing transcription by RNA polymerase II (Pol II).

.

Figure 3 | ***SINE B1* contains multiple transcription factor motifs.** A sequence logo of a super motif comprising the strong motifs #1–6 identified in a *de novo* motif search[28]. The super motif has 90% similarity to the *SINE B1 YR5* sequence, which is a variant of the *SINE B1* retrotransposon [28]. Oocyte-specific homeobox (OBOX) proteins binding motif (motif #5)[31,181] and two copies of the nuclear receptor subfamily 5 group A member 2 (NR5A2) binding motif (motif #1)[182,183] are highlighted and compared to the reference motifs.

Figure 4 | **Models of zygotic genome activation by transcription-factor binding to *SINE B1* elements. a** | Models of independent and/or cooperative binding of transcription factors upstream of zygotic genome activation (ZGA) genes. Model (1) any of the transcription factors (TF) can activate transcription of a ZGA gene; (2) a combination of the two transcription factors is required to activate transcription; and (3) co-binding of the two transcription factors would result in elevated levels of transcription. **b** | Model of cooperative binding of OBOX and orphan nuclear receptor (OrphNR) factors to a *SINE B1* element on a nucleosome. Multiple specific transcription factor binding sites (TFBSs) need to be occupied to efficiently activate transcription. Binding of adjacent TFBSs could enable more-efficient nucleosome eviction due to close spatial localization of the two transcription factors on the nucleosome dyad axis. Part **b** adapted from ref.[28], Springer Nature.

Box 1 | **Transcription by RNA polymerase II during zygotic genome activation**

Recent advances in profiling the genomic binding sites of RNA polymerase II have shown that Pol II localization changes during maternal-to-zygotic reprogramming in mice[184]. Following fertilization, Pol II is recruited to chromatin in the S and G2 phases of the zygote, and is not only found at transcription start sites of minor zygotic genome activation **[G]** (ZGA) genes, but also at those of genes that will be activated during the major zygotic genome activation **[G]** in 2-cell embryos[184]. **[Au: Repetition.]** Pol II "pre-configuration" (a relocation to future gene targets before major ZGA) is dependent on minor ZGA, requires OBOX proteins and is necessary for development[31,184]. The maternally deposited RNA-binding protein TDP-43 has been implicated in Pol II recruitment to its genomic locations prior to ZGA[185] and in suppressing retrotransposition to protect the genome from transposable elements derepressed during ZGA[186].

Glossary

- Heterozygous intercrosses – A mating set-up between two animals that are heterozygous at a particular locus, designed to generate embryos with three different genotypes (e.g., wild-type, homozygous mutants and heterozygotes).
- Homeodomain – A conserved DNA-binding protein domain that is encoded by homeobox DNA sequence.
- Major zygotic genome activation – A process of transcriptional activation of thousands of genes occurring in early embryos that is essential for development. The timing depends on the species, e.g. it occurs at the 2-cell stage of mouse embryos.
- Minor zygotic genome activation – A process of transcriptional activation that precedes major ZGA and involves fewer genes. The timing depends on the species, e.g. it occurs at the 1-cell stage of mouse embryos.

- Pluripotency – The developmental potential of a cell to differentiate into any of the three germ layers, but not into extraembryonic tissues.
- Pronucleus – This refers to the nucleus in a zygote. Fertilization of an egg by sperm results in a maternal and a paternal nucleus, each typically harboring a haploid number of chromosomes.
- “Yamanaka” factors – The set of four transcription factors: OCT4 (POU5F1), SOX2, KLF4 and c-MYC that are sufficient to reprogram cells into induced pluripotent stem cells (iPSCs).