

Review

Pioneer Transcription Factors: The First Domino in Zygotic Genome Activation

Bo Fu ^{1,2,†}, Hong Ma ^{1,2,†} and Di Liu ^{1,2,*}
¹ Institute of Animal Husbandry, HeiLongJiang Academy of Agricultural Sciences, Harbin 150086, China; fubo@haas.cn (B.F.); hongma@haas.cn (H.M.)

² Key Laboratory of Combining Farming and Animal Husbandry, Ministry of Agriculture and Rural Affairs, Harbin 150086, China

* Correspondence: liudi@haas.cn

† These authors contributed equally to this work.

Abstract: Zygotic genome activation (ZGA) is a pivotal event in mammalian embryogenesis, marking the transition from maternal to zygotic control of development. During the ZGA process that is characterized by the intricate cascade of gene expression, who tipped the first domino in a meticulously arranged sequence is a subject of paramount interest. Recently, Dux, Obox and Nr5a2 were identified as pioneer transcription factors that reside at the top of transcriptional hierarchy. Through co-option of retrotransposon elements as hubs for transcriptional activation, these pioneer transcription factors rewire the gene regulatory network, thus initiating ZGA. In this review, we provide a snapshot of the mechanisms underlying the functions of these pioneer transcription factors. We propose that ZGA is the starting point where the embryo's own genome begins to influence development trajectory, therefore in-depth dissecting the functions of pioneer transcription factors during ZGA will form a cornerstone of our understanding for early embryonic development, which will pave the way for advancing our grasp of mammalian developmental biology and optimizing in vitro production (IVP) techniques.

Keywords: pioneer transcription factors; preimplantation embryo; zygotic genome activation; retrotransposons



Citation: Fu, B.; Ma, H.; Liu, D. Pioneer Transcription Factors: The First Domino in Zygotic Genome Activation. *Biomolecules* **2024**, *14*, 720. <https://doi.org/10.3390/biom14060720>

Academic Editor: Carl A. Pinkert

Received: 26 April 2024

Revised: 31 May 2024

Accepted: 17 June 2024

Published: 18 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Following fertilization, the genome within the totipotent embryo remains transcriptionally silent, then initiates bursts of transcription, known as zygotic genome activation (ZGA). ZGA signifies the developmental governance from maternal genetic materials to zygotic control. This transition encompasses not only the degradation of maternal constituents but also the activation of the zygotic genome across two or more distinct transcriptional phases. Taking the early mouse embryo as an example, ZGA begins with the minor ZGA, which occurs in the one-cell stage embryo, then followed by the major ZGA, which is more pronounced and occurs at the two-cell stage. In fact, the timing and progression of ZGA vary across species, which reflects the diversity in embryonic development strategies employed by different mammals [1,2]. The ZGA process initiates a series of tightly regulated molecular events that are essential for the progression of embryogenesis. In particular, chromatin remodeling occurs during ZGA, which facilitates the accessibility of the embryonic genome to the transcriptional machinery [3,4]. Within this complex transcriptional machinery, transcription factors, particularly a distinct subgroup known as pioneer transcriptional actors, play a pivotal role.

In the architectural configuration of the nucleosome, genomic DNA is wrapped around an octameric assembly of histones, presenting a formidable impediment to the binding of transcription factors. While architectural configuration of the nucleosome is crucial for the organization and integrity of the genome, this structural conformation poses a

significant challenge to transcriptional activation processes, including those associated with ZGA [5]. Pioneer transcription factors, unlike other transcription factors that require accessible chromatin, can bind directly to nucleosomal DNA within compacted chromatin, thereby initiating chromatin remodeling and enabling gene expression. The term “pioneer” refers to transcription factors that meet specific criteria: they bind to their (partial) motifs within closed chromatin, are essential for local chromatin opening in vivo, and can bind to nucleosomes in vitro [6].

Pioneer transcription factors discussed in this review, including Dux, Obox and Nr5a2, are endowed with the unique capability to recognize and bind to specific sequences within their target motifs amidst the compact architecture of closed chromatin. This initial docking phenomenon is pivotal, as it triggers a sequence of chromatin remodeling activities that culminate in decondensation of chromatin. These pioneer transcription factors exhibit remarkable proficiency in infiltrating chromatin territories typically characterized by their repressive state within a living organism, catalyzing the transition of chromatin from an inaccessible to an open state within the physiological milieu in vivo. Such a transformative action on the chromatin landscape is mediated through diverse mechanisms, with each contributing to the facilitation of a more transcriptionally permissive environment [7–11]. Accordingly, it has been believed that this intrinsic attribute of these pioneer transcription factors to act as a trigger of chromatin accessibility is particularly critical in initiating the transcriptional awakening of the zygote.

Currently, the elucidation of ZGA mechanisms within mammalian species remains comparatively nascent, especially when compared to the extensive knowledge amassed in model organisms such as fruit flies and zebrafish, whose pioneer transcription factors have already been subject to detailed study [12–16]; and the pioneer transcription factors for triggering ZGA in mammals still remain elusive. Fascinatingly, recent investigations have illuminated the complexity of transcriptional regulation in mammalian embryonic development, revealing that not only a single, but multiple pioneer transcription factors (including Dux, Obox and Nr5a2) bind to the genome to open chromatin, thereby triggering the initiation of ZGA [17–21]. Surprisingly, all these three pioneer transcription factors use retrotransposons as hubs to control ZGA gene expression, thus initiating the ZGA process. In this review, we summarized how Dux, Obox and Nr5a2 function as activators of ZGA in mammals and outlined the interaction between these pioneer transcription factors and retrotransposons. We anticipate that further elucidation of the mechanisms by which these pioneer transcription factors operate will offer a novel perspective for further understanding the ZGA process.

2. The Role of Dux in ZGA

Dux in mice or DUX4 in humans, the double homeobox (DUX) gene family members that are conserved in mammals, have been identified as a pioneer transcription factor that initiates ZGA. Using nanopore-based ultra-long read sequencing, Mitsuhashi et al. revealed that *DUX4* gene loci are located within the 3.3-kb repeats of the D4Z4 array at the subtelomeric region of chromosome 4q35 in human [22]. The results derived from DNA fluorescence in situ hybridization (FISH) also confirmed that there is a hybridization signal of *Dux* repeats on chromosome 10 in mice [23]. The double homeobox gene, such as DUX4 tandem repeats located within pericentromeric regions, generally carries characteristic heterochromatin, which serves to repress expression in the vast majority of somatic cell types. Once the pathological misexpression of *DUX4* from these tandem repeats occurs, the aberrant expression usually results in the transcriptional activation of the early embryonic program, and then causes facioscapulohumeral muscular dystrophy (FSHD) [24,25]. However, in the developmental context, mouse DUX and its human counterpart, DUX4, are both expressed before ZGA in their respective species [19,26]. Dux expression rapidly increases after fertilization and peaks at the two-cell stage. This temporal expression pattern indicates that the transient expression of Dux may underscore a pivotal role in the orchestration of ZGA [27]. For instance, with regard to somatic cell nuclear transfer (SCNT) embryos

where the nucleus of a somatic cell was transferred into an enucleated oocyte to produce an embryo that is genetically identical to the donor cell, Dux is capable of rectifying aberrant H3K9ac acetylation and inducing local chromatin relaxation in SCNT embryos, thereby ensuring appropriate two-cell genome activation. Generally, abnormal H3K9ac patterns lead to diminished activation of the genome at the two-cell stage during the development of SCNT embryos. Nonetheless, overexpression of Dux corrected the aberrant H3K9ac signal at its target sites, thereby facilitating proper activation of the two-cell stage genome and improving SCNT efficiency. In addition, Yang et al. confirmed that the developmental advancements facilitated by TSA are notably hindered in the absence of Dux during the development of SCNT embryos, which also indicates that the rescue efficiency by TSA largely relies on Dux activation [28].

Investigations into the functional domains of Dux exhibit its structural complexity and reveal the critical domains responsible for its distinct regulatory actions, further substantiating the versatile roles of Dux in embryonic transcriptional regulation during ZGA. The mouse Dux protein, akin to the human DUX4 protein, comprises three functional domains. In detail, two adjacent homeodomains are located in the N-terminal (amino acids 29–72 and 113–166) and a singular domain is located in the C-terminal (amino acids 626–674), leaving amino acids 167–625 of Dux in the intermediate region. Huang et al. confirmed that Dux and DUX4 contain conserved N-terminal double homeodomains that are indispensable for its nuclear localization and any single homeodomain is also sufficient to show the nuclear location with the entire C-terminus. In addition, Huang et al. found that the intermediate region (AA167–625) of the Dux protein was responsible for activating the transcription of *Zscan4* and *MERVL* [29]. The terminal 49 amino acids at the C-terminus of the mouse Dux protein exhibit remarkable similarity to the corresponding C-terminal domain of the human DUX4 protein. This structural conservation underscores a potential evolutionary parallel in their functional roles, indicating that these C-terminal sequences may harbor critical transcriptional regulation capabilities [23]. Recently, Christina et al. identified 5 approximately 100 amino acid repeats followed by a single 14 amino acid highly-acidic tail in the C-terminus of Dux. Mechanistically, there exists cooperativity between active repeats and the acidic tail, which may facilitate cofactor recruitment, Dux-mediated opening of targets and transcription [30]. Concerning the role of DUX4 in facilitating chromatin accessibility for transcription, DUX4 also orchestrates chromatin dynamics through its C-terminal domain, thereby influencing gene expression. In more detail, DUX4 recruits the histone acetyltransferases p300/CBP to its target genes. This recruitment leads to increased H3K27 acetylation at these sites, as indicated by chromatin immunoprecipitation experiments. Specifically, it was observed that DUX4 binding resulted in significant acetylation changes at the *ZSCAN4* locus, with a marked increase in H3K27ac flanking the DUX4 binding site. This implies a reorganization of the nucleosome positioning around these loci. Furthermore, DUX4 is capable of binding to nucleosome-occupied DNA and facilitating chromatin remodeling. This is evidenced by the displacement of nucleosomes, particularly at previously inaccessible sites, following DUX4 binding. ChIP-seq data also reveals that DUX4 binding correlates with histone H3 depletion and increased H3K27 acetylation at target loci, suggesting a role in nucleosome repositioning [31]. Although DUX4 protein recruits the histone acetyltransferase P300 via its C-terminus and induces local chromatin opening, indicating that DUX4 may act as a pioneer factor, the direct binding of DUX4 to nucleosomes in vitro remains unconfirmed. Despite Dux's paramount position within the transcriptional regulatory hierarchy, we should also bear in mind that ZGA may be redundantly driven by other pioneer transcription factors. In Dux-deficient mouse embryo models, a minority of Dux target genes were affected, and the knockout embryos did not arrest at the two-cell stage but survived to adulthood, albeit with diminished developmental capacities [32–34]. It is well known that genes exhibiting sequence homology may display genetic redundancy [35], and multiple functionally overlapping factors typically aim collectively toward a teleological objective in order to counteract sporadic mutations

or defects [36,37]. Thus, it is reasonable to believe that such redundancy may represent an evolutionary adaptation or a fail-safe mechanism to guarantee the initiation of ZGA.

3. The Role of Obox in ZGA

Given that Dux is a homeobox gene, researchers are diligently investigating additional homeobox genes with the potential to trigger ZGA. Recently, Obox, a member of the homeobox gene family, has also been identified as a pioneer transcription factor that is functionally redundant to DUX during ZGA [20,38]. *Obox* genes, including Obox4, maps to proximal chromosome 7 in mouse [39]. Initially, both transcripts and translation levels of Obox4 are low within oocytes. During ZGA, Obox4 and its array of pseudogenes (Obox4-ps) exhibit notable dynamic in expression: the expression increased dramatically in early-two-cell embryos, then quickly declined in late-two-cell embryos [20]. The Obox4 protein is divided into three distinct regions: the N-terminal region, the central homeodomain (HD) region and the C-terminal region. Using a green fluorescent protein (GFP) fusion expression system, Park et al. revealed that intact HD region of Obox4 which harbors the presumptive DNA-binding motif essential for its transcription factor activity, may be responsible for the nuclear localization of Obox4 [40].

It is commonly acknowledged that ZGA in mice is marked by the pronounced upregulation of genes that are uniquely expressed during the two-cell (2C) stage, a significant number of which have co-opted MERVL long terminal repeats (LTRs) as promoters [41]. Thus, it is natural to raise the question whether Obox4 has the potential to trigger ZGA by activating MERVL, akin to the effects triggered by Dux. Through establishing a 2C::tdTomato reporter mESC line with the transgenic tdTomato red fluorescence protein driven by MERVL 5'-LTR, Guo et al. demonstrated that the ectopic expression of Obox4 led to a significant increase in tdTomato expression, mirroring the effect observed with Dux, thus highlighting the instrumental function Obox4 serves in the orchestration of ZGA [37]. Unexpectedly, when Obox4 genetic knockout mESCs were used as nuclear donor cells for SCNT, 54.4% of the reconstructed embryos still developed to the blastocyst stage. In addition, a single knockdown of Obox4 resulted in moderate developmental retardation in zygotes, with almost 50% of embryos advancing to the blastocyst stage. Nonetheless, the combined knockdown of Obox4 and Dux significantly hindered the formation of blastocysts, causing over 80% of the embryos to degenerate or stop developing before reaching the four-cell stage, and fewer than 20% displayed a morula-like structure. Presence of Obox4 can overcome the developmental arrest of Obox4/Dux double knockdown embryos. While RNA sequencing of two-cell stage embryos showed more severe dysregulation of two-cell specific genes and transposable elements due to the double knockdown of Obox4 and Dux compared to a single knockdown, reintroducing Obox4 can rectify the dysregulated transcriptome caused by the double knockdown. It should be noted that the knockout of the entire Obox cluster (including Obox1, Obox2, Obox3/4, Obox5, and Obox7) resulted in a 2–4 cell arrest in embryos. This phenomenon indicates that although Obox4 plays a redundant role in the absence of Dux to a certain extent, the functions of the Obox proteins cannot be fully compensated by Dux, as the complete knockout of the Obox cluster caused severe developmental arrest despite Dux's presence [20,38].

In addition to Obox4, Obox3 also plays a critical role in activating ZGA. Obox3 interacts with specific repetitive elements such as MERVL long terminal repeats. Using ChIP-seq to detect Obox3 binding sites across the genome, Sakamoto et al. discovered that Obox3 primarily binds to promoters of ZGA genes, facilitating RNA Polymerase II pre-configuration at these sites. In addition, Obox3 significantly enhances the acquisition of totipotency in somatic cell nuclear transfer (SCNT) embryos. Sakamoto et al. demonstrated that injection of Obox3 mRNA in SCNT embryos significantly increases the success rate of SCNT embryos reaching the blastocyst stage. This process is accompanied by the restorative activation of ZGA genes and specific repetitive elements such as MERVL, indicating that Obox3 plays a crucial role in correcting abnormal transcription programs in SCNT embryos [42].

From a mechanistic perspective, Obox guides the timely pre-configuration of Pol II and chromatin accessibility at regulatory elements, which initiates the activation of Pol II targets and triggers ZGA. During mouse ZGA, Pol II undergoes “loading, pre-configuration, and synthesis” phases [43], and Pol II binding correlates with CG density. In more detail, although CG-rich promoters were inherently accessible for Pol II loading [44], the Pol II peaks specific to late-two-cell embryos, characterized by their low CG content, require the involvement of additional transcription factors [20]. In late-two-cell-specific Pol II targets, the Obox motif is the most enriched motif, and Obox protein exhibits the highest expression among the inferred transcription factors; furthermore, Ji et al., through performing Stacc-seq, confirmed that Obox proteins are responsible for recruiting RNA polymerase II to CG-poor promoters of ZGA genes in two-cell stage embryos [20]. Considering that the failure to open promoters and enhancers correlates with the downregulation of nearby ZGA genes, this phenomenon suggests that Obox proteins facilitate chromatin opening at regulatory elements and guide the pre-configuration of Pol II at these sites. Consequently, the absence of Obox leads to defective ZGA [20]. Intriguingly, based on ATAC-seq results, it was shown that active enhancers, defined by distal H3K27ac, exhibited substantial decreases in chromatin accessibility in *Obox* maternal-zygotic KO late-two-cell embryos, indicating that Obox depletion decreased chromatin accessibility at the late-two-cell-specific Pol II binding sites [20]. Although Obox may be involved in chromatin opening, whether Obox proteins possess all the characteristics of pioneer factors remains an open question. Collectively, Obox proteins are instrumental in enhancing chromatin accessibility at CG-poor promoters and enhancers and orchestrating the temporal pre-configuration of RNA polymerase II, thereby facilitating the initiation of transcriptional activation of RNA polymerase II-dependent genes, which pave the way for onset of ZGA.

4. The Role of Nr5a2 in ZGA

Nuclear receptor subfamily 5 group A member 2 (Nr5a2), also known as liver receptor homolog-1 (LRH-1), is a zinc-finger transcription factor that belongs to the orphan nuclear receptor superfamily. Mouse Nr5a2, the ortholog of NR5A2 in human, maps to chromosome 7 in mouse. Previous research has shown that Nr5a2 is expressed in liver, exocrine pancreas, intestinal crypts and ovaries in adult mammals [45–48]. Recently, experimental evidence derived from immunofluorescence has shown that although Nr5a2 exhibits very low expression in GV (germinal vesicle) oocytes, it is significantly upregulated in the two-cell stage embryos [21]. Thus, although it is known that Nr5a2 regulates cholesterol and bile acid homeostasis, as well as cholesterol delivery and steroid production [49–52], we raise the question of whether it also acts as a candidate ZGA regulator.

To ascertain the potential role of Nr5a2 in early embryonic development, particularly in the context of ZGA, the developmental potential of fertilized embryos was monitored. SR1848, a chemical inhibitor, causes Nr5a2 to move from the nucleus to the cytoplasm, thereby preventing it from activating the transcription of its target genes. To specifically explore the function of Nr5a2, SR1848 was used [53,54]. When embryos were exposed to SR1848 commencing at 6 h post fertilization (hpf), these embryos failed to develop into blastocysts and exhibited extensive fragmentation. These embryos starkly contrast to control embryos, corresponding to Nr5a2’s role in sustaining naïve pluripotency in mouse ES cells [55]. Intriguingly, transient inhibition of Nr5a2 from 6 to 36 hpf led to an arrest at the two-cell stage, while embryos treated with SR1848 beginning at 36 hpf (two-cell stage) manifested a less severe phenotype than those treated from 6 hpf. These phenomena indicated that Nr5a2’s activity is crucial between fertilization and the two-cell stage, coinciding with the occurrence of ZGA [21]. Subsequently, to explore whether Nr5a2 initiates the onset of ZGA, Gassler et al. utilized single-molecule FISH analysis algorithms to quantify the de novo expression of ZGA genes. They treated zygotes with SR1848, then observed a dose-dependent reduction in nascent ZGA transcripts [21]. Furthermore, using the CUT&Tag technique optimized for ultra-low input samples, Gassler et al. also assessed whether Nr5a2 actively binds in the vicinity of ZGA gene regions in two-cell

embryos. In detail, using specific antibodies against Nr5a2, CUT&Tag results indicated a significant enrichment of Nr5a2 within these regions; and the motifs emerging from CUT&Tag peaks were characterized by an (A/G) (A/G) T sequence positioned upstream of the consensus sequences, which was also detected in a motif embedded in the *SINE B1/Alu* 5YR upstream of ZGA genes. This suggests that SINE B1 retrotransposons may serve as critical regulatory targets for Nr5a2 during the onset of ZGA [21]. Given the published ATAC-seq and ChIP-seq data (including histone modification H3K4me3 and H3K27ac) [56,57], it was also discovered that Nr5a2 binds to open chromatin, preferentially targeting enhancers in two-cell embryos [21].

Mechanistically, it is logical to pose the question of whether Nr5a2 performs its biological functions by promoting chromatin accessibility. Based on ChARM (Chromatin Accessibility Revealed by Microscopy) technology, both knockdown studies and chemical inhibition experiments confirmed that Nr5a2 is positively correlated with chromatin accessibility in early embryonic development; notably, treatment with SR1848 led to a substantial reduction in chromatin accessibility across a majority of Nr5a2 motifs. At the same time, Nr5a2 exhibits motif-dependent binding to nucleosomal DNA, preferentially at entry–exit sites, similar to known pioneer factors. Nr5a2’s interactions with nucleosomal DNA are both specific and competitive, which also align with the characteristics of pioneer transcription factors. Collectively, the combination of genome-wide binding profiles, chromatin accessibility assays and in vitro nucleosome binding has proven that Nr5a2 exhibits properties consistent with pioneer transcription factors [21]. Recently, Kobayashi et al. further demonstrated the mechanism underlying how Nr5a2 engages with nucleosomes to exert its pioneer activity [58]. The crystallographic analysis of the human NR5A2-DNA complex revealed that the α -helix of the zinc finger domain binds to the major groove, while the CTE (carboxy-terminal extension) binds to the minor groove [59]. Nr5a2, acting as a monomer, specifically binds to the DNA motif on both naked DNA and nucleosomes [21,60]. Through presenting a cryo-electron microscopy structure of human NR5A2 bound to a nucleosome, Kobayashi et al. observed the CTE loop of the NR5A2 DNA-binding domain interacts competitively with a DNA minor groove anchor of the nucleosome, facilitating the release of entry–exit site DNA, corresponding to the model in which histone–DNA interactions are compensated with pioneer transcription factor–DNA interactions [58,61]. Additionally, while the D159 residue of the CTE is not essential for DNA binding, it is crucial for maintaining stable nucleosome association and continuous DNA “unwrapping” [61]. Importantly, the RGGR motif within the CTE loop is a common characteristic among members of the NR5A and NR3B orphan nuclear receptor families, indicating that competition in the minor groove mediated by the CTE loop may be a conserved mechanism in these transcription factors [59].

Despite the conclusions from the study by Gassler et al. suggesting that Nr5a2 plays a certain role in the ZGA process, we shall also bear in mind that the role of Nr5a2 in ZGA remains a topic of ongoing debate. Recent studies utilizing more stringent genetic approaches have challenged these findings. For instance, Lai et al. demonstrated that Nr5a2 is largely dispensable for global ZGA but is essential for the 4–8 cell stage gene program.

Surprisingly, utilizing comprehensive ATAC-seq and CUT&RUN techniques to profile Nr5a2 binding, Lai et al. obtained high-quality datasets which highlighted Nr5a2’s binding to cis-regulatory elements, particularly in 2-cell and 8-cell embryos, and exhibited Nr5a2’s essential role in activating genes during the 4–8 cell stage transition. Their data showed that while Nr5a2 knockdown or knockout did not significantly impact ZGA, it led to impaired activation of 4–8 cell stage genes and resulted in embryonic arrest at the morula stage. Similarly, Festuccia et al. reported that Nr5a2 knockout embryos progressed beyond the two-cell stage but exhibited significant developmental delays and morphological abnormalities at the morula stage, highlighting its crucial role in morula development and first lineage specification [62,63]. We hypothesize that further studies using more specific inhibitors targeting Nr5a2 or the same genetic background mice may help resolve the debates about

the role of Nr5a2.5. Pioneer Transcription Factors Utilize Retrotransposons Elements as Hubs for ZGA Regulation

Retrotransposons, which originated from ancient retroviruses, have colonized genomes and now make up nearly 40% of mammalian genomes [64]. Retrotransposons are categorized into LTR (long terminal repeat) retrotransposons, also referred to as ERVs (endogenous retroviruses), and non-LTR retrotransposons; non-LTR retrotransposons are also further divided into LINEs (long interspersed elements) and SINEs (short interspersed elements) [65–68]. Given that the transposition activities of retrotransposons may compromise genomic integrity and potentially lead to cancer and autoimmune diseases [68,69], retrotransposons are generally silenced in most somatic cells through various epigenetic modifications, including DNA methylation and repressive histone modifications, to reduce the risk of retrotransposition [70,71]. Counterintuitively, bursts of retrotransposon transcription usually occur during ZGA in preimplantation embryos [72,73]. From an evolutionary perspective, we believe that there is a delicate balance between the activation of retrotransposon transcription and genome integrity in the developing mammalian embryo. Facing the pressures of natural selection, retroviruses and hosts compromise with each other. Taking endogenous retroviruses as an example, on one hand, throughout evolutionary history, recombination between LTRs has led to the loss of internal ERV genes, resulting in solo LTRs and preventing the harmful expression of viral proteins within the host genome [74]; on the other hand, during the reprogramming window, early embryos tolerate a massive onslaught of ERV expression, using cis-regulatory elements located in LTRs to rewire the gene regulatory network [75]. It is well recognized that despite the colonization by selfish retroviruses, host genomes have successfully domesticated retrotransposon sequences for the benefit of the host species. This process is also known as co-option [76,77]. In other words, in the context of host early development, retrotransposons have evolved cis-regulatory sequences that recruit host-encoded factors, such as transcription factors, to ensure the expression of essential developmental genes during ZGA, ultimately achieving host co-option.

During ZGA, pioneer transcription factors, which reside at the top of the transcriptional hierarchy, initiate cascading effects on gene expression using retrotransposon elements as hubs. These three pioneer transcription factors (such as Dux, Obox and Nr5a2) discussed in this review all achieve their functions by targeting motifs embedded within retrotransposons, as shown in Figure 1. The proliferation of retrotransposon and its associated cis-regulatory elements facilitates the expansion of transcription factor binding sites (TFBSs) throughout the genome, thereby driving evolutionary innovations in regulatory mechanisms.

Dux has been pinpointed as a central inducer of ZGA, largely through binding to the MERVL LTR [17–19]. A significant portion of the transcriptome at the two-cell stage initiates transcription within the MERVL LTR. The MERVL LTR seems to have been co-opted by the host genome as an alternative promoter, driving the expression of genes related to the two-cell stage during ZGA [73]. Using CRISPRi (CRISPR interference), Yang et al. perturbed 2,485 (93%) MERVL-LTRs (MT2-Mm) insertions in two-cell embryos, and then observed downregulation of hundreds of ZGA genes [78]. Importantly, analysis with a conservative approach of unique reads indicated that Dux directly interacted with at least 53% of all MERVL-LTRs (MT2-Mm) and at least 37% of regions that showed increased ATAC sensitivity in 2C-like cells in mESCs. Furthermore, a consensus DUX-binding motif (WGATTYAATCW) was identified, and this motif showed significant enrichment in regions exhibiting increased ATAC sensitivity following Dux overexpression. Intriguingly, MERVL-LTRs (MT2_Mm) exhibited no significant DUX4-motif enrichment, and HERVL-LTRs (MLT2A1/2) displayed only minimal DUX-motif enrichment [18]. This phenomenon suggests that although both Dux in mice and DUX4 in humans are intronless and located in tandem arrays across multiple chromosomes, indicating functional conservation [23], they have evolved to be species-specific, possibly in response to ERVs.

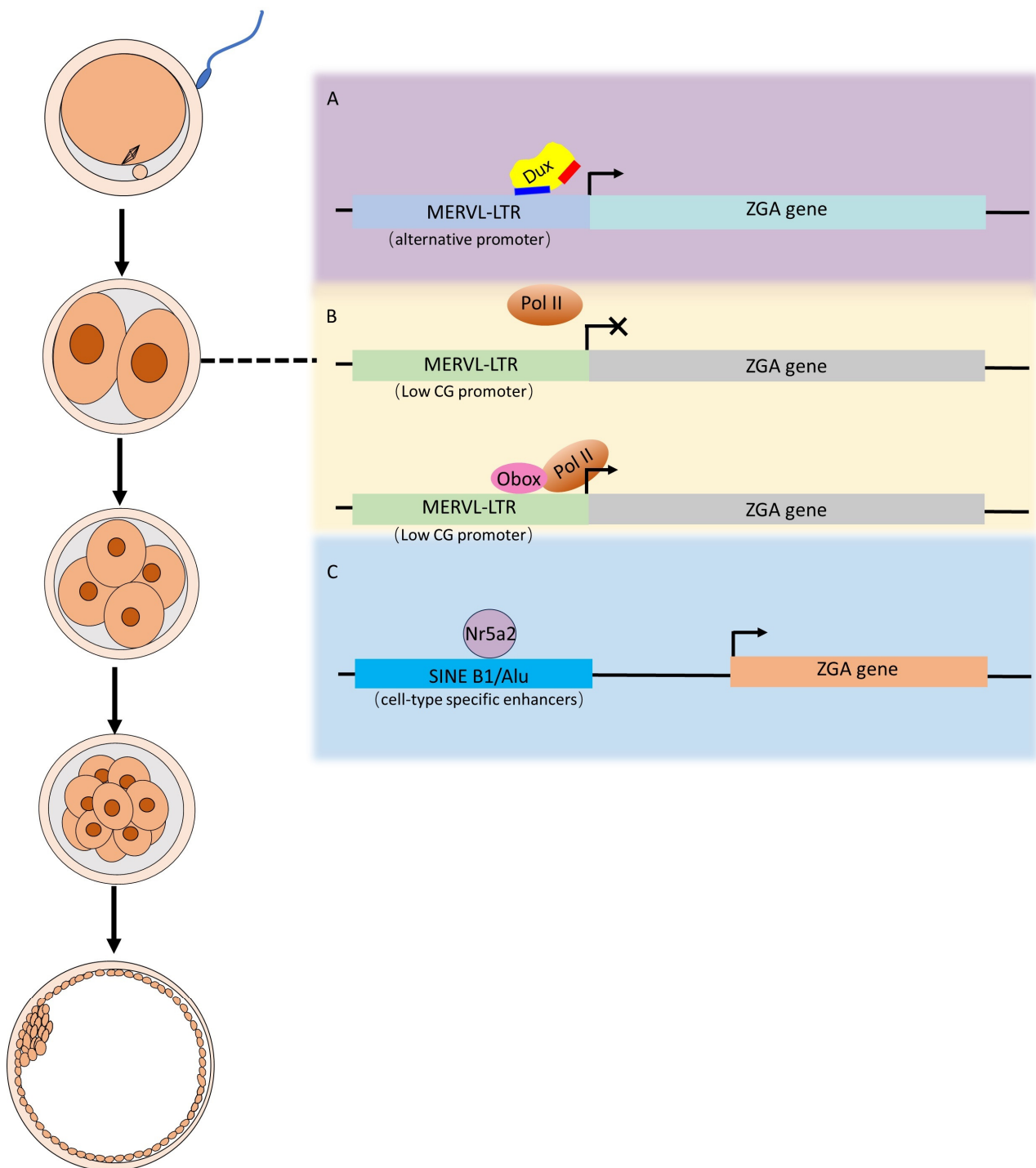


Figure 1. Scheme of pioneer transcription factors activating ZGA genes via retrotransposon elements. (A) ZGA gene activation by Dux. During ZGA, endogenous Dux is transiently expressed. Through the more divergent homeodomain (the first homeodomain) in DUX-C family proteins, Dux binds to the MERVL-LTR, which is co-opted by the host as an alternative promoter, thereby activating downstream ZGA genes. (B) ZGA gene activation by Obox. During ZGA, RNA pol II alone cannot locate CG-poor promoters of ZGA genes; therefore, Obox proteins are essential for recruiting RNA pol II to these promoters, which in turn activate downstream ZGA genes. (C) ZGA gene activation by Nr5a2. Nr5a2 specifically binds to its motif within a subtype of SINE B1/Alu transposable elements, which are prevalent in the enhancers of ZGA genes, ultimately activating downstream ZGA genes.

Obox preferentially binds to and regulates ZGA genes via the Obox motif in mouse embryos. Ji et al. investigated the DNA binding sites of Obox proteins in early mouse embryos by overexpressing Obox genes (Obox1, Obox5, and Obox3) tagged with a Flag label. Using next-generation sequencing and bioinformatics tools, they confirmed that Obox was enriched at MERVL and B1/B2/B4 repetitive elements. In more detail, they identified two main de novo motifs that closely matched the known Obox binding motif (TAATCCC) [20,79]. In addition, these de novo motifs are often closely spaced at the binding sites, forming an extended binding motif (ACNCCTTTAATCCCAG), with Obox1 displaying the longest consecutive version of this motif (CCTTTAATCCCAG). Notably, approximately 95.9% of the extended Obox motif is located in the B1 element, resulting in stronger gene activation than the reported 7-bp motif in reporter assays in embryos. Additionally, Guo et al. discovered that Obox4 directly binds to MERVL and MERVK loci and activates specific MERVL and MERVK elements in a Dux-independent manner, positioning *Obox4* as a redundant gene relative to *Dux* in the context of ZGA [38]. It is noteworthy that in the context of ZGA, genes specific to the two-cell stage exhibit a pronounced enrichment for Obox transcription factor binding motifs at promoter regions, with an observed increase in motifs relative to genes activated at subsequent developmental stages, which indicates a distinct regulatory landscape compared to genes activated at subsequent developmental stages [20].

Nr5a2 has been identified as a crucial activator of ZGA in mouse embryos, primarily due to its binding to specific motifs within a subtype of SINE B1/Alu transposable elements located in the cis-regulatory regions of ZGA genes. Through exploring the enrichment of transcription factor motifs in the cis-regulatory regions of ZGA genes, a specific motif sequence was found to be prevalent in a significant majority of ZGA genes, containing variable pyrimidine/purine (YR) stretches. Particularly, the CA variant of the fifth YR stretch, YR5, is associated with enhanced chromatin accessibility and increased histone acetylation, suggesting its significant function during ZGA. In detail, the CA variant of the motif includes the consensus sequence TCAAGGCCA, which was identified as the Nr5a2 motif. By comparing the binding sites of Nr5a2 to the overlap with SINE B1 transposable elements, it was found that 70% of Nr5a2 peaks overlapped with SINE B1, suggesting that SINE B1 transposons are major targets for Nr5a2. The Nr5a2 motif identified in the peaks includes the sequence (A/G)(A/G)T upstream of consensus sequences, a pattern similarly observed in motif 1 within the SINE B1/Alu 5YR upstream of ZGA genes [21].

5. Conclusions

Pioneer transcription factors, such as Dux, Obox and Nr5a2, orchestrate a complex regulatory landscape during ZGA, which highlights their central role in developmental biology and their potential utility in improving reprogramming efficiency. Host genomes have successfully domesticated retrotransposon elements that served both as challenges and aids by evolving cis-regulatory sequences, which ensures the expression of crucial developmental genes during ZGA and achieves a symbiotic use of these retrotransposon elements. Specifically, pioneer transcription factors target these retrotransposon elements and employ these elements as regulatory hubs for ZGA, then these pioneer transcription factors reside at the top of a transcriptional hierarchy. Dissecting pioneer transcription factor functions not only provides insights into the fundamental processes of life but also opens avenues for the application of this knowledge in reproductive medicine, genetic research, animal breeding and regenerative medicine domains. In the future, the combination of Hi-C with super-resolution imaging techniques, which can provide insights into the dynamic organization of chromatin in the nucleus, may reveal further implications for the function of pioneer transcription factors in gene regulation [80].

Author Contributions: Conceptualization, B.F. and H.M.; drafting the text, B.F. and H.M.; manuscript review and editing, B.F., H.M. and D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Heilongjiang Provincial Research Institutes Research Business Fund Project (CZKYF2024-1-B006), National Natural Science Foundation of China (U20A2052), National Center of Technology Innovation for Pigs (NCTIP-XD1C16), China Agricultural Research System (CARS-35).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jukam, D.; Shariati, S.A.M.; Skotheim, J.M. Zygotic Genome Activation in Vertebrates. *Dev. Cell* **2017**, *42*, 316–332. [\[CrossRef\]](#)
2. Schulz, K.N.; Harrison, M.M. Mechanisms regulating zygotic genome activation. *Nat. Rev. Genet.* **2019**, *20*, 221–234. [\[CrossRef\]](#)
3. Østrup, O.; Andersen, I.S.; Collas, P. Chromatin-linked determinants of zygotic genome activation. *Cell. Mol. Life Sci.* **2013**, *70*, 1425–1437. [\[CrossRef\]](#)
4. Wu, J.; Xu, J.; Liu, B.; Yao, G.; Wang, P.; Lin, Z.; Huang, B.; Wang, X.; Li, T.; Shi, S.; et al. Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **2018**, *557*, 256–260. [\[CrossRef\]](#)
5. Farkas, G.; Leibovitch, B.A.; Elgin, S.C.R. Chromatin organization and transcriptional control of gene expression in *Drosophila*. *Gene* **2000**, *253*, 117–136. [\[CrossRef\]](#)
6. Kobayashi, W.; Tachibana, K. Awakening of the zygotic genome by pioneer transcription factors. *Curr. Opin. Struct. Biol.* **2021**, *71*, 94–100. [\[CrossRef\]](#)
7. Iwafuchi-Doi, M.; Zaret, K.S. Pioneer transcription factors in cell reprogramming. *Genes Dev. J. Devoted Mol. Anal. Gene Expr. Eukaryotes Prokaryotes Viruses* **2014**, *28*, 2679–2692. [\[CrossRef\]](#)
8. Zaret, K.S.; Carroll, J.S. Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev.* **2011**, *25*, 2227–2241. [\[CrossRef\]](#)
9. Zaret, K.S. Pioneer transcription factors initiating gene network changes. *Annu. Rev. Genet.* **2020**, *54*, 367–385. [\[CrossRef\]](#)
10. Swinstead, E.E.; Paakinaho, V.; Presman, D.M.; Hager, G.L. Pioneer factors and ATP-dependent chromatin remodeling factors interact dynamically: A new perspective: Multiple transcription factors can effect chromatin pioneer functions through dynamic interactions with ATP-dependent chromatin remodeling factors. *Bioessays* **2016**, *38*, 1150–1157. [\[CrossRef\]](#)
11. Larson, E.D.; Marsh, A.J.; Harrison, M.M. Pioneering the developmental frontier. *Mol. Cell* **2021**, *81*, 1640–1650. [\[CrossRef\]](#)
12. Duan, J.; Rieder, L.; Colonna, M.M.; Huang, A.; Mckennedy, M.; Watters, S.; Deshpande, G.; Jordan, W.; Fawzi, N.; Larschan, E. CLAMP and Zelda function together to promote *Drosophila* zygotic genome activation. *Elife* **2021**, *10*, e69937. [\[CrossRef\]](#)
13. Liang, H.-L.; Nien, C.-Y.; Liu, H.-Y.; Metzstein, M.M.; Kirov, N.; Rushlow, C. The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* **2008**, *456*, 400–403. [\[CrossRef\]](#)
14. 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 *Drosophila* embryo. *Elife* **2021**, *10*, e66668. [\[CrossRef\]](#)
15. Lee, M.T.; Bonneau, A.R.; Takacs, C.M.; Bazzini, A.A.; DiVito, K.R.; Fleming, E.S.; Giraldez, A.J. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **2013**, *503*, 360–364. [\[CrossRef\]](#)
16. Leichenring, M.; Maes, J.; Mössner, R.; Driever, W.; Onichtchouk, D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **2013**, *341*, 1005–1009. [\[CrossRef\]](#)
17. Whiddon, J.L.; Langford, A.T.; Wong, C.-J.; Zhong, J.W.; Tapscott, S.J. Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* **2017**, *49*, 935–940. [\[CrossRef\]](#)
18. Hendrickson, P.G.; Dorais, J.A.; Grow, E.J.; Whiddon, J.L.; Lim, J.-W.; Wike, C.L.; Weaver, B.D.; Pflueger, C.; Emery, B.R.; Wilcox, A.L. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **2017**, *49*, 925–934. [\[CrossRef\]](#)
19. De Iaco, A.; Planet, E.; Coluccio, A.; Verp, S.; Duc, J.; Trono, D. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **2017**, *49*, 941–945. [\[CrossRef\]](#)
20. Ji, S.; Chen, F.; Stein, P.; Wang, J.; Zhou, Z.; Wang, L.; Zhao, Q.; Lin, Z.; Liu, B.; Xu, K. OBOX regulates mouse zygotic genome activation and early development. *Nature* **2023**, *620*, 1047–1053. [\[CrossRef\]](#)
21. Gassler, J.; Kobayashi, W.; Gáspár, I.; Ruangroengkulrith, S.; Mohanan, A.; Gómez Hernández, L.; Kravchenko, P.; Kümmecke, M.; Lalic, A.; Rifel, N. Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science* **2022**, *378*, 1305–1315. [\[CrossRef\]](#)
22. Mitsuhashi, S.; Nakagawa, S.; Takahashi Ueda, M.; Imanishi, T.; Frith, M.C.; Mitsuhashi, H. Nanopore-based single molecule sequencing of the D4Z4 array responsible for facioscapulohumeral muscular dystrophy. *Sci. Rep.* **2017**, *7*, 14789. [\[CrossRef\]](#)
23. Clapp, J.; Mitchell, L.M.; Bolland, D.J.; Fantes, J.; Corcoran, A.E.; Scotting, P.J.; Armour, J.A.; Hewitt, J.E. Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **2007**, *81*, 264–279. [\[CrossRef\]](#)
24. Geng, L.N.; Yao, Z.; Snider, L.; Fong, A.P.; Cech, J.N.; Young, J.M.; van der Maarel, S.M.; Ruzzo, W.L.; Gentleman, R.C.; Tawil, R. DUX4 activates germline genes, retroelements, and immune mediators: Implications for facioscapulohumeral dystrophy. *Dev. Cell* **2012**, *22*, 38–51. [\[CrossRef\]](#)
25. Hewitt, J.E. Loss of epigenetic silencing of the DUX4 transcription factor gene in facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* **2015**, *24*, 17–23. [\[CrossRef\]](#)

26. Nip, Y.; Bennett, S.R.; Smith, A.A.; Jones, T.I.; Jones, P.L.; Tapscott, S.J. Human DUX4 and porcine DUXC activate similar early embryonic programs in pig muscle cells: Implications for preclinical models of FSHD. *Hum. Mol. Genet.* **2023**, *32*, 1864–1874. [\[CrossRef\]](#)
27. Sugie, K.; Funaya, S.; Kawamura, M.; Nakamura, T.; Suzuki, M.G.; Aoki, F. Expression of Dux family genes in early preimplantation embryos. *Sci. Rep.* **2020**, *10*, 19396. [\[CrossRef\]](#)
28. Yang, G.; Zhang, L.; Liu, W.; Qiao, Z.; Shen, S.; Zhu, Q.; Gao, R.; Wang, M.; Wang, M.; Li, C. Dux-mediated corrections of aberrant H3K9ac during 2-cell genome activation optimize efficiency of somatic cell nuclear transfer. *Cell Stem Cell* **2021**, *28*, 150–163. [\[CrossRef\]](#)
29. Huang, X.; Hu, X.; Jiang, Q.; Cao, Q.; Wu, Y.; Lei, L. Functional study of distinct domains of Dux in improving mouse SCNT embryonic development. *Biol. Reprod.* **2021**, *105*, 1089–1103. [\[CrossRef\]](#)
30. Smith, C.M.; Grow, E.J.; Shadle, S.C.; Cairns, B.R. Multiple repeat regions within mouse DUX recruit chromatin regulators to facilitate an embryonic gene expression program. *bioRxiv* **2023**. bioRxiv:03.29.534786.
31. Choi, S.H.; Gearhart, M.D.; Cui, Z.; Bosnakovski, D.; Kim, M.; Schennum, N.; Kyba, M. DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.* **2016**, *44*, 5161–5173. [\[CrossRef\]](#)
32. De Iaco, A.; Verp, S.; Offner, S.; Grun, D.; Trono, D. DUX is a non-essential synchronizer of zygotic genome activation. *Development* **2020**, *147*, dev177725. [\[CrossRef\]](#)
33. Chen, Z.; Zhang, Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat. Genet.* **2019**, *51*, 947–951. [\[CrossRef\]](#)
34. Yang, L.; Liu, X.; Song, L.; Di, A.; Su, G.; Bai, C.; Wei, Z.; Li, G. Transient Dux expression facilitates nuclear transfer and induced pluripotent stem cell reprogramming. *EMBO Rep.* **2020**, *21*, e50054. [\[CrossRef\]](#)
35. Wagner, A. Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. *Biol. Cybern.* **1996**, *74*, 557–567. [\[CrossRef\]](#)
36. Nowak, M.A.; Boerlijst, M.C.; Cooke, J.; Smith, J.M. Evolution of genetic redundancy. *Nature* **1997**, *388*, 167–171. [\[CrossRef\]](#)
37. Vavouri, T.; Sempke, J.I.; Lehner, B. Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. *Trends Genet.* **2008**, *24*, 485–488. [\[CrossRef\]](#)
38. Guo, Y.; Kitano, T.; Inoue, K.; Murano, K.; Hirose, M.; Li, T.D.; Sakashita, A.; Ishizu, H.; Ogonuki, N.; Matoba, S. Obox4 secures zygotic genome activation upon loss of Dux. *bioRxiv* **2022**. bioRxiv:2022.07.04.498763. [\[CrossRef\]](#)
39. Rajkovic, A.; Yan, C.; Yan, W.; Klysik, M.; Matzuk, M.M. Obox, a family of homeobox genes preferentially expressed in germ cells. *Genomics* **2002**, *79*, 711–717. [\[CrossRef\]](#)
40. Park, G.T.; Lee, K.-A. Nuclear localization of Obox4 is dependent on its homeobox domain. *Clin. Exp. Reprod. Med.* **2013**, *40*, 1–6. [\[CrossRef\]](#)
41. Macfarlan, T.S.; Gifford, W.D.; Agarwal, S.; Driscoll, S.; Lettieri, K.; Wang, J.; Andrews, S.E.; Franco, L.; Rosenfeld, M.G.; Ren, B. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* **2011**, *25*, 594–607. [\[CrossRef\]](#)
42. Sakamoto, M.; Ito, A.; Wakayama, S.; Sasaki, H.; Wakayama, T.; Ishiuchi, T. Detection of newly synthesized RNA reveals transcriptional reprogramming during ZGA and a role of Obox3 in totipotency acquisition. *Cell Rep.* **2024**, *43*, 114118. [\[CrossRef\]](#)
43. Liu, B.; Xu, Q.; Wang, Q.; Feng, S.; Lai, F.; Wang, P.; Zheng, F.; Xiang, Y.; Wu, J.; Nie, J. The landscape of RNA Pol II binding reveals a stepwise transition during ZGA. *Nature* **2020**, *587*, 139–144. [\[CrossRef\]](#)
44. Fenouil, R.; Cauchy, P.; Koch, F.; Descostes, N.; Cabeza, J.Z.; Innocenti, C.; Ferrier, P.; Spicuglia, S.; Gut, M.; Gut, I. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.* **2012**, *22*, 2399–2408. [\[CrossRef\]](#)
45. Botrugno, O.A.; Fayard, E.; Annicotte, J.-S.; Haby, C.; Brennan, T.; Wendling, O.; Tanaka, T.; Kodama, T.; Thomas, W.; Auwerx, J. Synergy between LRH-1 and β -catenin induces G1 cyclin-mediated cell proliferation. *Mol. Cell* **2004**, *15*, 499–509. [\[CrossRef\]](#)
46. Galarneau, L.; Paré, J.-F.; Allard, D.; Hamel, D.; Lévesque, L.; Tugwood, J.D.; Green, S.; Bélanger, L. The α 1-fetoprotein locus is activated by a nuclear receptor of the Drosophila FTZ-F1 family. *Mol. Cell. Biol.* **1996**, *16*, 3853–3865. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Fayard, E.; Schoonjans, K.; Annicotte, J.-S.; Auwerx, J. Liver receptor homolog 1 controls the expression of carboxyl ester lipase. *J. Biol. Chem.* **2003**, *278*, 35725–35731. [\[CrossRef\]](#)
48. Boerboom, D.; Pilon, N.; Behdjani, R.; Silversides, D.W.; Sirois, J. Expression and regulation of transcripts encoding two members of the NR5A nuclear receptor subfamily of orphan nuclear receptors, steroidogenic factor-1 and NR5A2, in equine ovarian cells during the ovulatory process. *Endocrinology* **2000**, *141*, 4647–4656. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Fayard, E.; Auwerx, J.; Schoonjans, K. LRH-1: An orphan nuclear receptor involved in development, metabolism and steroidogenesis. *Trends Cell Biol.* **2004**, *14*, 250–260. [\[CrossRef\]](#)
50. Schoonjans, K.; Annicotte, J.-S.; Huby, T.; Botrugno, O.A.; Fayard, E.; Ueda, Y.; Chapman, J.; Auwerx, J. Liver receptor homolog 1 controls the expression of the scavenger receptor class B type I. *EMBO Rep.* **2002**, *3*, 1181–1187. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Hinshelwood, M.M.; Repa, J.J.; Shelton, J.M.; Richardson, J.A.; Mangelsdorf, D.J.; Mendelson, C.R. Expression of LRH-1 and SF-1 in the mouse ovary: Localization in different cell types correlates with differing function. *Mol. Cell. Endocrinol.* **2003**, *207*, 39–45. [\[CrossRef\]](#) [\[PubMed\]](#)

52. Saxena, D.; Safi, R.; Little-Ihrig, L.; Zeleznik, A.J. Liver receptor homolog-1 stimulates the progesterone biosynthetic pathway during follicle-stimulating hormone-induced granulosa cell differentiation. *Endocrinology* **2004**, *145*, 3821–3829. [[CrossRef](#)] [[PubMed](#)]
53. Busby, S.; Nuhant, P.; Cameron, M.; Mercer, B.A.; Hodder, P.; Roush, W.R.; Griffin, P.R. *Probe Reports from the NIH Molecular Libraries Program*; National Center for Biotechnology Information: Bethesda, MD, USA, 2010. [[PubMed](#)]
54. Corzo, C.A.; Mari, Y.; Chang, M.R.; Khan, T.; Kuruvilla, D.; Nuhant, P.; Kumar, N.; West, G.M.; Duckett, D.R.; Roush, W.R. Antiproliferation activity of a small molecule repressor of liver receptor homolog 1. *Mol. Pharmacol.* **2015**, *87*, 296–304. [[CrossRef](#)]
55. Festuccia, N.; Owens, N.; Chervova, A.; Dubois, A.; Navarro, P. The combined action of Esrrb and Nr5a2 is essential for murine naive pluripotency. *Development* **2021**, *148*, dev199604. [[CrossRef](#)] [[PubMed](#)]
56. Dahl, J.A.; Jung, I.; Aanes, H.; Greggains, G.D.; Manaf, A.; Lerdrup, M.; Li, G.; Kuan, S.; Li, B.; Lee, A.Y. Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* **2016**, *537*, 548–552. [[CrossRef](#)] [[PubMed](#)]
57. Wu, J.; Huang, B.; Chen, H.; Yin, Q.; Liu, Y.; Xiang, Y.; Zhang, B.; Liu, B.; Wang, Q.; Xia, W. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **2016**, *534*, 652–657. [[CrossRef](#)] [[PubMed](#)]
58. Kobayashi, W.; Sappier, A.H.; Bollschweiler, D.; Kümmecke, M.; Basquin, J.; Arslantas, E.N.; Ruangroengkulrith, S.; Hornberger, R.; Duderstadt, K.; Tachibana, K. Nucleosome-bound NR5A2 structure reveals pioneer factor mechanism by DNA minor groove anchor competition. *Nat. Struct. Mol. Biol.* **2024**, *31*, 757–766. [[CrossRef](#)] [[PubMed](#)]
59. Solomon, I.H.; Hager, J.M.; Safi, R.; McDonnell, D.P.; Redinbo, M.R.; Ortlund, E.A. Crystal structure of the human LRH-1 DBD–DNA complex reveals Ftz-F1 domain positioning is required for receptor activity. *J. Mol. Biol.* **2005**, *354*, 1091–1102. [[CrossRef](#)]
60. Seacrist, C.D.; Kuenze, G.; Hoffmann, R.M.; Moeller, B.E.; Burke, J.E.; Meiler, J.; Blind, R.D. Integrated structural modeling of full-length LRH-1 reveals inter-domain interactions contribute to receptor structure and function. *Structure* **2020**, *28*, 830–846. [[CrossRef](#)]
61. Felipe, C.; Shin, J.; Kolomeisky, A.B. How pioneer transcription factors search for target sites on nucleosomal DNA. *J. Phys. Chem. B* **2022**, *126*, 4061–4068. [[CrossRef](#)]
62. Lai, F.; Li, L.; Hu, X.; Liu, B.; Zhu, Z.; Liu, L.; Fan, Q.; Tian, H.; Xu, K.; Lu, X. NR5A2 connects zygotic genome activation to the first lineage segregation in totipotent embryos. *Cell Res.* **2023**, *33*, 952–966. [[CrossRef](#)]
63. Festuccia, N.; Vandormael-Pournin, S.; Chervova, A.; Geiselman, A.; Langa-Vives, F.; Coux, R.-X.; Gonzalez, I.; Cohen-Tannoudji, M.; Navarro, P. Nr5a2 is essential for morula development. *bioRxiv* **2023**. bioRxiv:2023.01.16.524255.
64. Waterston, R.H.; Toh, K.L.; Birney, E.; Lander, E.S. Initial sequencing and comparative analysis of the mouse genome. *Nature* **2002**, *420*, 520–562. [[PubMed](#)]
65. Ostertag, E.M.; Kazazian, H.H., Jr. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **2001**, *35*, 501–538. [[CrossRef](#)] [[PubMed](#)]
66. Mager, D.L.; Stoye, J.P. Mammalian endogenous retroviruses. In *Mobile DNA III*; ASM Press: Karnataka, India, 2015; pp. 1079–1100.
67. Cordaux, R.; Batzer, M.A. The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **2009**, *10*, 691–703. [[CrossRef](#)] [[PubMed](#)]
68. Lamprecht, B.; Walter, K.; Kreher, S.; Kumar, R.; Hummel, M.; Lenze, D.; Köchert, K.; Bouhrel, M.A.; Richter, J.; Soler, E. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat. Med.* **2010**, *16*, 571–579. [[CrossRef](#)] [[PubMed](#)]
69. Nexø, B.A.; Christensen, T.; Frederiksen, J.; Møller-Larsen, A.; Oturai, A.B.; Villesen, P.; Hansen, B.; Nissen, K.K.; Laska, M.J.; Petersen, T.S. The etiology of multiple sclerosis: Genetic evidence for the involvement of the human endogenous retrovirus HERV-Fc1. *PLoS ONE* **2011**, *6*, e16652. [[CrossRef](#)] [[PubMed](#)]
70. Kao, T.-H.; Liao, H.-F.; Wolf, D.; Tai, K.-Y.; Chuang, C.-Y.; Lee, H.-S.; Kuo, H.-C.; Hata, K.; Zhang, X.; Cheng, X. Ectopic DNMT3L triggers assembly of a repressive complex for retroviral silencing in somatic cells. *J. Virol.* **2014**, *88*, 10680–10695. [[CrossRef](#)] [[PubMed](#)]
71. Reichmann, J.; Crichton, J.H.; Madej, M.J.; Taggart, M.; Gautier, P.; Garcia-Perez, J.L.; Meehan, R.R.; Adams, I.R. Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells. *PLoS Comput. Biol.* **2012**, *8*, e1002486. [[CrossRef](#)]
72. Fadloun, A.; Le Gras, S.; Jost, B.; Ziegler-Birling, C.; Takahashi, H.; Gorab, E.; Carninci, P.; Torres-Padilla, M.-E. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* **2013**, *20*, 332–338. [[CrossRef](#)]
73. Peaston, A.E.; Esvikov, A.V.; Graber, J.H.; De Vries, W.N.; Holbrook, A.E.; Solter, D.; Knowles, B.B. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **2004**, *7*, 597–606. [[CrossRef](#)] [[PubMed](#)]
74. Hughes, J.F.; Coffin, J.M. Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: Implications for human and viral evolution. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1668–1672. [[CrossRef](#)] [[PubMed](#)]
75. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **2017**, *18*, 71–86. [[CrossRef](#)] [[PubMed](#)]
76. Sundaram, V.; Wysocka, J. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philos. Trans. R. Soc. B* **2020**, *375*, 20190347. [[CrossRef](#)] [[PubMed](#)]

77. Fueyo, R.; Judd, J.; Feschotte, C.; Wysocka, J. Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 481–497. [[CrossRef](#)]
78. Yang, J.; Cook, L.; Chen, Z. Systematic evaluation of retroviral LTRs as cis-regulatory elements in mouse embryos. *Cell Rep.* **2024**, *43*, 113775. [[CrossRef](#)] [[PubMed](#)]
79. Berger, M.F.; Badis, G.; Gehrke, A.R.; Talukder, S.; Philippakis, A.A.; Pena-Castillo, L.; Alleyne, T.M.; Mnaimneh, S.; Botvinnik, O.B.; Chan, E.T. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **2008**, *133*, 1266–1276. [[CrossRef](#)]
80. Gagnon, L. Imaging the genome in 3D at super resolution. *Microsc. Today* **2020**, *28*, 18–26. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.