The transcription factor SRF regulates MERVL retrotransposons and gene expression during zygotic genome activation

Clara Hermant,[1,3] Carlos Michel Mourra-Díaz,[1,3] Marlies E. Oomen,[1] Luis Altamirano-Pacheco,[1] Mrinmoy Pal,[1] Tsunetoshi Nakatani,[1] and Maria-Elena Torres-Padilla[1,2]

[1]Institute of Epigenetics and Stem Cells (IES), Helmholtz Zentrum München, D-81377 München, Germany; [2]Faculty of Biology, Ludwig-Maximilians Universität, D-81377 München, Germany

The regulatory circuitry of cell-specific transcriptional programs is thought to be influenced by transposable elements (TEs), whereby TEs serve as raw material for the diversification and genome-wide distribution of genetic elements that contain *cis*-regulatory activity. However, the transcriptional activators of TEs in relevant physiological contexts are largely unknown. Here, we undertook an evolutionary approach to identify regulators of two main families of MERVL, a major regulator of transcription during early mouse development. Using a combination of phyloregulatory, transcriptomic, and loss-of-function approaches, we demonstrate that SRF is a novel regulator of MERVL and embryonic transcription during zygotic genome activation. By resolving the phylogenetic history of two major MERVL families, we delineate the evolutionary acquisition of SRF and DUX binding sites and show that the acquisition of the SRF site precedes that of DUX. SRF contributes to embryonic transcription through the regulation of MERVLs, which in turn serve as promoters for host genes. Our work identifies new transcriptional regulators and TEs that shape the gene expression programs in early embryos and highlights the process of TE domestication via the sequential acquisition of transcription factor binding sites and coevolution with the host.

[*Keywords*: MERVL; mouse embryos; retrotransposons; transcription]

Supplemental material is available for this article.

Received September 2, 2024; revised version accepted January 30, 2025.

Transposable elements (TEs) and their remnants dominate the content of mammalian genomes (Lander et al. 2001; Mouse Genome Sequencing Consortium 2002). While their potential in driving genome evolution and regulatory innovation is increasingly recognised, the molecular mechanisms that enable such attributions remain obscure. TEs can shape developmental programs across species, primarily by acting as and spreading *cis*-regulatory elements with the capacity to drive specific transcriptional programs (Bourque et al. 2008; Feschotte 2008; Chuong et al. 2017) or by modulating chromatin accessibility (Jachowicz et al. 2017). While the mechanisms underlying TE transcriptional repression have been extensively studied, the factors that activate TE transcription remain largely unknown.

In mammals, fertilization of the oocyte by the sperm initiates development. The earliest developmental stages after fertilization are characterized by a robust transcriptional activation of TEs. TEs from all classes, including LINE1, SINE, and long-terminal repeat (LTR) elements, become expressed (Evsikov et al. 2004; Peaston et al. 2004; Fadloun et al. 2013). Although there is rich specificity in terms of the temporal kinetics and specific subclasses of TEs expressed, in general, the most abundant transcriptional activity occurs at the 2-cell stage in mice, the time at which the major wave of zygotic genome activation (ZGA) occurs (Fadloun et al. 2013; Rodriguez-Terrones and Torres-Padilla 2018). In addition, ~4000 "host" genes are transcriptionally activated simultaneously during ZGA (Aoki et al. 1997; Hamatani et al. 2004; Jukam et al. 2017). MERVL, a mouse-specific ERVL, has emerged as a key TE expressed during ZGA, and its promoter, the MT2_Mm LTR, regulates the expression of ~1000 genes during this developmental window (Sakashita et al. 2023). Thus, TEs have been

[3]**These authors contributed equally to this work.**
Corresponding author: torres-padilla@helmholtz-muenchen.de
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.352270.124.

demonstrated to play a role during the transition from the maternal to the embryonic control. However, the regulators of MERVL themselves are not fully characterized. In vitro, several transcription factors (TFs) such as GATA2, DUX, ZSCAN4C, and OBOX can bind to and activate MERVL and consequently ZGA genes, also referred to as the "2C" program, in 2-cell-like cells (2CLCs) (Choi et al. 2017; De Iaco et al. 2017; Hendrickson et al. 2017; Whiddon et al. 2017; Zhang et al. 2019; Ji et al. 2023; Guo et al. 2024). However, in vivo, deletion of DUX or OBOX in mouse embryos, although resulting in a strong ZGA phenotype, does not fully abrogate transcription of MERVL and ZGA genes (Chen and Zhang 2019; Guo et al. 2019, 2024; De Iaco et al. 2020; Bosnakovski et al. 2021; Ji et al. 2023). This indicates that additional TFs play redundant roles in regulating TE transcription at these stages. In addition, the evolutionary path that led to the co-option of MERVL as a key regulator of ZGA is not known.

Here, we have reconstructed the evolutionary history of MERVL and its transcriptional regulators. By performing phyloregulatory analyses combining phylogenetics and transcription factor binding site profiling, we establish the evolutionary point at which the DUX motif emerged in MERVL. In addition, we identify SRF as a transcription factor with motifs widespread in the MT2 family, which appeared prior to the DUX motifs. We demonstrate that SRF regulates MERVL, thus establishing SRF as a new TF for TEs. By performing loss-of-function experiments in mouse embryos, we further demonstrate that SRF contributes to gene and TE expression during ZGA. Our work highlights co-option in the deployment of developmental programs and expands our conceptual understanding of the evolutionary arms race between TEs and the host beyond repression strategies toward adaptation of (host) TF binding sites for transcriptional activation.

Results

Phylogenetic analysis of the MERVL LTR uncovers sequence heterogeneity linked to transcription factor binding motifs

In order to investigate the regulation of TEs, we undertook an evolutionary approach with a specific focus on TEs with roles in early embryogenesis. We first sought to determine the evolutionary history of MERVL. MERVL and its LTR, MT2_Mm, are *Mus*-specific. Early studies on a handful of genomic insertions indicated that MERVL/MT2_Mm experienced a first wave of expansion soon after the divergence between *Mus* and *Rattus* ~14 million years ago and a second wave occurred ~2 million years ago (Bénit et al. 1999; Costas 2003). We performed a phylogenetic analysis (Carter et al. 2022) using all MT2_Mm insertions in the mouse genome ($n = 2776$), which we first subjected to a size distribution analysis (Supplemental Fig. S1A). This analysis revealed that most MT2_Mm insertions correspond to an intact LTR and display a median length of 493 bp ($n = 2307$) (Supplemental Fig. S1B). Multiple sequence alignments of all complete MT2_Mm LTRs, followed by reconstruction of a phylogenetic tree (Supplemental Fig. S1A), revealed the existence of new subfamilies, which were defined as such if they formed a cluster of at least 30 sequences with branch length >0.015 and supported by a node with confidence >95%. Although these parameters are semiarbitrary, we tested several parameters in order to obtain a clear distinction of visible, defined clades but without subfamilies with too few insertions, as they would be likely uninformative. Using these criteria, we classified MT2_Mm insertions into five new MT2_Mm subfamilies that we named MT2_Mm_i to MT2_Mm_v based on their order of appearance on the tree as well as their number of insertions (Fig. 1A). These families vary in their number of genomic insertions—from MT2_Mm_v, which contains 47 insertions, to the largest, MT2_Mm_ii, which included ~55% of all complete MT2_Mm insertions in the genome ($n = 1291$) (Fig. 1A). The subfamily MT2_Mm_i, composed of 568 sequences, is the most evolutionarily distant from all the other subfamilies (Fig. 1A). Thus, using phylogenetics, we defined new subfamilies and revealed a previously unknown heterogeneity of the LTR of MERVL.

To explore the potential significance of such heterogeneity, we asked whether all MT2_Mm subfamilies are expressed in mouse embryos. For this, we reanalyzed our RNA-seq data sets from SMART-seq +5′, which captures the 5′ end of transcripts, allowing us to precisely map the transcription start site (TSS) from each TE insertion and thereby measure their autonomous transcription levels accurately (Oomen et al. 2025). Expression of all the MT2_Mm subfamilies displays the same temporal dynamics as the complete MT2_Mm family (Fig. 1B; Supplemental Fig. S1D-H), which shows the highest transcript abundance at the late 2-cell stage (Supplemental Fig. S1C), as described previously (Kigami et al. 2003; Svoboda et al. 2004; Macfarlan et al. 2012; Ishiuchi et al. 2015; Rodriguez-Terrones and Torres-Padilla 2018; Sakashita et al. 2023). Interestingly, however, insertions from the different MT2_Mm subfamilies show varying levels of expression, with MT2_Mm_i displaying lowest expression levels on average per insertion (Fig. 1B). These differences were also evident when we examined expression levels per individual insertion by plotting the expression of all insertions from each subfamily (Fig. 1C). These findings prompted us to investigate whether differences in expression levels of MT2_Mm subfamilies are related to their evolutionary age, because it is known that for other TEs, such as LINE1, younger elements are more transcriptionally active than older ones (DeBerardinis and Kazazian 1999; Goodier et al. 2001). To address this, we established the genetic divergence of MT2_Mm subfamilies, which represents the genetic distance from each sequence and thus a proxy for their evolutionary age. We used an "outgroup rooting" method (Steel 2010; Kinene et al. 2016), which relies on rooting the phylogenetic tree using a known near ancestor and determines the genetic distance of each MT2_Mm insertion relative to the root. For phylogenetic inference, we used iQ-Tree because it estimates ancestral sequences for a more accurate computation of

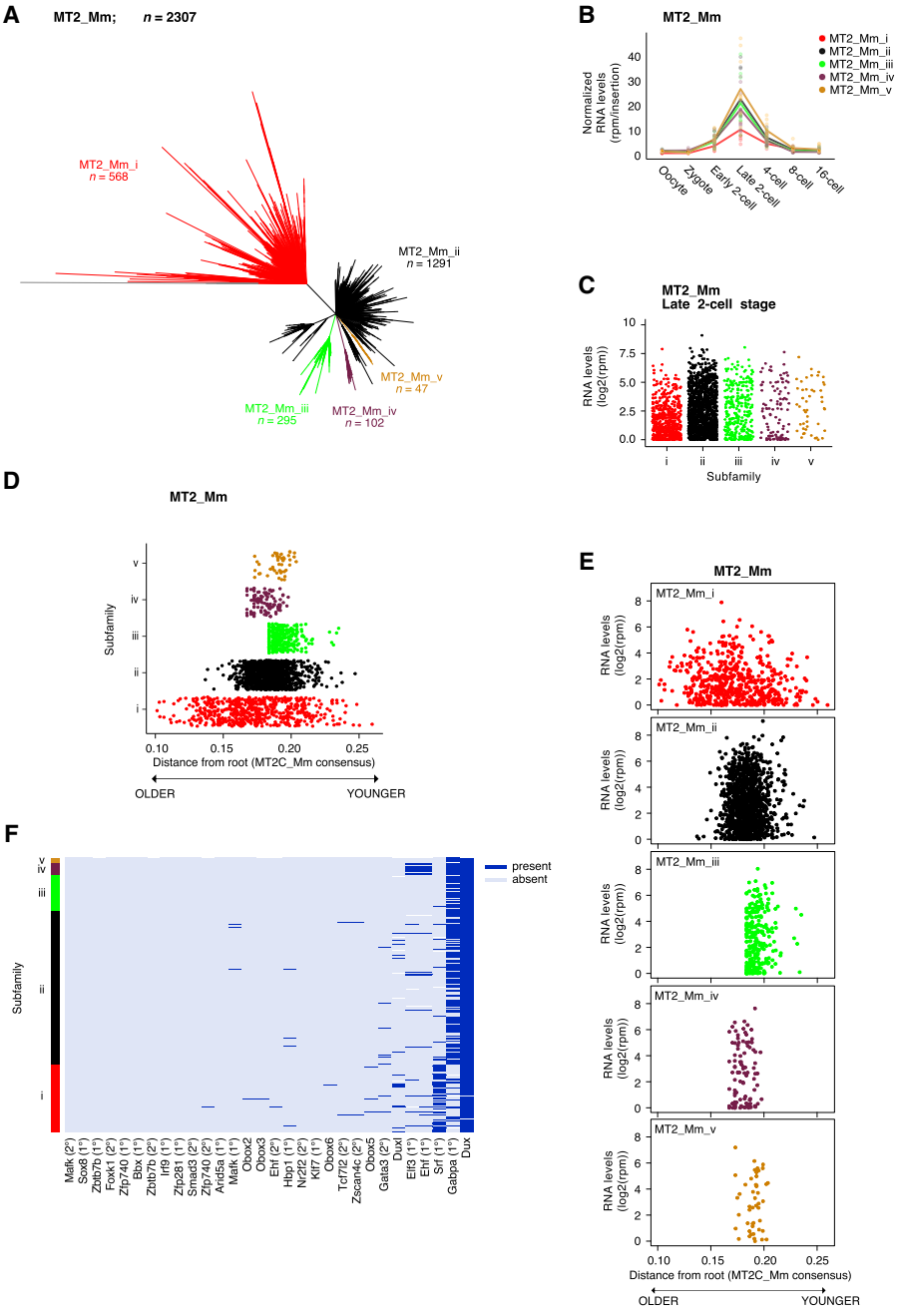

Figure 1. Phylogenetic analysis of the MERVL LTR reveals sequence heterogeneity linked to specific transcription factor binding motifs. (*A*) Unrooted phylogenetic tree of the insertions of the MERVL LTR MT2_Mm, comprising all 5′, 3′, and solo LTRs, ranging between 400 and 586 bp in length. Subfamilies are highlighted in color; *n* is the number of insertions per group indicated. Sequences in gray correspond to outgroups (in this case, only four sequences, which therefore did not qualify as new subfamilies). (*B*) Expression of the five MT2_Mm subfamilies across preimplantation development. Each dot represents the sum reads per million (RPM) of all MT2_Mm insertions belonging to one subfamily per single embryo at the indicated stage, normalized by the number of insertions of each subfamily. The trend line connects the mean values across embryos for each stage. (*C*) Expression of single MT2_Mm insertions at the late 2-cell stage. Each dot is the $\log_2$ transformed mean RPM of single insertions in all the embryos ($n = 10$) (Oomen et al. 2025), categorized by the subfamily to which each insertion belongs. (*D*) Divergence analysis of MT2_Mm single insertions. Each dot is an insertion, and the position of the *X*-axis corresponds to the computed genetic distance to the root of the tree (namely, MT2C_Mm). Single insertions are ordered and colored by subfamily. The distribution of MT2_Mm subfamily iii (green) and, to some extent, subfamily iv (bordeaux) is influenced by polytomies, which reflect the fact that a single ancestral lineage gives rise to three or more descendant lineages simultaneously, and thus the genetic distances to the ancestor occupy the same position on the *X*-axis. This is most likely due to the lack of enough data to robustly conclude how some elements within these subfamilies (lineages) are related. (*E*) Expression levels against genetic divergence of single MT2_Mm insertions. Each dot is an insertion, and the position on the *X*-axis corresponds to the computed genetic distance to the root of the tree, while the position on the *Y*-axis corresponds to the $\log_2$ transformed mean RPM of each insertion in all late 2-cell stage embryos ($n = 10$) (Oomen et al. 2025). Insertions are categorized by the subfamily to which they belong. (*F*) Heat map showing the presence/absence of transcription factor (TF) binding sites (TFBSs) in individual MT2_Mm insertions. Each line is an individual insertion, ordered by subfamily as indicated by the bar legend at the *left*. The presence of a binding site for the indicated TF is displayed in dark blue. TFs are clustered based on the number of insertions containing a binding site, from none (*left*) to the most insertions (*right*), and subsequently the heat map was sorted by subfamily without additional specific order. Primary (1°) and secondary (2°) binding sites are indicated as per UniPROBE terminology: The 1° site refers to a higher-confidence site prediction than the 2° site. For DUX, we added the motif manually because it is absent in UniPROBE, and for OBOX, we used the shorter motif from UniPROBE instead of the extended de novo motif recently identified for OBOX1/3/5 (Ji et al. 2023).

phylogenetic relationships, models the best parameters without a priori input knowledge, and most importantly, compares between several substitution models and chooses the model that provides the highest robustness to the phylogeny. We first confirmed that MT2C_Mm is a *Mus*-specific close ancestor that shares common ancestry

with MT2_Mm (Supplemental Fig. S1I), as suggested previously (Franke et al. 2017). Computing the genetic divergence of all individual insertions of MT2_Mm subfamilies revealed that the oldest elements, which are the closest to MT2C_Mm, belong to MT2_Mm_i (Fig. 1D). These analyses also indicate that the emergence of MT2_Mm_i was followed by the expansion of MT2_Mm_ii, followed by iv and v and then by subfamily iii. MT2_Mm subfamilies i, ii, and iii all have recent insertions (Fig. 1D). Subfamily i spreads substantially across genetic distances compared with subfamilies iv and v, which appear to have ceased expansion despite being younger than MT2_Mm_i (Fig. 1D). This can be appreciated based on the distribution pattern along the *X*-axis in Figure 1D but also by the number of insertions, which is therefore not necessarily correlated with their age. Thus, there are differences in the spread of the predicted ages between subfamilies. This suggests that the colonization of the latter two subfamilies in the mouse genome has been less efficient and was restricted to a limited period in time. Additionally, the overall expression level of each subfamily is not related to their evolutionary age. Indeed, plotting expression levels of individual insertions against their age indicated that specific insertions vary in their expression levels regardless of their age, and this is consistent across all MT2_Mm subfamilies (Fig. 1E). Interestingly, subfamily i, which is also the oldest subfamily, is mostly composed of solo LTRs (80% of the insertions), whereas the other four MT2_Mm subfamilies only comprise between 5% and 30% solo LTRs (Supplemental Fig. S1J). This suggests that the internal MERVL sequences have been lost during evolution. The evolutionary expansion pattern of MT2_Mm that we report here differs from that of other TEs, such as LINE1, which typically expand through sequential waves in the host genomes (Castro-Diaz et al. 2014). We conclude that the evolutionary trajectory of MT2_Mm is characterized by an older subfamily that has expanded and prevailed in the mouse genome, mostly as solo LTRs, suggesting distinctive evolutionary pressures and/or colonization strategies by the LTR MT2_Mm.

The distinct phylogeny and the different expression levels of MT2_Mm subfamilies prompted us to investigate whether their sequence evolution is linked to the acquisition of specific transcription factor (TF) binding sites (TFBSs). To address this, we used publicly available ATAC-seq data sets (Wu et al. 2016) to identify TF footprints on the MT2_Mm consensus LTR at the late 2-cell stage. We then searched for the most likely TFs bound to these footprints based on their annotated DNA binding motifs in the UniPROBE database (Newburger and Bulyk 2009; Tanaka et al. 2011; Bailey et al. 2015; Hume et al. 2015). We further filtered the list of TFs identified and considered only those expressed at the 2-cell stage according to RNA-seq data (Deng et al. 2014). This resulted in a list of 25 TFs belonging to different families, which included known ZGA regulators such as ZSCAN4C and OBOX TFs (Fig. 1F; Zhang et al. 2019; Ji et al. 2023; Guo et al. 2024). Indeed, OBOX1/3/5 have been shown to directly bind MT2_Mm in 2-cell embryos (Ji et al. 2023), which is consistent with our TFBS prediction. We also added the DUX binding motif, previously identified in MT2_Mm (Hendrickson et al. 2017) but not included in UniPROBE. We then searched all individual MT2_Mm insertions within the subfamilies that we identified for the presence of TFBSs for these 26 TFs (Fig. 1F; Santana-Garcia et al. 2022). Interestingly, we found that the presence of the binding motif for specific TFs demarcates the MT2_Mm subfamilies identified phylogenetically. For example, MT2_Mm insertions in subfamily iv are dominated by the presence of TFBSs for ELF3 and EHF (Fig. 1F). Remarkably, all subfamilies contain the DUX binding site, but subfamily i is distinguished from all other subfamilies by the presence of a TFBS for SRF (Fig. 1F). Instead, all other MT2_Mm subfamilies (ii–v) contain a GABPA motif, which appears to be mutually exclusive to the SRF motif (Fig. 1F). Thus, this analysis suggests that the phylogeny of MT2_Mm, and thus its evolution in the mouse genome, are linked to the presence of specific TFBSs.

The evolutionarily older MT2C_Mm contains a DUX binding site and an SRF motif that appeared prior to DUX in evolution

To better understand how and whether TFBSs have shaped the evolution of MT2_Mm, we turned to the closest ancestor of MT2_Mm, which is known to be MT2C_Mm (Franke et al. 2017), which has a LTR size distribution and number of insertions in the mouse genome ($n = 2046$) similar to those of MT2_Mm (Supplemental Fig. S2A). Only solo LTRs for MT2C_Mm are found in the mouse genome, presumably due to its older evolutionary age, and we excluded fragmented elements from this analysis (Supplemental Fig. S2A). We performed phylogenetic analysis on MT2C_Mm as above, which led to the classification of MT2C_Mm into nine new MT2C_Mm subfamilies (Fig. 2A). The number of insertions per subfamily is relatively homogeneous and ranges from 45 in the smallest subfamily to 206 in the largest subfamily (Fig. 2A). Computing the genetic divergence of MT2C_Mm insertions indicated that the oldest insertions belong to subfamily i and the youngest insertions primarily belong to subfamily vii (Fig. 2B). This analysis also revealed that the colonization of MT2C_Mm in the mouse genome occurred through sequential waves of expansion, initially by subfamily i to the newest subfamily, which is subfamily ix (Fig. 2B). This expansion was closely followed by the appearance of MT2_Mm, in particular MT2_Mm subfamily i (Supplemental Fig. S2B). Interestingly, plotting the genetic distance of MT2_Mm and MT2C_Mm insertions combined resulted in a slightly different pattern. Indeed, in the combined phylogenetic tree, the genetic distance is based on one root for all sequences, which has to be the closest ancestor, MT2B (Supplemental Fig. S1I; Franke et al. 2017). Thus, because the genetic distances are relative, the structure of the tree changes because the older MT2C_Mm insertions cluster further away from MT2_Mm. In fact, the combined structure of the tree suggests that MT2C_Mm_i and MT2C_Mm_ii are derived from the same common ancestor, but that

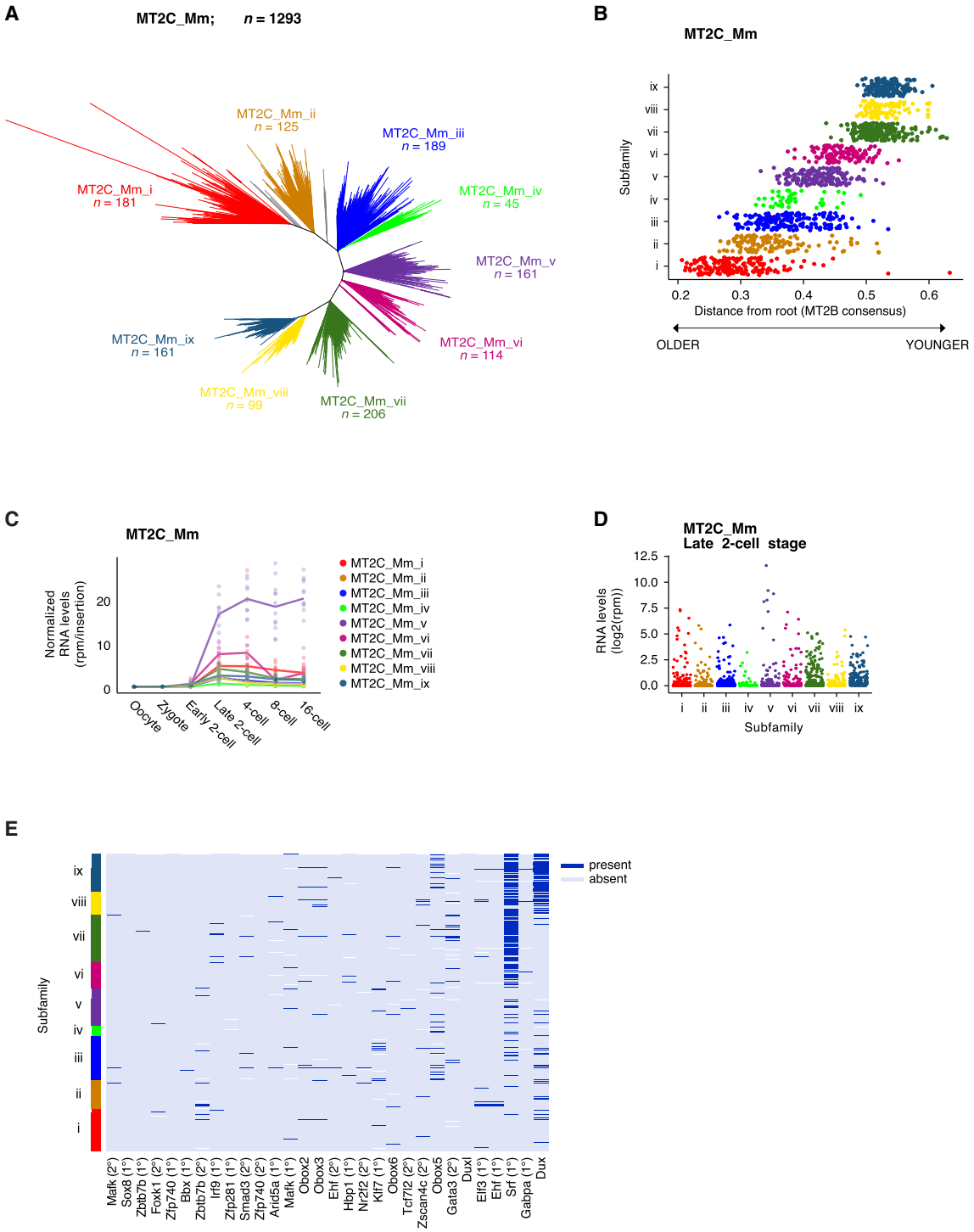

Figure 2. The evolutionarily older MT2C_Mm contains a DUX binding site and an SRF motif that appeared prior to DUX in evolution. (*A*) Unrooted phylogenetic tree of MT2C_Mm insertions, ranging between 385 and 565 bp in length. Subfamilies are highlighted in color; *n* is the number of insertions per group indicated. Sequences in gray correspond to outgroups (in this case, only 12 sequences, which therefore did not qualify as new subfamilies). (*B*) Divergence analysis of MT2C_Mm single insertions. Each dot is an insertion, and the position on the *X*-axis corresponds to the computed genetic distance to the root of the tree (namely, MT2B consensus). Single insertions are ordered and colored by subfamily. (*C*) Expression of the nine MT2C_Mm subfamilies across preimplantation development. Each dot represents the sum RPM of all MT2C_Mm insertions belonging to one subfamily per single embryo at the indicated stage, normalized by the number of insertions of each subfamily. The trend line connects the mean values across embryos for each stage. (*D*) Expression of single MT2C_Mm insertions at the late 2-cell stage. Each dot is the $\log_2$ transformed mean RPM of single insertions in all the embryos (n = 10) (Oomen et al. 2025), categorized by the subfamily to which each insertion belongs. (*E*) Heat map showing the presence/absence of TFBSs in individual MT2C_Mm insertions. Each line is an individual insertion, ordered by subfamily as indicated by the legend at the *left*. The presence of the binding site for the indicated TF is displayed in dark blue. TFs are arranged in the same order as for the MT2_Mm TF heat map in Figure 1F, and the heat map was sorted by subfamily without additional specific ordering.

MT2C_Mm_i does not follow the same evolutionary lineage as the other MT2C_Mm subfamilies (Fig. 2A; Supplemental Fig. S2B). We obtained overall similar results by using a molecular clock-based method using the Jukes–Cantor substitution model (Supplemental Fig. S2C; Jukes and Cantor 1969). It is noteworthy that the pattern of genome expansion of MT2C_Mm is thus different from that of MT2_Mm. Interestingly, interrogating our SMART-seq +5′ data sets demonstrated that MT2C_Mm is also transcribed in early mouse embryos, with a clear ZGA profile reflected by its transcriptional induction at the late 2-cell stage (Supplemental Fig. S2D). However, the expression levels varied between subfamilies. This is evident when plotting normalized expression levels (reads per million [RPM]) per insertion, which indicates that subfamily v displays the highest transcriptional activity per insertion, and subfamily iv displays the lowest (Fig. 2C). Additionally, we found that individual insertions from all MT2C_Mm subfamilies are expressed at the 2-cell stage, as seen when plotting expression levels per insertion (Fig. 2D). Similarly to MT2_Mm, we found no correlation between age and expression of individual insertions (Supplemental Fig. S2E). Globally, the greatest number of MT2C_Mm transcripts in 2-cell embryos derive from MT2C_Mm subfamily v, due primarily to one insertion expressed very highly [Fig. 2D, note the $\log_2$(RPM) scale]. Notwithstanding, all individual MT2C_Mm subfamilies undergo transcriptional activation at the late 2-cell stage, indicating that MT2C_Mm becomes activated at ZGA (Supplemental Fig. S2F–N). Thus, considering the different expression levels of MT2C_Mm, we also wondered whether the different subfamilies are characterized by distinctive TF binding motifs. In light of our evolutionary analysis of MT2_Mm, we specifically asked whether the TFBSs that we identified in MT2_Mm are also present in the MT2C_Mm subfamilies. Compared with MT2_Mm, MT2C_Mm displays only a few insertions with EHF, ELF3, GABPA, or OBOX binding motifs, but none of these appear to be distributed in specific MT2C_Mm subfamilies (Fig. 2E). However, our TFBS analysis revealed differences in the presence of the motifs of TF across the different MT2C_Mm subfamilies (Fig. 2E). Namely, the SRF binding site is present throughout all MT2C_Mm subfamilies, though it is more prevalent in subfamilies vi–ix, suggesting that the SRF binding site prevailed on MT2C_Mm from subfamily vi onward, following MT2C_Mm's evolutionary dynamics (Fig. 2E). Interestingly, we noted that a binding site for DUX is present predominantly in the most recent subfamilies (viii and ix) only, suggesting that the well-studied DUX binding site in MERVL first arose in the youngest MT2C_Mm insertions, only after the appearance of the SRF binding site (Fig. 2E). Indeed, our observations showing that the oldest MT2C_Mm families do not possess a DUX binding motif but the younger families do (Fig. 2E) indicate that DUX arose during the evolution of MT2C_Mm. Thus, these analyses suggest both that SRF may be a TF regulating the MT2 classes MT2C_Mm and MT2_Mm and that the DUX binding site was acquired prior to the expansion of MT2_Mm in the mouse genome.

The DUX binding site originated from a deletion during MT2C_Mm evolution prior to the expansion of MT2_Mm

The above results prompted us to investigate more specifically how the DUX binding site emerged during the evolution of MT2 and whether SRF is a regulator of MT2C_Mm and MT2_Mm. We first asked whether we could delineate the evolutionary time and the manner in which the DUX binding site originated. For this, we generated a majority rule consensus sequence for all subfamilies of MT2C_Mm and MT2_Mm (nine plus five, respectively) (Fig. 3A). Aligning and scanning these 14 consensus sequences with the matrix-scan tool from RSAT (Santana-Garcia et al. 2022) revealed the origin of the DUX binding site in MT2_Mm and showed that it first arose in MT2C_Mm subfamily viii through a 9 bp deletion event (Fig. 3A, purple text). We also observed two changes at the single-nucleotide level that are predicted to change the affinity of the SRF binding site: A C-to-T change from consensus of MT2C_Mm subfamily vi to subfamily vii creates a stronger predicted SRF binding site, and a T-to-A change between MT2_Mm subfamily i and subfamily ii is thought to result in a lower-affinity SRF binding site (Fig. 3A, cyan text). We next reconstructed the complete phylogeny of MT2C_Mm and MT2_Mm together using a median-joining network analysis, which confirmed the genetic divergence analysis. Namely, MT2C_Mm_i is the oldest subfamily and closest to MT2B, and a stepwise amplification commences with MT2C_Mm_ii up until the appearance of the most recent MT2_Mm_iii subfamily (Fig. 3B). Incorporating the TFBS analysis into this phylogeny confirmed the temporal origin of the DUX binding site in MT2C_Mm and suggests a potential reduction in the affinity of SRF in the more recent MT2_Mm subfamilies (Fig. 3B).

SRF can activate transcription of MT2C_Mm and MT2_Mm

Next, we addressed directly whether SRF can activate transcription of MT2C_Mm and MT2_Mm and whether our predictions on the evolutionary origin of the DUX binding site do indeed reflect its transactivation capacity on MT2C_Mm. For this, we first used a heterologous system based on a luciferase reporter plasmid in which the LTRs of MT2C_Mm or MT2_Mm drive luciferase expression in HEK293 (human embryonic kidney) cells with cotransfection of either SRF or DUX expression vectors (Fig. 3C). We chose HEK293 as an unrelated cell type to minimize the potential confounding effects of endogenous transcription factors and because, as a human cell line, it does not contain endogenous MT2 elements. We verified that both SRF and DUX were expressed upon transfection in HEK293 cells (Supplemental Fig. S3A). We chose to use LTRs from specific MT2C_Mm and MT2_Mm subfamilies harboring different combinations of TFBSs: MT2C_Mm with SRF but no DUX site (subfamily vii) or with SRF and DUX sites (subfamily ix), and MT2_Mm with SRF and DUX sites (subfamily i) or with DUX but no SRF site (subfamily ii) (Fig. 3A,C). For all

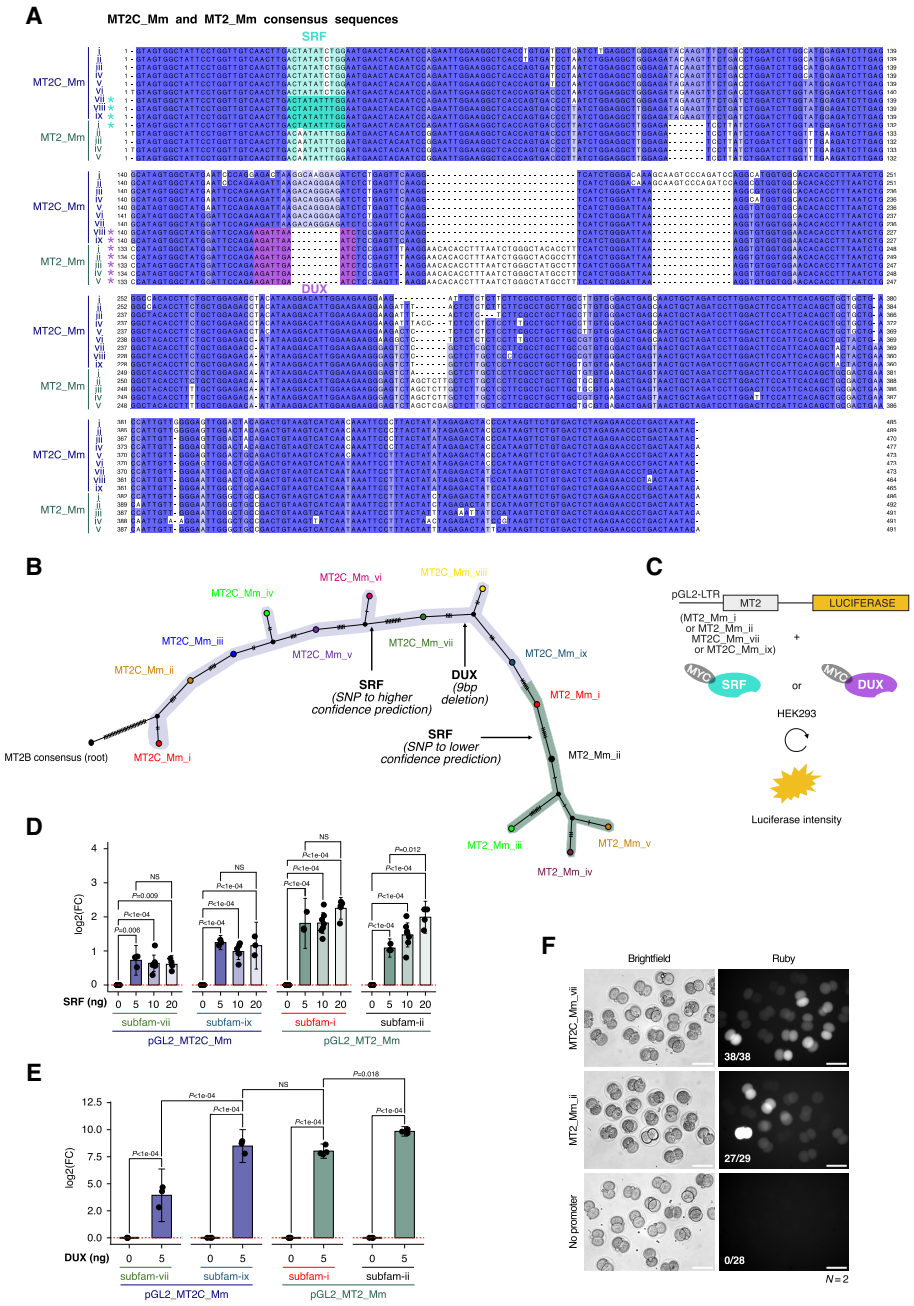

Figure 3. The DUX binding site originated from a deletion prior to the expansion of MT2_Mm, and SRF can activate MERVL families MT2C_Mm and MT2_Mm. (*A*) Multiple sequence alignment of MT2_Mm and MT2C_Mm subfamily consensus sequences. The positions of binding motifs for SRF and DUX are indicated in cyan and purple, respectively. The SRF site of higher-confidence prediction is indicated in darker cyan, and the SNPs leading to lower-confidence prediction are in white. (*B*) Median-joining network analysis of MT2_Mm and MT2C_Mm consensus sequences. The evolutionary orientation of the network was inferred by rooting with the Dfam MT2B consensus. The number of ticks between each consensus represents the number of mutations at nongaps. MT2C_Mm subfamilies are highlighted in purple, and MT2_Mm subfamilies are in dark green. The TFBS acquisitions/gains in prediction confidence points are indicated. (*C*) Schematic overview of the experimental design used to test the capacity of DUX and SRF to transcriptionally activate four MT2 subfamilies: MT2C_Mm_vii, MT2C_Mm_ix, MT2_Mm_i, and MT2_Mm_ii. (*D*) Log_2 transformed fold change of the normalized luciferase activity of four subfamilies in the presence of 5, 10, and 20 ng of pCMV-encoding SRF coding sequence over control (pCMV-empty). Bar plots show the mean ± 95% confidence intervals ($N \geq 3$). For statistics, preselected hypotheses compared each group with 0, and 20 ng was compared with 5 ng. (*E*) Log_2 transformed fold change of the normalized luciferase activity of four subfamilies in the presence of 5 ng of pCMV-encoding DUX coding sequence over control (pCMV-empty). Bar plots show the mean ± 95% confidence intervals ($N \geq 3$). For statistics, preselected hypotheses compared each group with 0, and log_2(FC) of MT2_Mm_ii over control was compared with that of MT2_Mm_i, MT2_Mm_i was compared with MT2C_Mm_ix, and finally, MT2C_Mm_ix was compared with MT2C_Mm_vii. (*F*) Representative images of late 2-cell stage embryos after microinjection with the indicated reporter plasmid. The numbers indicate the number of embryos with fluorescent signal over the total number of embryos injected, analyzed from two independent replicates ($N = 2$). Scale bars, 100 µm. We note that these experiments are prone to microinjection variability, and therefore levels of fluorescence intensity should not be compared quantitatively.

LTRs, we used the consensus sequence. Cotransfection with SRF resulted in transactivation of both MT2C_Mm LTRs reporters at comparable levels (Fig. 3D). However, and in line with our TFBS and median-joining network analyses, SRF did not activate a reporter with MT2C_Mm_i (Supplemental Fig. S3B,C). SRF activated luciferase expression from both the MT2_Mm reporters, though the LTR from subfamily i reached saturation at lower SRF concentrations, suggesting a potentially higher capacity of SRF to more readily activate MT2_Mm_i compared with MT2_Mm_ii (Fig. 3D). Interestingly, transfection with a DUX expression vector resulted in activation of the MT2C_Mm reporter plasmid, in concordance with the predictions based on our TFBS analysis (Fig. 3E). Namely, transfection with a DUX expression vector resulted in activation of the MT2C_Mm reporter plasmid of both MT2C_Mm_ix and MT2C_Mm_vii, but the extent of the transcriptional activation of MT2C_Mm_ix (which contains the deletion that leads to the creation of the DUX binding site) was on average 25 times higher than the induction of MT2C_Mm_vii (Fig. 3E). Importantly, the transcriptional induction of MT2C_Mm_ix was of an extent similar to that of MT2_Mm_i and MT2_Mm_ii, which are known to be robustly transcriptionally activated by DUX (Fig. 3E; Hendrickson et al. 2017; Whiddon et al. 2017). Of note, DUX overexpression did not induce transcriptional activation of MT2C_Mm_i (Supplemental Fig. S3C). These data confirm the acquisition of a functional DUX binding site in MT2C_Mm_ix. We note that the transactivation activity of DUX is much higher compared with SRF, which is consistent with the fact that DUX is a pioneer TF with strong transactivation capacity. Last, DUX was able to robustly activate expression of both MT2_Mm reporters (Fig. 3E), in line with the known activity of DUX over MT2_Mm (Hendrickson et al. 2017; Whiddon et al. 2017; De Iaco et al. 2020). Importantly, control, empty reporter plasmid without the MT2 LTRs or a scrambled LTR sequence did not show transactivation by DUX or SRF (Supplemental Fig. S3D–F). We conclude that a DUX functional site emerged on MT2C_Mm and that SRF can activate transcription of MT2C_Mm and MT2_Mm. Because of the predicted GABPA binding motif in MT2_Mm, specifically from the emergence of subfamily ii onward (Fig. 1F), we also assessed whether GABPA can transactivate MT2_Mm. However, we did not detect luciferase activity on the MT2_Mm_i and MT2_Mm_ii reporters upon expression of GABPA and its obligatory partner, GABPB1 (Supplemental Fig. S3G–I), and thus we focused on SRF for subsequent experiments. Our data indicate that the DUX binding site appeared prior to the MT2_Mm expansion. In addition, the above data indicate that SRF is an ancient transcription factor of MERVL and that SRF can activate transcription of MT2_Mm and MT2C_Mm in a cell culture model. Last, we asked whether an MT2C_Mm reporter from a subfamily without a DUX binding site is capable of driving transcription in mouse embryos. For this, we generated a Ruby fluorescence reporter downstream from the LTR of MT2C_Mm subfamily vii, the latest subfamily of MT2C_Mm that does not contain a DUX TFBS but does contain an SRF motif (Fig. 3A,B), and microinjected this MT2C_Mm_vii reporter into mouse zygotes. Imaging these embryos at the 2-cell stage indicated that MT2C_Mm_vii drives transcription of the Ruby reporter in mouse embryos at the 2-cell stage (Fig. 3F), in contrast to a "no promoter" control (Fig. 3F). As expected, MT2_Mm_ii, which is the largest MT2_Mm subfamily and contains a DUX motif, also led to transcription of Ruby in 2-cell stage mouse embryos (Fig. 3F). Thus, these observations suggest that the younger MT2C_Mm subfamilies that do not contain a DUX motif but contain a functional SRF site are active in vivo.

SRF regulates embryonic transcription in early mouse embryos

SRF is a conserved, ubiquitously expressed MADS-box-containing transcription factor that binds serum response elements and is best known for its role in mediating cell proliferation and immediate early response (Greenberg and Ziff 1984; Treisman 1986, 1987; Greenberg et al. 1987; Miano 2010). SRF is largely conserved across mammals, and SRF-like genes are found in many organisms, including *Xenopus* and *Drosophila* (Mohun et al. 1991; Affolter et al. 1994). The MADS-box, which is the DNA-binding domain of SRF and SRF-like proteins, is highly conserved (Shore and Sharrocks 1995). Initially identified for its ability to mediate serum responses in cell culture, SRF is also known to be activated by different stimuli, including calcium release, antioxidants, and others (Chai and Tarnawski 2002; Deshpande et al. 2022). Mice homozygous for *Srf* deletion exhibit gastrulation defects and die by E12.5 (Arsenian et al. 1998), but whether SRF plays a role earlier during development has not been addressed, as maternally inherited mRNA could contribute to development prior to zygotic transcription. Indeed, we found that *Srf* exists as a maternally inherited transcript in early embryos, with mRNA present in the mature oocyte, and expression levels increase from the early 2-cell stage, peaking in late 2-cell embryos (Fig. 4A). Interestingly, SRF protein is practically undetectable in the nuclei of zygotes and early 2-cell stage embryos but becomes readily detectable in the nucleus by the late 2-cell stage (Fig. 4B).

Because our results above indicate that SRF is a transcriptional regulator of MERVL families, we next addressed whether SRF regulates transcription in preimplantation embryos, specifically focusing on ZGA. For this, we undertook a loss-of-function approach. Because SRF is ubiquitously expressed and its function is likely required across multiple cell types, we sought to implement a loss of function for SRF specifically in mouse embryos after fertilization and spanning the ZGA period. Accordingly, we first performed siRNA against *Srf* by microinjecting mouse zygotes with siRNA at the time of fertilization. However, using this approach, we observed efficient protein depletion only after the 4-cell stage, which is not suitable to address potential effects on ZGA at the 2-cell stage (data not shown). Thus, we turned to using a dominant negative strategy, taking advantage of a previously characterized, widely used dominant

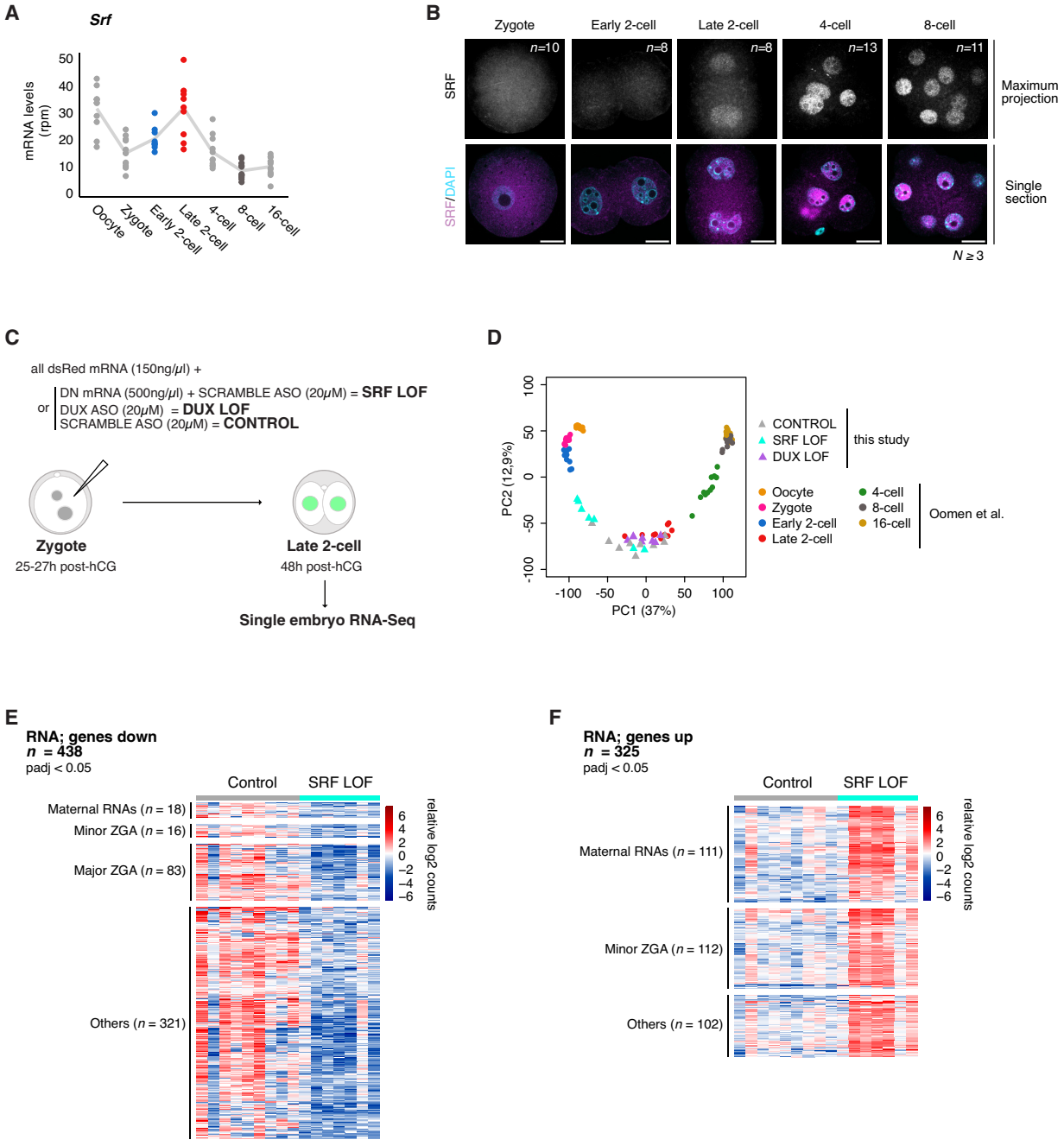

Figure 4. SRF regulates gene transcription during ZGA in mouse embryos. (*A*) Expression of *Srf* during mouse preimplantation development. Each dot represents *Srf* mRNA levels (RPM) in individual embryos at the indicated stages. The trend line connects the mean values across embryos for each stage. (*B*) Representative confocal images of SRF immunostainings at the indicated developmental stage. The intensity of the fluorescent signal is comparable, as all embryos within each replicate were processed and acquired using the same conditions. SRF images are maximum intensity projections, and merged images with DAPI staining to visualize DNA are single confocal sections. *n* is the total number of embryos analyzed per stage. *N* is the number of independent replicates. Scale bars, 20 µm. (*C*) Schematic representation of the experimental design for single-embryo RNA sequencing of DUX and SRF loss of function (LOF). (*D*) Principal component analysis (PCA) of the transcriptional profiles of SRF LOF, DUX LOF, and control embryos in comparison with wild-type nonmanipulated embryos from oocytes to 16-cell stage (Oomen et al. 2025). Each point is an embryo, and the conditions are displayed by the indicated color code. Circles are wild-type embryos from Oomen et al. (2025). Triangles are embryos from this study. The variance explained (percentage) is indicated along the PC1 and PC2 axes. (*E*,*F*) Heat maps showing significantly downregulated (*E*) or upregulated (*F*) genes in SRF LOF late 2-cell stage embryos clustered according to the gene categories as indicated. Values are the $\log_2$ of normalized counts centered on the row mean. *n* is the number of genes in the indicated category. Each column is an individual embryo from either the control or SRF LOF group as indicated. The variability in the extent of the phenotype in SRF LOF embryos (namely, embryos 1 and 6) may be due to slightly different developmental stages.

negative of SRF (SRF DN) that lacks the transactivation domain but contains the DNA binding domain (Supplemental Fig. S4A; Belaguli et al. 1999). We first confirmed that this truncated form of SRF behaves as dominant negative toward wild-type SRF on MT2_Mm using the luciferase reporter assay in HEK293 cells (Supplemental Fig. S4B). We then expressed the SRF DN in mouse zygotes through mRNA microinjection (Fig. 4C). As a technical positive control for successful microinjection we used mRNA for dsRed (Fig. 4C). In addition, we controlled for effective expression of SRF DN by using a GFP fusion version of SRF DN (Supplemental Fig. S4A), which allowed us to distinguish embryos that adequately expressed SRF DN based on nuclear GFP fluorescence (Supplemental Fig. S4C). In parallel, we knocked down DUX using an antisense oligo (ASO) (Guo et al. 2024) as a positive control for potential effects on MT2_Mm expression. We refer to these as SRF and DUX loss of function (LOF), respectively. For all samples, we used embryos microinjected with mRNA for dsRed and scramble (ASO) as negative controls (Fig. 4C). We cultured zygotes until the late 2-cell stage, at the time corresponding to major ZGA, and processed them for single-embryo RNA-seq (Fig. 4C). Following quality control checks based on standard metrics, we analyzed a total of 25 single embryos (Supplemental Fig. S4D,E). Overall, SRF LOF led to changes in expression of 763 genes, with slightly more downregulated genes than upregulated genes (438 downregulated vs. 325 upregulated; $P\text{adj} < 0.05$) (Supplemental Fig. S4F; Supplemental Table S1). These data indicate that SRF regulates embryonic transcription during ZGA. As expected, DUX LOF also led to changes in gene expression, and these were more extensive than those caused by SRF LOF (2592 DE genes; 1543 upregulated and 1049 downregulated; $P\text{adj} < 0.05$) (Supplemental Fig. S4G; Supplemental Table S1). Among the genes downregulated upon DUX depletion, we found all *Zscan4* paralogs (*Zscan4a–f*), *Zfp352*, *Eif4e3*, *Tmem92*, and *Pramef6*, all known to be DUX targets, in agreement with previous work (Supplemental Fig. S4G; Supplemental Table S1; De Iaco et al. 2017; Hendrickson et al. 2017). Interestingly, we found little overlap between the differentially expressed genes in SRF LOF and DUX-depleted embryos (Supplemental Fig. S4H). Thus, SRF and DUX regulate a different set of genes in 2-cell stage embryos.

Principal component analysis (PCA) of all our samples combined with a data set of nonmanipulated, wild-type oocytes and embryos up to the 16-cell stage (Oomen et al. 2025) indicated that control embryos clustered together with late 2-cell stage embryos, as expected (Fig. 4D). Likewise, 2-cell stage embryos depleted of DUX clustered primarily with late 2-cell stage embryos (Fig. 4D), in line with previous work showing that DUX-depleted embryos can develop to the blastocyst stage despite a reduction in the expression of ~25% of ZGA genes (Hendrickson et al. 2017; Chen and Zhang 2019; De Iaco et al. 2020). In contrast, SRF LOF embryos clustered in an intermediate position between early and late 2-cell stage (Fig. 4D). Twenty percent of the top genes contributing to PC1 were differentially expressed in SRF LOF embryos, and only 9% were differentially expressed in DUX LOF (Supplemental Fig. S4I,J). Although it is impossible to formally ascertain to what each PC corresponds, the positions of the oocytes at the leftmost part of PC1 and of the 16-cell stage embryos at the rightmost part suggest that PC1 corresponds to developmental progression and PC2 to ZGA. Thus, the position of the SRF LOF along the PCA suggests a transcriptional defect during the 2-cell stage and coincident with ZGA. Further analysis of differentially expressed genes using the Database of Transcriptome in Mouse Early Embryos (DBTMEE) (Park et al. 2015) indicated that, although SRF LOF resulted in downregulation of 99 ZGA genes (22% of downregulated genes; 16 minor and 83 major ZGA genes), the downregulated genes were not restricted to ZGA genes (Fig. 4E). Upregulated genes included maternal genes (111; 34% of upregulated genes) and minor ZGA genes (112; 34% of upregulated genes) but no major ZGA genes (Fig. 4F). A specific analysis using the maternal gene list confirmed that a large proportion of upregulated genes correspond to maternal transcripts already present in the oocyte (Supplemental Fig. S4K). This suggests that SRF LOF embryos fail to undergo proper maternal-to-zygotic transition, which is reflected by defects in ZGA gene expression but also failure to fully degrade maternal RNA. A similar analysis on DUX LOF embryos indicated that 403 DE genes correspond to maternal transcripts, but these are both downregulated and upregulated (Supplemental Fig. S4L). We conclude that SRF contributes to the regulation of embryonic transcription during ZGA in mouse embryos, with a more widespread transcriptional phenotype not restricted to ZGA genes.

SRF regulates a specific set of TEs during ZGA and drives expression of host chimeric transcripts

Next, we asked whether SRF regulates TE expression in mouse embryos. We first investigated potential changes in TE expression at the family level. SRF LOF resulted in altered expression of 14 different TE families (Supplemental Fig. S5A; Supplemental Table S2). Among them, we found TEs from all superfamilies, including evolutionarily older LINE-1 (L1MCb) but also SINEs (B1_Mur), MaLRs (ORR1C2), ERVL (MLT1N2), ERVK, and DNA transposons (Charlie1a:hA) (Supplemental Table S2). MT2_Mm was largely unaffected ($P\text{adj} = 0.98$) (Fig. 5A; Supplemental Fig. S5A; Supplemental Table S2). In contrast, MT2C_Mm was downregulated in SRF LOF (2.7-fold), but this did not reach significance at a cutoff of 0.05 ($P\text{adj} = 0.059$) (Fig. 5B; Supplemental Fig. S5A; Supplemental Table S2). However, performing the same differential expression analysis at the subfamily level, we found that some MT2C_Mm subfamilies are significantly downregulated in late 2-cell stage SRF LOF embryos (Fig. 5C; Supplemental Table S3). In particular, MT2C_Mm subfamily vii was the most downregulated (Supplemental Table S3). This is in line with the fact that MT2C_Mm_vii is the oldest subfamily that contains the SRF binding site but does not contain the DUX binding motif. (Fig. 2E). The apparent discrepancy between the

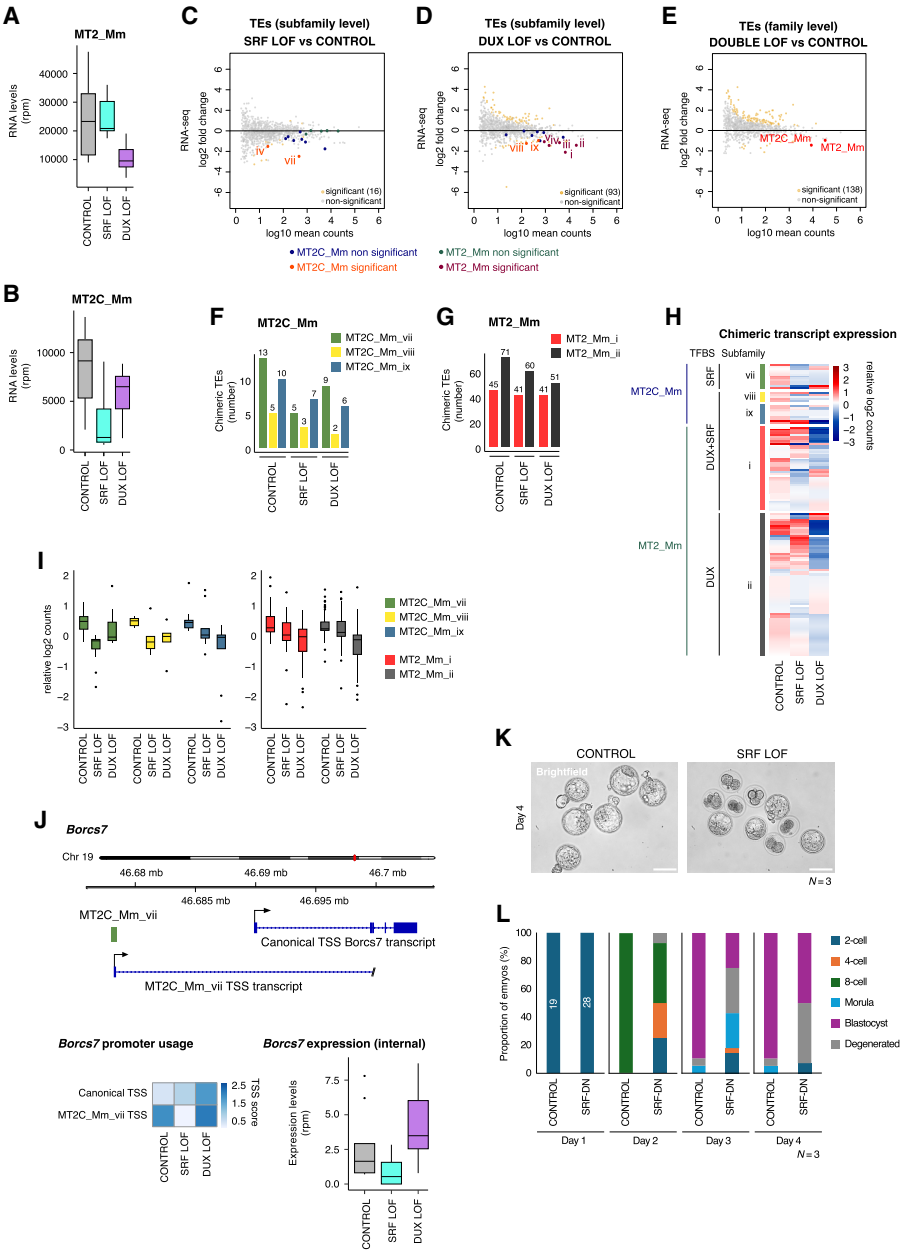

Figure 5. SRF regulates TE expression and contributes to host genome regulation in mouse embryos by driving expression of chimeric transcripts. (*A*,*B*) Expression of MT2_Mm (*A*) and MT2C_Mm (*B*) in control, SRF LOF, and DUX LOF embryos. Box plots show the median and interquartile ranges of MT2C_Mm- or MT2_Mm-derived RPM across embryos, and whiskers display the highest and lowest values within 1.5 times the interquartile range (IQR). Only complete MT2C_Mm and MT2_Mm insertions were considered. (*C*,*D*) MA plot comparing $\log_2$ fold change in SRF LOF and control (*C*) and DUX LOF and control (*D*) embryos against $\log_{10}$ RNA mean counts. Differentially expressed TEs (from subfamily DEseq object; see the Materials and Methods) are in orange (Padj < 0.05), and nondifferentially expressed TEs are in gray. Nonsignificant MT2_Mm and MT2C_Mm subfamilies are in green and blue, respectively. MT2_Mm and MT2C_Mm subfamilies significantly differentially expressed are labeled with the subfamily number in pink and orange, respectively. (*E*) MA plot comparing $\log_2$ fold change in double LOF and control embryos against $\log_{10}$ RNA mean counts. Differentially expressed TEs (from general DEseq object; see the Materials and Methods) are in orange (Padj < 0.05), and nondifferentially expressed TEs are in gray. Families of interest (MT2_Mm and MT2C_Mm; considering only complete insertions) significantly differentially expressed are in red. (*F*) The number of MT2C_Mm-derived chimeric TE genes found in control, SRF LOF, and DUX LOF, ordered and colored by subfamily (in MT2C_Mm_vii, MT2C_Mm_viii, and MT2C_Mm_ix). The number of chimeric TE genes is indicated *above* each bar of the bar plot. (*G*) The number of MT2_Mm-derived chimeric TE genes found in control, SRF LOF, and DUX LOF, ordered and colored by subfamily (in MT2_Mm_i and MT2_Mm_ii). The number of chimeric TE genes is indicated *above* each bar of the bar plot. (*H*) Heat map showing the expression of chimeric TE genes in control, SRF LOF, and DUX LOF embryos (chimeric TEs found in the control data set are represented). Values are $\log_2$-normalized counts for each condition centered on the row mean. Rows are organized by MT2_Mm or MT2C_Mm subfamilies of the TE-derived chimeric gene. Individual insertions were hierarchically clustered within all subfamilies, and the resulting five plots were combined. (*I*) Box plot showing the median and interquartile ranges of chimeric TE genes in control, SRF LOF, and DUX LOF embryos (chimeric TEs found in the control data set are represented) by subfamily. Values are $\log_2$-normalized counts for each condition centered on each chimeric TE mean (row mean from the heat map). Whiskers are the highest and lowest value within 1.5 times the IQR. (*J*) Promoter usage and expression of the *Borcs7* gene at the 2-cell stage. Browser snapshot of the genomic region of chromosome 19 containing the canonical gene transcript and the TE-derived (MT2C_Mm_vii) transcript. The promoter usage is indicated by the heat map of the TSS score of both the canonical and the TE-derived TSSs in control, SRF LOF, and DUX LOF embryos. The box plot shows the median RPM and the interquartile range of *Borcs7* transcripts from internal sequencing reads across embryos, and whiskers display the highest and lowest values within 1.5 times the interquartile range. (*K*) Bright-field representative images at day 4 after injection of control and SRF LOF (see the Materials and Methods). N is the number of biological replicates. Scale bars, 100 µm. (*L*) Developmental progression (in percentage) of control and SRF LOF embryos at days 1–4. The numbers in the bar plot at day 1 are the total number of embryos for each condition. N is the number of biological replicates.

effects of SRF on transcriptional regulation of TEs at a family level versus a subfamily level indicates heterogeneity in the dependence on a specific TF—in this case, SRF within a given MT2C_Mm subfamily. Alternatively, but not mutually exclusively, it highlights the need for a combinatorial action of several TFs that ensure robust gene

expression in vivo. This is in line with our findings indicating that the subfamilies of MT2C_Mm and MT2_Mm are related to the presence of specific TFBSs. Our data indicating that only some MT2C_Mm subfamilies are deregulated in the SRF LOF are consistent with the presence of the SRF motif alone versus SRF and DUX motifs. Loss of DUX, in contrast to SRF, led to a strong reduction in MT2_Mm levels to ~30% of the levels in control embryos (Fig. 5A; Supplemental Fig. S5B; Supplemental Table S2), in agreement with previous work (De Iaco et al. 2017, 2020; Guo et al. 2019, 2024). This is a downregulation of MT2_Mm levels similar to that described using a similar LOF approach, which led to undetectable levels of DUX (Guo et al. 2024). Although DUX loss led to a slight reduction of MT2C_Mm transcript levels (Fig. 5B), this effect was not significant (Padj = 0.43) (Supplemental Fig. S5B; Supplemental Table S2). Interestingly, analysis at the subfamily level indicated that depletion of DUX led to a significant downregulation of all MT2_Mm subfamilies (Fig. 5D; Supplemental Table S3), in agreement with the presence of the DUX binding site across all MT2_Mm subfamilies. Importantly, the two youngest subfamilies of MT2C_Mm, viii and ix, which uniquely possess the DUX binding site, were also significantly downregulated upon DUX depletion (Fig. 5D; Supplemental Table S3). These results are in line with our observations above, suggesting a complementary role for both SRF and DUX in regulating transcription of the MT2 classes of ERVL and a shift from an SRF dependence to a DUX dependence during MT2C_Mm-to-MT2_Mm evolution. In addition, they predict that the combined loss of DUX and SRF would result in downregulation of both MT2C_Mm and MT2_Mm. Accordingly, we found that the double LOF of SRF and DUX leads to a significant reduction in both MT2_Mm and MT2C_Mm transcript levels (Fig. 5E). We conclude that SRF regulates the expression of a distinctive set of TEs in mouse embryos. Thus, our data indicate that both SRF and DUX regulate MT2 expression collaboratively at the 2-cell stage.

Having established that SRF regulates MT2C_Mm subfamily expression in mouse embryos, we then asked whether MT2C_Mm expression is relevant for early embryo biology. Specifically, we investigated whether MT2C_Mm has a potential regulatory role in 2-cell stage embryos. We reasoned that the identification of chimeric TE host transcripts would provide direct evidence of regulatory TE function, as this would indicate TE-driven host gene transcription. For this, we leveraged our single-embryo RNA-seq protocol, which enables the capture of 5′ transcript ends and can therefore assign the TSSs of individual transcripts. This allows the identification of chimeric transcripts that initiate within a TE and continue into host genes. Importantly, we found chimeric transcripts initiating at all MT2C_Mm subfamilies in control embryos (n = 64) (Fig. 5F; Supplemental Fig. S5C; Supplemental Table S4). This indicates that MT2C_Mm contributes to the regulation of host genes in mouse embryos. Interestingly, we also identified a total of 134 chimeric transcripts initiating at MT2_Mm, which likewise derive from all subfamilies (Fig. 5G; Supplemental Fig. S5D; Supplemental Table S4). SRF LOF led to a reduction in the number of chimeric transcripts initiating at MT2C_Mm (from 64 to 34) and affected all subfamilies, albeit to different extents (Fig. 5F; Supplemental Fig. S5C). We next focused on those chimeric TE transcripts that are driven by the youngest MT2C_Mm families and the oldest MT2_Mm families according to whether they contain an SRF motif only, SRF and DUX motifs, or DUX only, thereby representing the evolutionary transition between the acquisition of SRF and DUX TFBSs (Fig. 5H). SRF LOF led to a reduction of chimeric gene expression of all these MT2C_Mm subfamilies (subfamilies vii, viii, and ix) (Fig. 5H). Notably, among the 64 genes that form chimeric transcript with MT2C_Mm, eight are significantly downregulated upon SRF LOF, but none of the 134 genes that form chimeric transcripts with MT2_Mm is downregulated upon SRF LOF (Fig. 5H). Although the number of chimeric transcripts is low, thereby precluding strong statistical conclusions, these data may suggest that the chimeric transcripts affected by SRF LOF are primarily driven by MT2C_Mm. In contrast, depletion of DUX affected the expression levels of most MT2_Mm chimeric transcripts (Fig. 5H). Interestingly, for the oldest MT2_Mm subfamily (i), which still contains a high-confidence site for SRF (Fig. 1E), we observed that while roughly half of the transcripts are dependent on SRF, the other half are dependent on DUX (Fig. 5H). This was also evident when we computed the expression levels of all the chimeric transcripts associated with these three MT2C_Mm subfamilies (vii, viii, and ix) and two MT2_Mm subfamilies (i and ii) together (Fig. 5I). Although we cannot formally rule out a nonspecific effect of our dominant negative approach, these results are in line with our phylogenetic analysis pinpointing the emergence of the DUX binding site and suggest that the appearance of the DUX motif led to a dominating role of DUX over SRF in regulating those MT2 insertions containing TFBSs for both factors.

The above data prompted us to analyze whether the promoter usage of chimeric transcripts in embryos is dependent on the presence of SRF and DUX sites. We found that genes that form chimeric transcripts with MT2C_Mm in control embryos rely substantially on the usage of MT2C_Mm for transcription initiation at the 2-cell stage; for example, *Borcs7* (Fig. 5J, control). Interestingly, SRF LOF completely switches off this TSS and leads to overall reduced transcript levels of *Borsc7*, which we quantified based on the Smart-seq internal fragments (Fig. 5J). In general, SRF LOF either reduces the TSS usage at MT2C_Mm or leads to a switch in TSS usage to the canonical gene TSS, leading to an overall downregulation of gene expression. We depict two examples of this in Supplemental Figure S5, E and F. These observations may suggest that MT2C_Mm, through SRF function, could provide robustness to gene expression in embryos. Likewise, we found genes that rely primarily on MT2_Mm to initiate transcription in control embryos and where loss of DUX results in the reduced usage of MT2_Mm as a promoter, ultimately leading to decreased gene expression (Supplemental Fig. S5G). We conclude that both DUX and SRF contribute to host genome regulation in mouse embryos at least in part through driving the

expression of chimeric transcripts. We propose that MT2C_Mm and MT2_Mm provide a platform for TSS usage, which is regulated by DUX and SRF and determined by the phylogeny and the amplification of their LTR sequences during evolution.

Finally, we asked whether SRF function may be required for early developmental progression. For this, we monitored the development of SRF LOF until the blastocyst stage. We found that although most control embryos (17 out of 19) reach the blastocyst stage, SRF LOF embryos do so in a lower proportion, with only about half reaching the blastocyst stage (14 out of 28) (Fig. 5K–L; Supplemental Fig. S5H). All SRF LOF embryos reached the 2-cell stage in a timely manner, but the proportion of embryos developing beyond the 2-cell stage decreased compared with controls (Fig. 5K–L; Supplemental Fig. S5H; Supplemental Table S5). Importantly, microinjection of the same amount of mRNA for GFP as the SRF DN mRNA did not affect developmental progression (Supplemental Fig. S5I), thereby ruling out that the developmental phenotype of the loss-of-function SRF that we report is due to a toxic effect linked to the concentration of the mRNA that we used. Among the genes misregulated upon SRF LOF, Cyclin 1 and Cdkn1a, both required for cell cycle progression, are downregulated (Supplemental Table S1), in line with the known literature of SRF's role in proliferation. Although the effect of SRF on developmental progression may be attributed to multiple genes and/or repeats, our data indicate that SRF is a crucial regulator of gene and TE expression in mouse embryos at the time of ZGA.

Discussion

Evolutionary approaches can reveal the molecular mechanisms for potential co-option of TEs by the host and for regulatory strategies involved in this process. MT2_Mm has been extensively studied as a TE that can drive expression of host ZGA genes in mouse embryos and activate a ZGA program in 2CLCs (Evsikov et al. 2004; Peaston et al. 2004; Macfarlan et al. 2012). Our work indicates that MT2C_Mm is an ancestral *Mus* LTR to MT2_Mm that can drive expression of host genes in mouse embryos, contributing to establishing cell-specific transcriptional programs at the beginning of development. We also demonstrate that MT2_Mm and MT2C_Mm function as alternative promoters and generate chimeric transcripts. Our analysis of TSS usage suggests that both LTRs are used at the time of ZGA and therefore could work as modules to provide temporal coordination for gene expression control. This provides support for the model in which TEs, through the distribution of *cis*-regulatory elements across the genome, drive the evolution of developmental transcription programs.

We found that the expression of individual insertions does not correlate with their evolutionary age, and thus relatively older insertions are transcribed at levels similar to more recent ones. Of note, recent work has documented that, indeed, old TEs are robustly expressed in early mammalian embryos despite having lost structural integrity (Oomen et al. 2025). Our phylogenetic analysis has revealed the sequential acquisition of transcription factor binding sites in the MT2 family over evolutionary time, including a striking 9 bp deletion in MT2C_Mm prior to the emergence of MT2_Mm, leading to the acquisition of the DUX binding site. Our data indicate that only some MT2C_Mm subfamilies are deregulated in the SRF LOF, in agreement with the presence of SRF motifs. However, there are likely additional TFs contributing to the fine-tuned regulation of these TEs, providing additional granularity to their regulation in vivo. This is comparable with the known combinatorial, and cooperativity of, transcription factors regulating expression of genes across cell types. It is interesting that genes regulated by SRF and those regulated by DUX in 2-cell embryos are largely nonoverlapping. This is in line with the concept that instead of one master TF regulator, several factors contribute to the regulatory network underlying ZGA, potentially allowing for robustness and fine-tuning of transcriptional programs. SRF therefore adds to the other factors so far identified in this process, including, for example, OBOX (Ji et al. 2023; Guo et al. 2024). The existence of additional TFs and or regulatory TEs is also in agreement with the indication that DUX depletion leads to a defect in ZGA but not its full abrogation (Chen and Zhang 2019; De Iaco et al. 2020). Thus, we view the role of SRF as an additional contribution to transcription in early embryos, a process that is likely to be regulated by multiple transcription factors, some of which are redundant. Although we cannot rule out nonspecific effects of our dominant negative approach, which would need to be validated with additional methods, our combined data in vitro are consistent with SRF regulating transcription of MERVL. Further genetic approaches to deplete SRF at a defined temporal window in 2-cell embryos will be necessary to further shed light on the role of SRF in embryos.

Last, we note that only between ~3% (MT2C_Mm) and ~6% (MT2_Mm) of all the complete LTR insertions in the genome form detectable chimeric transcripts. Thus, although alternative promoter usage demonstrates regulatory functions of these TEs at these stages, the co-option strategies may not be limited to this mode of action. It would be important to investigate whether MT2 family members play additional roles; for example, through chromatin or epigenome remodeling.

Overall, our work showcases an outstanding example of phyloregulatory adaptation during mammalian embryogenesis and suggests that waves of acquisition of TFBSs have contributed to waves of amplification of TEs within our current genomes.

Materials and methods

Phylogenetic analyses

For the LTR selection and filtering, LTRs (MT2_Mm and MT2C_Mm) and internal regions (MERVL-int) were extracted from the RepeatMasker annotation for the mouse genome (mm10; RepeatMasker open-4.0.5; repeat library 20140131; http://www.repeatmasker.org). We used

OneCodeToFindThemAll.pl (Bailly-Bechet et al. 2014) and rename_mergedLTRelements.pl (Thomas et al. 2018) for MT2_Mm and MERVL-int to identify LTRs and internal sequences belonging to the same elements and assign them as 5′, 3′, or solo LTRs. The LTR size distributions were visualized by density plot using the ggplot2 (version 3.5.1) library in R (Wickham 2016). All MT2_Mm (solo, 5′, and 3′) and MT2C_Mm with a length between 400 and 586 bp and 385 and 565 bp, respectively, were used for further analysis. The size-selected sequences were retrieved from mm10 using the getfasta function from the Bedtools package (v2.31.1) (Quinlan and Hall 2010) and aligned with MUSCLE (version 3.8.1551) (Sievers et al. 2011) with default parameters. The alignment was trimmed with TrimAl (version 1.4. rev15) (Capella-Gutiérrez et al. 2009) using the option -tg 0.01 to remove columns where >99% of the sequences had a gap. A maximum-likelihood phylogeny was generated using IQ-Tree (version 2.1.4-beta) (Minh et al. 2020) with the options --seed 42 -T AUTO -m MFP -B 6000 --ancestral --sup-min 0.95. The consensus tree files (.contree) output from IQ-Tree were visualized using iTOL (Letunic and Bork 2024). Subfamilies were defined as clusters of a minimum of 30 sequences, supported by a node with UFBootstrap >0.95, with branch length >0.015. These thresholds were defined empirically based on previous work (Carter et al. 2022) and on visual examination of the trees generated by IQ-Tree. Therefore, the criteria are semiarbitrary, as there is no standardized manner to perform these analyses across TE families. For example, changing the branch length threshold to 0.006 or the minimum number of insertions to 10 led to several additional subfamilies with only 10–12 insertions and without a clear branching, which are unlikely to be informative. The parameters chosen for Figures 1A and 2A resulted in a clear distinction of visible clades in both MT2_Mm and MT2C_Mm. Insertions that did not qualify as subfamilies because there were too few per group but were separated from the subfamilies with a long enough genetic distance were classified as outgroup and are labeled in gray in the figures.

Divergence analysis

Consensus sequences for MT2A, MT2B, MT2C_Mm, and MT2_Mm were recovered from Dfam (Storer et al. 2021). These consensus sequences were aligned and trimmed, and a phylogenetic tree was reconstructed as described above. The obtained consensus tree file (.contree) was visualized in rectangular mode using iTOL (Letunic and Bork 2024). Based on this phylogenetic tree, the Dfam consensus sequences of the closest ancestor were used to root the phylogenetic trees and establish the genetic distance of each insertion to that root. MT2C_Mm consensus sequence was used to root the MT2_Mm tree. MT2B consensus sequence was used to root the MT2C_Mm tree and a tree containing both MT2_Mm and MT2C_Mm insertions. The consensus sequence was added to the fasta file containing the single insertions, which were aligned, trimmed, and used to reconstruct a phylogenetic tree as described above. The obtained tree files (.treefile) from IQ-Tree were parsed, and the consensus sequence was assigned as root using the phylo package from the biopython module (version 1.79) (Cock et al. 2009). Genetic distances of each insertion to the root were calculated using the distance function available in the phylo package. The distances were visualized using the geom_jitter function from ggplot2 in R.

Molecular clock age analysis

The RepeatMasker annotation file for mm10 (rmsk.txt) was obtained from the UCSC genome browser (https://hgdownload.cse.ucsc.edu/goldenPath/mm10/database) and modified to contain the subfamily annotation information by matching genome coordinates for each insertion. Not all insertions were matched from the phylogenetic analysis, leading to slight insertion number differences for each subfamily. The age of each insertion was computed by extracting the millidivergence and converting it to a *p*-distance. The *p*-distance values were then transformed to genetic distances using the Jukes–Cantor nucleotide substitution model (Jukes and Cantor 1969). The distances were converted to age by multiplying by the mouse substitution rate per million years (1 nt change per 4.5 million years ago). The age was visualized using the geom_jitter function from the ggplot2 package in R.

Transcription factor binding site mapping

The primary and/or secondary position weight matrices (PWMs) were downloaded from UniPROBE (Newburger and Bulyk 2009; Hume et al. 2015) for the selected TFs, except for DUX, whose PWM was based on the DUX binding site found to bind MT2_Mm by Hendrickson et al. (2017). These matrices were used to scan MT2_Mm or MT2C_Mm single insertions using the matrix-scan command line from RSAT (version 2020.02.29) (Santana-Garcia et al. 2022) with the following options: -pseudo 1 -decimals 1 -2str -origin start -bgfile 2nt_upstream-noorf_Mus_musculus_GRCm38-noov-1str.freq -bg_pseudo 0.01 -return limits -return sites -return pval -return rank -lth score 1 -uth pval 1×10^{-4}. The obtained matrix was converted to a binary file, and results were displayed using the heat map function from ggplot2 in R (version 4.2.3). The consensus sequences of the MT2_Mm and MT2C_Mm subfamilies were scanned for TFBS presence as described above.

Generation of consensus sequences and median-joining network analysis

For each new subfamily, the consensus sequence was established using the majority rule for each nucleotide position using the seqinr package (version 4.2.36) in R (Charif and Lobry 2007). The resulting consensus sequences were aligned with MUSCLE, and alignments were visualized with Jalview (version 2.11.3.3) (Waterhouse et al. 2009). Median-joining network analysis (Bandelt et al. 1999) was reconstructed using Popart (Leigh and Bryant 2015).

Plasmid construction and in vitro transcription

The pCR8/GW/TOPO-SRF (Addgene 98618) was purchased from Addgene, and the cDNA was cloned in a pCMV vector with a MYC tag in the N terminus. The DUX cDNA was a gift from Didier Trono (De Iaco et al. 2017) and was cloned into the same pCMV vector. The GABPA and GABPB1 CDSs were amplified from mouse ES cell cDNA and cloned to the same pCMV vector. The SRF dominant negative mutant of SRF was generated by Q5 site-directed mutagenesis (NEB E0554S) in pCVM-MYC. sfGFP was cloned in-frame into the 3′ end of the SRF dominant negative. The final insert (referred to as "SRF-DN") was further cloned to the pRN3p-HA construct. For renilla luciferase, we used the pCDH-E1Fa-Ren-T2A-mCherry vector (Addgene 104833). The scrambled LTR was designed using the Random DNA Sequence Generator (https://users-birc.au.dk/palle/php/fabox/random_sequence_generator.php) with the following criteria: size similar to that of an LTR (500 bp) and GC content similar to that of the mouse genome (42%); minimal TFBSs compared with MT2_Mm was controlled using RSAT. The sequence was synthesized by Eurofins and amplified using primers introducing KpnI and XhoI restrictions sites on the 5′ and 3′ ends, respectively. The consensus sequences (for MT2_Mm_i, MT2_Mm_ii, MT2C_Mm_i, MT2C_Mm_vii, and MT2C_Mm_ix) were synthesized by Eurofins in pEX-A128 with KpnI and XhoI restriction sites in the 5′ and 3′ ends, respectively. All sequences were subsequently cloned to the firefly luciferase-containing vector pGL2-basic. The mRuby plasmid without promoter was described previously (Oomen et al. 2025). The consensus sequences for MT2_Mm_ii and MT2C_Mm_vii were cloned upstream of the mRuby. Sanger sequencing was used to verify all plasmids. A NucleoBond Xtra Midi kit (MN 740410.50) was used to isolate DNA before transfection or in vitro transcription. mRNAs were transcribed in vitro using the T3 mMessage mMachine transcription kit (Ambion AM1348).

Luciferase reporter assay

HEK293 (human embryonic kidney) Tet-on cells (Clontech 631182) were cultured in Dulbecco's modified Eagles's medium (DMEM; Gibco 41966-029) complemented with 10% Tet system approved FBS (Takara 631106) and 1% penicillin–streptomycin–glutamine (Gibco 103780 16) at 37°C and 5% CO_2. Twenty-four hours before transfection, 2×10^5 cells were seeded in 6 well plates. Cells were transfected with 1.5 µg of the firefly luciferase plasmid, 20 ng of the renilla luciferase plasmid, and 5–500 ng of the pCMV-TF plasmid for the corresponding transcription factor using the amount indicated in the figures. The pCMV-TF plasmid with no insert (pCMV-empty) was used to adjust the levels of DNA transfected to 2 µg in all conditions except for GABPA and GABPB1, where it was adjusted to 2.5 µg. Transfections were performed using JetPrime (PolyPlus 101000015) using a 1:1 ratio (DNA:JetPrime). The medium was replaced on the day following the transfections, and cells were lysed 48 h after transfection. Luciferase activity was measured using the dual-luciferase reporter assay kit (Promega E1980) following the manufacturer's instructions. The ratios of luciferase to renilla were computed for all experiments, and fold change over control for each replicate was calculated. The fold change values were $\log_2$ transformed and plotted in R using ggplot2 (version 3.5.1). For statistical analysis, a linear model was fitted to the data excluding the intercept to evaluate group differences (R version 4.1.2). All MT2 luciferase assays performed were used together to fit the model for SRF and DUX separately. All control luciferase assays performed were used together to fit the model for SRF and DUX separately. Preselected hypotheses were tested and corrected for multiple comparisons using the glht function from the Multcomp package (version 1.4-25). To determine significance, an adjusted *P*-value threshold set to 0.05 was used.

Western blot analysis

HEK293 Tet-on cells were cultured and transfected as above, except with 500 ng of pCMV-TF vector. Cell lysates containing proteins were recovered as performed for luciferase assay (using dual-luciferase reporter assay kit lysis buffer). Proteins were separated on 12% polyacrylamide gel, which was subsequently transferred to 0.2 µm PVDF membranes (Cytiva 10600022) previously activated in methanol. The membranes were blocked in 3% BSA in TBST for 1 h at room temperature. Membranes were then incubated with primary antibodies in blocking solution overnight at 4°C on a nutator. The antibodies used were anti-Myc tag (1/10,000 dilution; Cell Signaling Technology 2276) and anti-histone H3 (dilution 1/100,000; Abcam ab1791). Three 5 min washes with TBST were followed by 1 h incubation with HRP-conjugated secondary antibodies in blocking solution for 1 h at room temperature. The secondary antibodies used were antirabbit (dilution 1/20,000; Thermo Fisher A16110) and antimouse (dilution 1/20,000; Thermo Fisher A16078). After three 5 min washes in TBST, membranes were visualized using SuperSignal West Pico Plus chemiluminescent substrate (Thermo Fisher 34580) with a Chemi-Doc touch imaging system (Bio-Rad).

Embryo collection, manipulation, and culture

All animal experiments were performed in compliance with regulations from the government of Upper Bavaria. For immunostainings, CD1 females (6–10 weeks old) were mated with CD1 males (2–8 months old). Zygotes and early 2-cell, late 2-cell, 4-cell, and 8-cell embryos were collected at ~16, ~32–33, ~41–42, ~48, and ~56–57 h postcoitum, respectively. For microinjections, embryos were collected from F1 (C57BL/6J × CBA/H) females (<10 weeks of age) mated with F1 males (3–6 months old). Ovulation was induced by injection of pregnant mare serum gonadrotropin (PMSG; Ceva) followed 48 h later by human chorionic gonadotropin (hCG; MSD Animal Health). For RNA-seq, zygotes were collected between 24 and 25 h after hCG from the oviducts of the females, and cumulus cells were removed by brief incubation in M2 containing hyaluronidase

(Sigma-Aldrich). Zygotes were microinjected with 150 ng/µL dsRed mRNA with, in addition, 20 µM either scramble ASO or anti-Dux ASO (Guo et al. 2024) with or without 500 ng/µL SRF-DN mRNA. Embryos were cultured in K-modified simplex optimized (KSOM) drops covered with paraffin oil (Sigma) at 37°C and 5% CO_2 until collection at 48 h after hCG for single-embryo RNA-seq. Nine control, nine DUX LOF, seven SRF LOF, and 10 double LOF 2-cell stage embryos from at least three independent experiments were collected for Smart-seq+5′, and one embryo from the DUX LOF group was removed after quality control analysis. For development experiments, SRF LOF and control embryos were collected between 24 and 25 h after hCG; injected with 500 ng/µL SRF-DN mRNA and 250 ng/µL GFP mRNA, respectively; cultured; and scored on days 1, 2, 3, and 4 for developmental progression. To control for potential nonspecific or toxic effects due to mRNA concentration, zygotes were microinjected with a mixture of 500 ng/µL GFP mRNA combined and 150 ng/µL dsRed mRNA as described above and scored for developmental progression. Developmental progression was plotted with either Excel or GraphPad Prism (version 10.4.0). For reporter assay in embryos, zygotes were collected 18–19 h after hCG and injected with 40 ng/µL either MT2_Mm_ii or MT2C_vii or no promoter reporter plasmids. The embryos were cultured until 48 h after hCG, when the Ruby signal was observed with an epifluorescent microscope and the number of positive embryos was counted.

Immunostainings

Immunostainings were performed as described previously (Torres-Padilla et al. 2006). Briefly, the zona pellucida was removed with acid tyrode (Sigma-Aldrich). Embryos were then washed in PBS and fixed in 4% PFA, 0.04% Triton, 0.3% Tween-20, and 0.2% sucrose in a glass-bottom dish for 20 min at 37°C. Embryos were subsequently washed three times in PBS and permeabilized with 0.5% Triton X-100 for 20 min at room temperature. Embryos were then washed three times in PBST (0.1% Tween 20 in PBS), incubated for 10 min in 2.6 mg/mL ammonium chloride in PBS solution, washed twice again in PBST, and blocked for 4–5 h in 3% BSA in PBST at 4°C. Embryos were then incubated in primary antibody in 3% BSA overnight at 4°C. The antibody used was anti-SRF (1/500 dilution; Abcam ab252868). Embryos were subsequently washed three times in PBST, briefly blocked again for 20 min in 3% BSA, and incubated for 3–4 h at room temperature in 3% BSA in PBST with secondary antibody labeled with Alexa-546 fluorophore (goat antirat; 1/500 dilution; Invitrogen A11081). Finally, embryos were washed twice in PBST and once in PBS for 20 min and then mounted in VectaShield containing DAPI (Vector Laboratories H-2000). Confocal microscopy was done using a 63× oil objective on a Leica SP8 microscope (Leica).

RNA-seq (SMART-seq+5′)

SMART-seq+5′ was modified from the Smart-seq2 protocol (Picelli et al. 2013, 2014) as described previously (Oomen et al. 2025). All samples were collected in the same lysis buffer, which was stored at −80°C until use (Clontech 10× lysis buffer [635015] diluted to 1× in H_2O supplemented with ERCC RNA spike-ins diluted to 1:581,000 and aliquoted in PCR tubes at 5.8 µL/tube). Embryos were washed three times in PBS, transferred to tubes containing lysis buffer, snap-frozen in liquid nitrogen, and kept at −80°C until further processing. RNA was extracted using AMPure RNA magnetic beads (Beckman Coulter A63987), washed with 80% ethanol, and resuspended in 1 µL of dNTP mix (Thermo Fisher R0192), 1 µL of 10 µM oligo-dT30V (5′-AAGCAGTGGTATCAACGCAGAG TACT30V-3′; Sigma), and 1 µL of nuclease-free H_2O containing 5% RNase inhibitor (Takara 2313A). The samples were first incubated for 3 min at 72°C and kept on ice until further processing. In the meantime, the reverse transcription solution was prepared: 2 µL of SuperScript II 5× RT buffer (Thermo Fisher 18064014), 1.6 µL of 40% PEG-8000 (Sigma P1458), 0.5 µL of DTT, 0.25 µL of RNase inhibitors (Takara 2313A), 0.1 µL of 100 µM TSO (5′-AAG CAGTGGTATCAACGCAGAGTACATrGrG+G-3′; TIB MolBiol), 0.06 µL of 1 M $MgCl_2$ (Sigma M1028), 2 µL of 5 M betaine (Sigma B0300-1VL), and 0.5 µL of SuperScript II RT. Seven microliters of the reverse transcription mix was added to the 3 µL of annealed RNA mix and incubated for 90 min at 42°C followed by 15 min at 70°C. Preamplification of the obtained cDNA was performed using KAPA HiFi readymix (KM2605) for 14 cycles with 10 µM ISPCR primers (5′-AAGCAGTGGTATCAACG-CAGAGT-3′; Sigma) and purified using Agencourt Ampure XP beads (Beckman Coulter A63881). For each sample, 2.5 µL of 120 µg/µL cDNA was used for tagmentation, which was performed with the Nextera XT kit (Illumina 15032354). The preamplified cDNA was mixed with 5 µL of tagment DNA buffer and 2.5 µL of Amplicon tagment mix and incubated for 5 min at 55°C. The tagmentation reaction was stopped with 2.5 µL NT buffer, and samples were incubated for 5 min at room temperature. Tagmented DNA was then amplified for 12 cycles using the two standard i5 and i7 Nextera unique double indexes together with a tailed i7 index containing an overhang, enabling the capture of the 5′ of the transcripts (Oomen et al. 2025). The libraries were verified using an Agilent 5200 fragment analyzer system (Agilent). A 150 bp paired-end sequencing protocol was used on an Illumina NovaSeq 6000 platform.

Mapping and processing of Smart-seq+5′

Smart-seq+5′ libraries were processed as described previously (Oomen et al. 2025). In brief, sequence quality was verified using FastQC. Trimmomatic (Bolger et al. 2014) configured in paired-end (PE) mode was used to remove adaptor and low-quality sequences. A custom Python script, available on GitHub, was used to sort between 5′ transcript ends and internal transcript fragments according to their adaptor sequences. Unless stated otherwise, all analyses were performed with the 5′ reads. The reads were mapped to GRCm38 using STAR (v2.7.11a). BAM files were modified to keep only the read2 using a custom

Perl script available on GitHub and the SAMtools package (Li et al. 2009). TEcount or TElocal from the TEtranscript toolkit (Jin et al. 2015) was used to count TEreads. The RepeatMasker annotation file for mm10 (mm10_rmsk.gtf) was modified to include the information of the new identified phylogenetic subfamilies and used to count TE reads. Insertion coordinates corresponding to specific subfamilies were identified using bedfiles generated during the phylogenetic analysis. These coordinates were then used to modify the annotation file by modifying the "gene_id" field for TEcount analysis and the "transcript_id" field for TElocal analysis. Not all insertions were matched from the phylogenetic analysis, leading to slight insertion number differences for each subfamily. Expression analysis across development stages was performed using the BAM files from Oomen et al. (2025) containing the 5′ reads using the previously mentioned modified TE annotation file as TE annotation. The poly(A)-based nature of Smart-seq+5′ captures efficiently the ERVLs studied here and has been experimentally assessed (Oomen et al. 2025). Plots were made using ggplot2 in R. To analyze single insertion expression, the fastq files of the 2-cell stage embryos from Oomen et al. (2025) were processed as described above, but TElocal from the TEtranscript toolkit (Jin et al. 2015) was used to count reads from single TE insertions. For this, we created an index file by processing the modified annotation file containing the phylogenetic subfamilies with the TElocal_indexer script (Jin et al. 2015). These modified index files were then used to quantify single insertion expression using TElocal. The data were visualized using jitter dot plot from the ggplot2 package in R. Single-insertion TElocal values were combined in a table with genetic distances obtained in the divergence analysis, and expression against age was plotted using ggplot2 in R.

Differential expression analysis

The sample count tables generated by TEcount from the 5′ or internal read data were merged into a single table using a bash script and then loaded in R. Reads per million (RPM) were calculated for each sample. The following quality thresholds were applied: a minimum of 5×10^5 reads and a minimum of 1000 detected genes; the maximum percentages of reads assigned to mitochondrial DNA and ERCC spike-ins were set to 10% and 15%, respectively. Expression levels of the TFs SRF and DUX were also assessed, and one DUX LOF sample was removed due to higher *Dux* expression (compared with controls), indicating that downregulation of *Dux* had not worked in this embryo. Differential expression analysis was performed using DESeq2 (v1.38.3) (Love et al. 2014) with read counts per gene and TEs calculated by TEcount, taking as input the alignments performed with STAR. We performed two DESeq objects. For the first one, which we refer to as general DESeq object in our pipeline, we used the TE read counts at the family level, for which we compiled all the subfamilies that we identified in our phylogeny as one single family comprising all the insertions considered used for the phylogeny. In this annotation, the label "MT2_Mm:ERVL:LTR:OTHERS" (or MT2C_Mm:OTHERS) comprises all the (fragmented) sequences that were not included in the size selection. In the second DESeq object, we used TE read counts for individual subfamilies as determined in our phylogenetic analyses (subfamily DESeq object). Differential expression analysis was presented using MA plots showing $\log_2$ fold change between each LOF experiment and the control experiment. For gene MA plots, the scattermore package (v1.2) was used for plotting. Comparisons of the DUX and SRF DE genes were visualized using scatter plots (using ggplot2). For significance, adjusted *P*-value (*P*adj) threshold was set to 0.05 (*P*-values for genes and TEs were obtained from the general DESeq object, and *P*-values for subfamilies were obtained from the subfamily DESeq object). Genes significantly upregulated or downregulated in SRF LOF were assigned to maternal RNAs, minor ZGA, major ZGA, or other using the Database of Transcriptome in Mouse Early Embryos (DBTMEE) (Park et al. 2015). Changes between individual embryos were visualized in heat maps displaying the $\log_2$ of normalized counts centered on the row mean using pheatmap (v1.0.12) in R, with hierarchical clustering applied to rows for each group of genes. For embryonic PCA, embryos from the different embryonic stages from Oomen et al. (2025) and SRF LOF embryo, DUX LOF embryo, and control embryo counts were $\log_2$ transformed to generate the PCA. The 100 genes contributing the most to PC1 were extracted from the PCA and called the "top 100 PC1 genes."

Identification of TE-initiated gene transcripts

Chimeric TE–gene interactions were identified with ChimeraTE (Oliveira et al. 2023) using mode-1 and only keeping TE-initiated chimeric transcripts. First, mapped quality-passed 5′ fragment read pairs were converted back to fastq format using SAMtools fastq. Converted fastq files were used as input for ChimeraTE with the parameters –strand rf-stranded and --window 150000. For quantification, we used the modified TE annotation containing the MT2C_Mm and MT2_Mm subfamilies and the GRCm38 gene annotation. Only TE-initiated chimeric transcripts present in two or more replicates per experimental condition were used in downstream analysis and visualization. Using chimeric transcripts present in control embryos, we plotted heat maps of the $\log_2$ of normalized counts in each condition, centered on the row mean using pheatmap (v1.0.12) in R with hierarchical clustering applied to rows for each MT2_Mm or MT2C_Mm subfamily. For data representation, the combined relative $\log_2$ counts of all chimeric reads within each subfamily and for each condition were plotted as box plots using ggplot2 in R. Next, we identified transcript isoforms de novo with Bambu (Chen et al. 2023) using the junction files output by ChimeraTE as input. Last, we visualized promoter usage of canonical or chimeric promoters using Proactive (Demircioğlu et al. 2019).

Data availability

The RNA-seq data from this study are available from the Gene Expression Omnibus (GEO) database under accession number GSE271983. All other data supporting the findings of this study are available on reasonable request. Previous RNA-seq data sets reanalyzed here are available under accession code GSE225056.

Competing interest statement

M.-E.T.-P. is a member of the Advisory Ethics Panel of Merck. All of the other authors declare no competing interests.

Acknowledgments

We thank I. de la Rosa Velazquez and the Genomics Facility at Helmholtz Munich for sequencing, T. Schauer for support in the RNA-seq and statistical analysis, and J. Mo for providing the scramble LTR sequence. We thank T. Schauer and A. Burton for critical reading of the manuscript, C. Feschotte and T.A. Carter for their advice on the phylogenetic analysis, A. Hart-Fernandez for technical help, and all members of the Torres-Padilla laboratory for helpful suggestions. M.-E.T.-P. acknowledges funding from the Helmholtz Association and the German Research Foundation (DFG) Project-ID 213249687 (SFB 1064). M.E.O is a recipient of an EMBO long-term fellowship (ALTF 91-2021), and C.M.M.-D. is a graduate of the Genomics Programme of the National University of Mexico.

Author contributions: C.H. performed all experimental work. C.M.M.-D., M.E.O., and L.A.-P. performed computational work. M.P. and T.N. performed embryo experimental work. M.-E.T.-P. supervised and designed the study.

References

Affolter M, Montagne J, Walldorf U, Groppe J, Kloter U, LaRosa M, Gehring WJ. 1994. The *Drosophila* SRF homolog is expressed in a subset of tracheal cells and maps within a genomic region required for tracheal development. *Development* **120:** 743–753. doi:10.1242/dev.120.4.743

Aoki F, Worrad DM, Schultz RM. 1997. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. *Dev Biol* **181:** 296–307. doi:10.1006/dbio.1996.8466

Arsenian S, Weinhold B, Oelgeschläger M, Rüther U, Nordheim A. 1998. Serum response factor is essential for mesoderm formation during mouse embryogenesis. *EMBO J* **17:** 6289–6299. doi:10.1093/emboj/17.21.6289

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. *Nucleic Acids Res* **43:** W39–W49. doi:10.1093/nar/gkv416

Bailly-Bechet M, Haudry A, Lerat E. 2014. "One code to find them all": a Perl tool to conveniently parse RepeatMasker output files. *Mob DNA* **5:** 13. doi:10.1186/1759-8753-5-13

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16:** 37–48. doi:10.1093/oxfordjournals.molbev.a026036

Belaguli NS, Zhou W, Trinh T-HT, Majesky MW, Schwartz RJ. 1999. Dominant negative murine serum response factor: alternative splicing within the activation domain inhibits transactivation of serum response factor binding targets. *Mol Cell Biol* **19:** 4582–4591. doi:10.1128/MCB.19.7.4582

Bénit L, Lallemand JB, Casella JF, Philippe H, Heidmann T. 1999. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *J Virol* **73:** 3301–3308. doi:10.1128/JVI.73.4.3301-3308.1999

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:** 2114–2120. doi:10.1093/bioinformatics/btu170

Bosnakovski D, Gearhart MD, Ho Choi S, Kyba M. 2021. Dux facilitates post-implantation development, but is not essential for zygotic genome activation†. *Biol Reprod* **104:** 83–93. doi:10.1093/biolre/ioaa179

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew J-L, Ruan Y, Wei C-L, Ng HH, et al. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18:** 1752–1762. doi:10.1101/gr.080663.108

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25:** 1972–1973. doi:10.1093/bioinformatics/btp348

Carter TA, Singh M, Dumbović G, Chobirko JD, Rinn JL, Feschotte C. 2022. Mosaic *cis*-regulatory evolution drives transcriptional partitioning of HERVH endogenous retrovirus in the human embryo. *eLife* **11:** e76257. doi:10.7554/eLife.76257

Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, Duc J, Jang SM, Turelli P, Trono D. 2014. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev* **28:** 1397–1409. doi:10.1101/gad.241661.114

Chai J, Tarnawski AS. 2002. Serum response factor: discovery, biochemistry, biological roles and implications for tissue injury healing. *J Physiol Pharmacol* **53:** 147–157.

Charif D, Lobry JR. 2007. Seqinr 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution: molecules, networks, populations* (ed. Bastolla U, et al.), pp. 207–232. Springer, Berlin. doi:10.1007/978-3-540-35306-5_10

Chen Z, Zhang Y. 2019. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat Genet* **51:** 947–951. doi:10.1038/s41588-019-0418-7

Chen Y, Sim A, Wan YK, Yeo K, Lee JJX, Ling MH, Love MI, Göke J. 2023. Context-aware transcript quantification from long-read RNA-seq data with Bambu. *Nat Methods* **20:** 1187–1195. doi:10.1038/s41592-023-01908-w

Choi YJ, Lin C-P, Risso D, Chen S, Kim TA, Tan MH, Li JB, Wu Y, Chen C, Xuan Z, et al. 2017. Deficiency of microRNA *miR-34a* expands cell fate potential in pluripotent stem cells. *Science* **355:** eaag1927. doi:10.1126/science.aag1927

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18:** 71–86. doi:10.1038/nrg.2016.139

Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25:** 1422–1423. doi:10.1093/bioinformatics/btp163

Costas J. 2003. Molecular characterization of the recent intragenomic spread of the murine endogenous retrovirus MuERV-L. *J Mol Evol* **56:** 181–186. doi:10.1007/s00239-002-2392-3

DeBerardinis RJ, Kazazian HH. 1999. Analysis of the promoter from an expanding mouse retrotransposon subfamily. *Genomics* **56:** 317–323. doi:10.1006/geno.1998.5729

De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. 2017. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* **49:** 941–945. doi:10.1038/ng.3858

De Iaco A, Verp S, Offner S, Grun D, Trono D. 2020. DUX is a non-essential synchronizer of zygotic genome activation. *Development* **147:** dev177725. doi:10.1242/dev.177725

Demircioğlu D, Cukuroglu E, Kindermans M, Nandi T, Calabrese C, Fonseca NA, Kahles A, Lehmann K-V, Stegle O, Brazma A, et al. 2019. A pan-cancer transcriptome analysis reveals pervasive regulation through alternative promoters. *Cell* **178:** 1465–1477.e17. doi:10.1016/j.cell.2019.08.018

Deng Q, Ramsköld D, Reinius B, Sandberg R. 2014. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343:** 193–196. doi:10.1126/science.1245316

Deshpande A, Shetty PMV, Frey N, Rangrez AY. 2022. SRF: a seriously responsible factor in cardiac development and disease. *J Biomed Sci* **29:** 38. doi:10.1186/s12929-022-00820-3

Evsikov AV, de Vries WN, Peaston AE, Radford EE, Fancher KS, Chen FH, Blake JA, Bult CJ, Latham KE, Solter D, et al. 2004. Systems biology of the 2-cell mouse embryo. *Cytogenet Genome Res* **105:** 240–250. doi:10.1159/000078195

Fadloun A, Le Gras S, Jost B, Ziegler-Birling C, Takahashi H, Gorab E, Carninci P, Torres-Padilla M-E. 2013. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat Struct Mol Biol* **20:** 332–338. doi:10.1038/nsmb.2495

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* **9:** 397–405. doi:10.1038/nrg2337

Franke V, Ganesh S, Karlic R, Malik R, Pasulka J, Horvat F, Kuzman M, Fulka H, Cernohorska M, Urbanova J, et al. 2017. Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. *Genome Res* **27:** 1384–1394. doi:10.1101/gr.216150.116

Goodier JL, Ostertag EM, Du K, Kazazian HH. 2001. A novel active L1 retrotransposon subfamily in the mouse. *Genome Res* **11:** 1677–1685. doi:10.1101/gr.198301

Greenberg ME, Ziff EB. 1984. Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. *Nature* **311:** 433–438. doi:10.1038/311433a0

Greenberg ME, Siegfried Z, Ziff EB. 1987. Mutation of the c-*fos* gene dyad symmetry element inhibits serum inducibility of transcription in vivo and the nuclear regulatory factor binding in vitro. *Mol Cell Biol* **7:** 1217–1225. doi:10.1128/mcb.7.3.1217-1225.1987

Guo M, Zhang Y, Zhou J, Bi Y, Xu J, Xu C, Kou X, Zhao Y, Li Y, Tu Z, et al. 2019. Precise temporal regulation of Dux is important for embryo development. *Cell Res* **29:** 956–959. doi:10.1038/s41422-019-0238-4

Guo Y, Kitano T, Inoue K, Murano K, Hirose M, Li TD, Sakashita A, Ishizu H, Ogonuki N, Matoba S, et al. 2024. Obox4 promotes zygotic genome activation upon loss of Dux. *eLife* **13:** e95856. doi:10.7554/eLife.95856

Hamatani T, Carter MG, Sharov AA, Ko MSH. 2004. Dynamics of global gene expression changes during mouse preimplantation development. *Dev Cell* **6:** 117–131. doi:10.1016/S1534-5807(03)00373-3

Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim J-W, Wike CL, Weaver BD, Pflueger C, Emery BR, Wilcox AL, et al. 2017. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* **49:** 925–934. doi:10.1038/ng.3844

Hume MA, Barrera LA, Gisselbrecht SS, Bulyk ML. 2015. UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. *Nucleic Acids Res* **43:** D117–D122. doi:10.1093/nar/gku1045

Ishiuchi T, Enriquez-Gasca R, Mizutani E, Bošković A, Ziegler-Birling C, Rodriguez-Terrones D, Wakayama T, Vaquerizas JM, Torres-Padilla M-E. 2015. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat Struct Mol Biol* **22:** 662–671. doi:10.1038/nsmb.3066

Jachowicz JW, Bing X, Pontabry J, Bošković A, Rando OJ, Torres-Padilla M-E. 2017. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat Genet* **49:** 1502–1510. doi:10.1038/ng.3945

Ji S, Chen F, Stein P, Wang J, Zhou Z, Wang L, Zhao Q, Lin Z, Liu B, Xu K, et al. 2023. OBOX regulates mouse zygotic genome activation and early development. *Nature* **620:** 1047–1053. doi:10.1038/s41586-023-06428-3

Jin Y, Tam OH, Paniagua E, Hammell M. 2015. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31:** 3593–3599. doi:10.1093/bioinformatics/btv422

Jukam D, Shariati SAM, Skotheim JM. 2017. Zygotic genome activation in vertebrates. *Dev Cell* **42:** 316–332. doi:10.1016/j.devcel.2017.07.026

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In *Mammalian protein metabolism* (ed. Munro HN), pp. 21–132. Academic Press, New York. doi:10.1016/B978-1-4832-3211-9.50009-7

Kigami D, Minami N, Takayama H, Imai H. 2003. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol Reprod* **68:** 651–654. doi:10.1095/biolreprod.102.007906

Kinene T, Wainaina J, Maina S, Boykin LM. 2016. Rooting trees, methods for. In *Encyclopedia of evolutionary biology* (ed. Kliman RM), pp. 489–493. Academic Press, New York. doi:10.1016/B978-0-12-800049-6.00215-8

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921. doi:10.1038/35057062

Leigh JW, Bryant D. 2015. Popart: full-feature software for haplotype network construction. *Methods Ecol Evol* **6:** 1110–1116. doi:10.1111/2041-210X.12410

Letunic I, Bork P. 2024. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* **52:** W78–W82. doi:10.1093/nar/gkae268

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25:** 2078–2079. doi:10.1093/bioinformatics/btp352

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550. doi:10.1186/s13059-014-0550-8

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487:** 57–63. doi:10.1038/nature11244

Miano JM. 2010. Role of serum response factor in the pathogenesis of disease. *Lab Invest* **90:** 1274–1284. doi:10.1038/labinvest.2010.104

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-Tree 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* **37:** 1530–1534. doi:10.1093/molbev/msaa015

Mohun TJ, Chambers AE, Towers N, Taylor MV. 1991. Expression of genes encoding the transcription factor SRF during early development of *Xenopus laevis*: identification of a CArG box-binding activity as SRF. *EMBO J* **10:** 933–940. doi:10.1002/j.1460-2075.1991.tb08027.x

Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520–562. doi:10.1038/nature01262

Newburger DE, Bulyk ML. 2009. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res* **37:** D77–D82. doi:10.1093/nar/gkn660

Oliveira DS, Fablet M, Larue A, Vallier A, Carareto CMA, Rebollo R, Vieira C. 2023. ChimeraTE: a pipeline to detect chimeric transcripts derived from genes and transposable elements. *Nucleic Acids Res* **51:** 9764–9784. doi:10.1093/nar/gkad671

Oomen ME, Rodriguez-Terrones D, Kurome M, Zakhartchenko V, Mottes L, Simmet K, Noll C, Nakatani T, Mourra-Diaz CM, Aksoy I, et al. 2025. An atlas of transcription initiation reveals regulatory principles of gene and transposable element expression in early mammalian development. *Cell* **188:** 1156–1174.e20. doi:10.1016/j.cell.2024.12.013

Park S-J, Shirahige K, Ohsugi M, Nakai K. 2015. DBTMEE: a database of transcriptome in mouse early embryos. *Nucleic Acids Res* **43:** D771–D776. doi:10.1093/nar/gku1001

Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB. 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* **7:** 597–606. doi:10.1016/j.devcel.2004.09.004

Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. 2013. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* **10:** 1096–1098. doi:10.1038/nmeth.2639

Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9:** 171–181. doi:10.1038/nprot.2014.006

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26:** 841–842. doi:10.1093/bioinformatics/btq033

Rodriguez-Terrones D, Torres-Padilla M-E. 2018. Nimble and ready to mingle: transposon outbursts of early development. *Trends Genet* **34:** 806–820. doi:10.1016/j.tig.2018.06.006

Sakashita A, Kitano T, Ishizu H, Guo Y, Masuda H, Ariura M, Murano K, Siomi H. 2023. Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nat Genet* **55:** 484–495. doi:10.1038/s41588-023-01324-y

Santana-Garcia W, Castro-Mondragon JA, Padilla-Gálvez M, Nguyen NTT, Elizondo-Salas A, Ksouri N, Gerbes F, Thieffry D, Vincens P, Contreras-Moreira B, et al. 2022. RSAT 2022: regulatory sequence analysis tools. *Nucleic Acids Res* **50:** W670–W676. doi:10.1093/nar/gkac312

Shore P, Sharrocks AD. 1995. The MADS-box family of transcription factors. *Eur J Biochem* **229:** 1–13. doi:10.1111/j.1432-1033.1995.tb20430.x

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol* **7:** 539. doi:10.1038/msb.2011.75

Steel M. 2010. The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing edited by Lemey, P., Salemi, M., and Vandamme, A.-M. *Biometrics* **66:** 324–325. doi:10.1111/j.1541-0420.2010.01388.x

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA* **12:** 2. doi:10.1186/s13100-020-00230-y

Svoboda P, Stein P, Anger M, Bernstein E, Hannon GJ, Schultz RM. 2004. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev Biol* **269:** 276–285. doi:10.1016/j.ydbio.2004.01.028

Tanaka E, Bailey T, Grant CE, Noble WS, Keich U. 2011. Improved similarity scores for comparing motifs. *Bioinformatics* **27:** 1603–1609. doi:10.1093/bioinformatics/btr257

Thomas J, Perron H, Feschotte C. 2018. Variation in proviral content among human genomes mediated by LTR recombination. *Mob DNA* **9:** 36. doi:10.1186/s13100-018-0142-3

Torres-Padilla M-E, Bannister AJ, Hurd PJ, Kouzarides T, Zernicka-Goetz M. 2006. Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. *Int J Dev Biol* **50:** 455–461. doi:10.1387/ijdb.052073mt

Treisman R. 1986. Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. *Cell* **46:** 567–574. doi:10.1016/0092-8674(86)90882-2

Treisman R. 1987. Identification and purification of a polypeptide that binds to the c-fos serum response element. *EMBO J* **6:** 2711–2717. doi:10.1002/j.1460-2075.1987.tb02564.x

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25:** 1189–1191. doi:10.1093/bioinformatics/btp033

Whiddon JL, Langford AT, Wong C-J, Zhong JW, Tapscott SJ. 2017. Conservation and innovation in the DUX4-family gene network. *Nat Genet* **49:** 935–940. doi:10.1038/ng.3846

Wickham H. 2016. *Ggplot2*. Springer International Publishing, Cham, Switzerland.

Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, Zhang B, Liu B, Wang Q, Xia W, et al. 2016. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534:** 652–657. doi:10.1038/nature18606

Zhang W, Chen F, Chen R, Xie D, Yang J, Zhao X, Guo R, Zhang Y, Shen Y, Göke J, et al. 2019. Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. *Nucleic Acids Res* **47:** 8485–8501. doi:10.1093/nar/gkz594