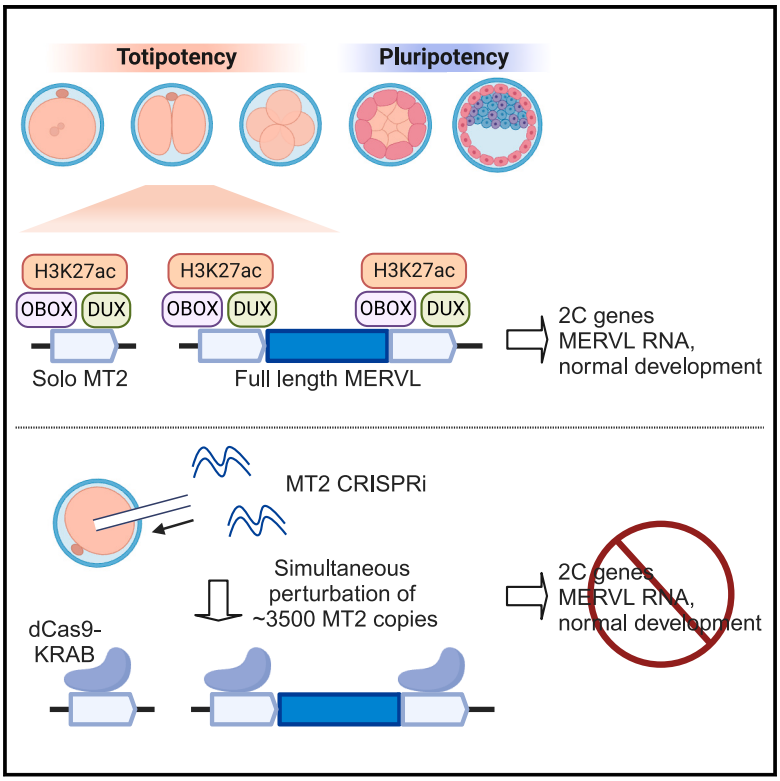
Cell Reports

Systematic evaluation of retroviral LTRs as *cis*-regulatory elements in mouse embryos

Graphical abstract



Highlights

- Zygotic injection of dCas9-KRAB mRNA achieves robust CRISPRi in 2-cell embryos
- MT2 CRISPRi efficiently perturbs thousands of MT2/MERVL copies during mouse ZGA
- MT2 CRISPRi causes ZGA and preimplantation defects
- MT2-derived promoters/enhancers function downstream of OBOX/DUX proteins

Authors

Jian Yang, Lauryn Cook, Zhiyuan Chen

Correspondence

zhiyuan.chen@cchmc.org

In brief

MT2 retrotransposon is transiently activated in mouse 2-cell embryos and is a well-known ZGA and totipotency marker. Yang et al. utilize a robust epigenome editing approach to directly evaluate biological significance of MT2 activation and systematically define *cis*-regulatory activities of ~3,500 MT2 copies in mouse preimplantation embryos.

Yang et al., 2024, Cell Reports *43*, 113775
March 26, 2024
https://doi.org/10.1016/j.celrep.2024.113775

Cell Reports

CellPress
OPEN ACCESS

Report

Systematic evaluation of retroviral LTRs as *cis*-regulatory elements in mouse embryos

Jian Yang,[1,2] Lauryn Cook,[1,2] and Zhiyuan Chen[1,2,3,*]
[1]Reproductive Sciences Center, Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA
[2]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA
[3]Lead contact
*Correspondence: zhiyuan.chen@cchmc.org
https://doi.org/10.1016/j.celrep.2024.113775

SUMMARY

In mammals, many retrotransposons are de-repressed during zygotic genome activation (ZGA). However, their functions in early development remain elusive largely due to the challenge to simultaneously manipulate thousands of retrotransposon insertions in embryos. Here, we applied CRISPR interference (CRISPRi) to perturb the long terminal repeat (LTR) MT2_Mm, a well-known ZGA and totipotency marker that exists in ~2,667 insertions throughout the mouse genome. CRISPRi robustly perturbed 2,485 (~93%) MT2_Mm insertions and 1,090 (~55%) insertions of the closely related MT2C_Mm in 2-cell embryos. Remarkably, such perturbation caused downregulation of hundreds of ZGA genes and embryonic arrest mostly at the morula stage. Mechanistically, MT2 LTRs are globally enriched for open chromatin and H3K27ac and function as promoters/enhancers downstream of OBOX/DUX proteins. Thus, we not only provide direct evidence to support the functional importance of MT2 activation in development but also systematically define *cis*-regulatory function of MT2 in embryos by integrating functional perturbation and multi-omic analyses.

INTRODUCTION

Retrotransposons comprise ~40% of the human genome.[1] Retrotransposons are generally silenced in somatic cells but some of them are highly expressed in germ cells and early embryos.[2,3] It is believed that these retrotransposons act as *cis*-regulatory elements to enable successful gametogenesis and early development. For example, MTC- and MT2B2-derived alternative promoters have been shown to be essential for mouse oogenesis and preimplantation development, respectively.[4,5] In these two studies, experiments were carried out by knocking out single MTC and MT2B2 insertions. In addition to single-copy knockout (KO), small interfering RNA (siRNA), antisense oligonucleotides (ASOs), TALEN- and CRISPR-based methods have been used to simultaneously target many retrotransposon insertions.[6–12]

MT2_Mm is the long terminal repeat (LTR) of the mouse endogenous retroviruses with leucine tRNA primer (MERVL) element.[13] MT2C_Mm is closely related to MT2_Mm; thus these two LTRs together are referred to as MT2 in this study. Both MT2 and MERVL have been used as zygotic genome activation (ZGA) and totipotency markers in mice since 2012.[14,15] However, direct evidence to support the functional importance of MT2 activation in embryos remains lacking. This is largely due to the technical challenge to perturb thousands of MT2 insertions throughout the mouse genome. Recently, a remarkable study efficiently knocked down MERVL and the flanking MT2_Mm by ASOs.[6] However, ASOs used in this study exclusively target MERVL and do not perturb solo MT2 (see below for details). MT2 LTRs have been reported to form a couple of hundred or fewer LTR chimeric transcripts in 2-cell (2C) embryos,[5,14,16] but a comprehensive evaluation of *cis*-regulatory function of MT2 by functional perturbation and integrative epigenomic analyses has not been achieved. Here, we address these questions by developing a computational and experimental framework to systematically define *cis*-regulatory function of retroviral LTRs in preimplantation embryos.

RESULTS

Dynamic expression of MT2 single insertions in preimplantation embryos

Although MT2 LTRs are generally considered to be 2-cell specific, whether different copies have similar expression dynamics has not been assessed. We thus classified MT2 LTRs into different categories such as solo MT2 or MERVL associated and determined their expression levels using uniquely mapped total and poly-A RNA-seq reads.[17–19] We defined 1,062 solo MT2_Mm that do not have any MERVL nearby and 591 full-length MERVL. The rest of the MT2_Mm LTRs (423) are associated with truncated MERVL. For MT2C_Mm LTRs, most of them (1,715) are solo LTRs, and only 267 of them are associated with MERVL. The majority of MT2 insertions have similar activation kinetics in preimplantation embryos with some having higher expression levels than others (Figures 1A, S1A, and S1B).

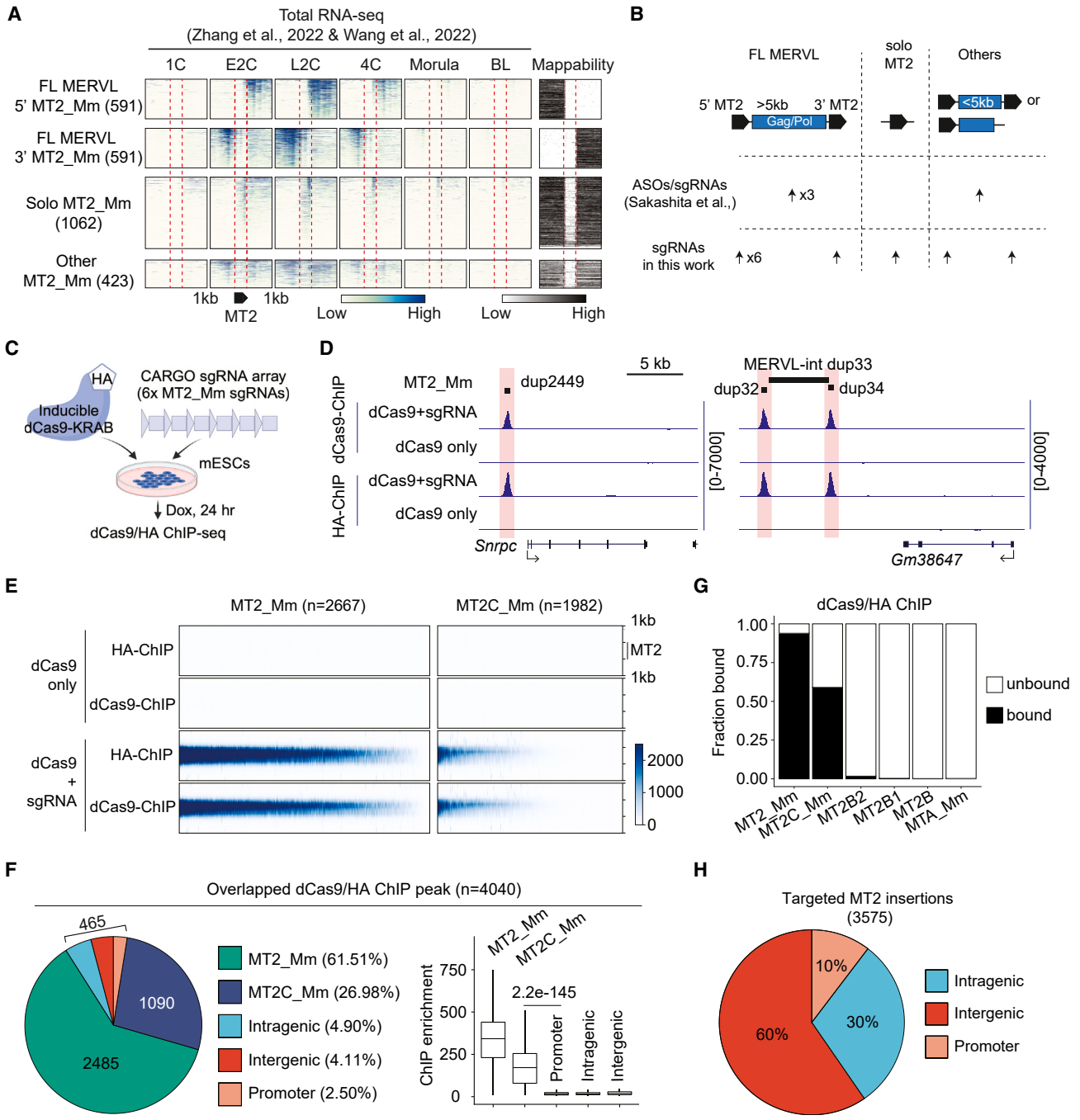


Figure 1. Selective dCas9 targeting to MT2 by CARGO

(A) Expression dynamics of single MT2_Mm copies in preimplantation embryos. FL: full-length; E2C/L2C: early/late 2-cell; BL: blastocyst.

(B) Schematics illustrating the different targeting designs between the two studies.

(C) Experimental design of dCas9/HA ChIP-seq.

(D) Genome browser views of dCas9 and HA ChIP signals at a solo MT2_Mm and a full-length MERVL.

(E) Heatmap illustrating ChIP signals over MT2_Mm and MT2C_Mm insertions.

(F) Pie chart showing genomic distributions of high-confidence ChIP-seq peaks (left panel). Boxplot showing ChIP signals at the targeted MT2 insertions and the off-target sites (right panel). The middle lines in the boxes represent medians. Box hinges indicate the 25th and 75th percentiles, and whiskers indicate the hinge ±1.5 × interquartile range. p value: two-sided Wilcoxon rank-sum test.

(G) Stacked plot showing the fractions of LTRs bound by dCas9.

(H) Pie chart showing genomic distributions of the targeted MT2 insertions. See also Figure S1.

Targeting thousands of retroviral LTRs by multiplexed epigenome editing

To perturb MT2 by epigenome editing, six single-guide RNAs (sgRNAs) were designed to target ~94.5% (2,521/2,667) of MT2_Mm insertions (Figure S1C). Notably, ~91.5% of MT2_Mm insertions are predicted to be targeted by at least two sgRNAs with zero mismatch (Figure S1C). This design is critical because multiple sgRNA targeting increases the likelihood of robust epigenome editing.[20] About 5.5% (146) of MT2_Mm insertions are not recognized by any sgRNAs. This is likely because they are too truncated compared to the MT2_Mm consensus (Figure S1D). Since MT2C_Mm consensus is highly similar to MT2_Mm (85% identity), ~28% of MT2C_Mm copies are recognized by the sgRNAs with zero mismatch (Figure S1E). These sgRNAs are predicted to have minimal off-targeting to other LTR subfamilies including MT2B, MT2B1, and MT2B2 (Figure S1E). The key difference between this work and the previous study[6] is the different target designs (Figure 1B). The sgRNAs here are expected to perturb MT2, regardless of whether they are associated with MERVL. In contrast, ASOs/sgRNAs used by the previous study exclusively target MERVL and should not affect solo MT2.

We next evaluated sgRNA on- and off-target effects in mouse embryonic stem cells (ESCs), an *in vitro* approximate of early embryos. We multiplexed the six sgRNAs using the CARGO (chimeric array of gRNA oligonucleotides) method[21] and generated a transgenic ESC line constitutively expressing these sgRNAs (Figures 1C, S1F, and S1G). The ESC line also harbors a doxycycline (dox)-inducible catalytically dead Cas9 fused to a KRAB domain and an hemagglutinin (HA)-tag (dCas9-KRAB-HA). After dox induction, chromatin immunoprecipitation followed by sequencing (ChIP-seq) was performed using antibodies against dCas9 and HA. The parental ESC line with the inducible dCas9-KRAB-HA, but not the MT2 sgRNAs, served as a control.

The dCas9 and HA ChIP peaks were highly overlapped at MT2 (Figures 1D and 1E). We identified 4,040 high-confidence ChIP peaks (Table S1), with most of the dCas9 binding occurring at MT2_Mm (2,485, 93.2% of MT2_Mm insertions) and MT2C_Mm sites (1,090, 55.0% of MT2C_Mm insertions) (Figures 1F and 1G). Note that the MT2C_Mm copies bound by dCas9 had higher expression levels than those not bound by dCas9 at the 2C stage (Figure S1H). These targeted MT2 LTRs were mostly localized in intergenic (60.0%) and intragenic regions (30.0%) (Figure 1H). The remaining 465 peaks were classified as off-targets but had a very low level of ChIP signals (Figure 1F), suggesting weak and/or transient dCas9 binding. Lastly, no significant bindings were observed to other LTRs such as MT2B and MTA (Figure 1G). Thus, these data indicate that the CARGO-dCas9 selectively targets MT2.

Rapid, robust, and specific CRISPRi in mouse embryos

We next injected dCas9-KRAB mRNA and sgRNAs into 1-cell (1C) embryos (referred to as MT2i) and collected 1C/2C embryos for analyses (Figure 2A). Both the non-injected embryos and the non-targeting sgRNA/dCas9-KRAB injection (referred to as CTRi) served as negative controls. The dCas9 protein was detected at ~3 h post injection (Figure S2A), which is before the dramatic increase of MT2 at the early 2C (E2C) stage (Figure 1A). Remarkably, both MT2_Mm and MERVL transcripts were efficiently repressed at the E2C stage (Figure S2B). The ZGA genes *Zscan4* and *Obox3* were not affected by CTRi, suggesting minimal non-specific global transcriptional repression. In addition, the MERVL-Gag protein became undetectable in MT2i L2C embryos (Figures 2B and S2C). The second polar bodies (arrows) served as internal immunostaining positive controls (Figure 2B). It has been shown that the second polar body has both transcriptional and translational activities.[22] The reason why MERVL expression in the polar body was not perturbed is likely due to that the polar body had been extruded during micro-injection between 2 and 4 hpi (Figure 2A). Consistently, there was no dCas9 protein detected in polar bodies (Figure S2A).

Next, total RNA-seq was performed to evaluate specificity and efficiency of CRISPRi in embryos. After confirming data reproducibility (Figure S2D), we compared gene expression between CTRi and non-injected groups. Consistent with qRT-PCR, transcriptomes of these two groups were highly correlated (Figures S2E and S2F), indicating that CTRi did not have a global transcription repression effect. Notably, MT2_Mm and MERVL-int were downregulated over 35-fold at both E2C and L2C stages (Figure 2C). In addition, other related subfamilies including MT2C_Mm, MERVL_2A-int, and MER89 were among the top downregulated repeats (Figure 2D). As expected, MT2B1, MT2B, and MT2B2 remained unchanged. Thus, these data demonstrated rapid (<16 h), robust (>35-fold reduction), and specific CRISPRi-mediated perturbation of retrotransposons in embryos.

We next compared MT2i to the MERVL ASO knockdown (KD) and MERVL CRISPRi (MERVLi).[6] Consistent with the design (Figure 1B), MERVL ASO KD and MERVLi perturbed MERVL but had minimal effects on solo MT2_Mm (Figures 2E and 2G). In addition, ASO KD and MERVLi did not affect MT2C_Mm LTRs, which mostly exist as solo LTRs (Figures 2F, 2G, and S2H). MERVLi showed overall less efficient KD than ASO and MT2i (Figures 2E and S2G). Thus, although both studies perturbed full-length MERVL, the current study achieved a more comprehensive depletion of MT2/MERVL by additionally perturbing solo MT2.

ZGA and preimplantation defects in MT2i embryos

We next sought to evaluate how MT2i may affect early development. Both CTRi and non-injected embryos cultured *ex vivo* developed to the blastocyst stage normally. However, most MT2i embryos were arrested at the morula stage (Figures 3A and 3B). To determine how MT2i may cause the preimplantation defect, we identified differentially expressed genes (DEGs) at E2C and L2C stages (fold change [FC] > 2, false discovery rate [FDR] < 0.05). For both stages, more genes were downregulated (341 for E2C and 1,111 for L2C) than upregulated (158 for E2C and 481 for L2C) in the MT2i group (Figure S3A, Table S2). This indicates that the predominant role of the targeted MT2 is to drive gene activation at these stages. Importantly, most downregulated genes (87.7% for E2C and 67.0% for L2C) are within 50 kb of the targeted repeats (Figure 3C), suggesting that these repeats mainly activate genes *in cis*. In contrast, most upregulated genes (78.5% for E2C and 87.9% for L2C) are >50 kb away from

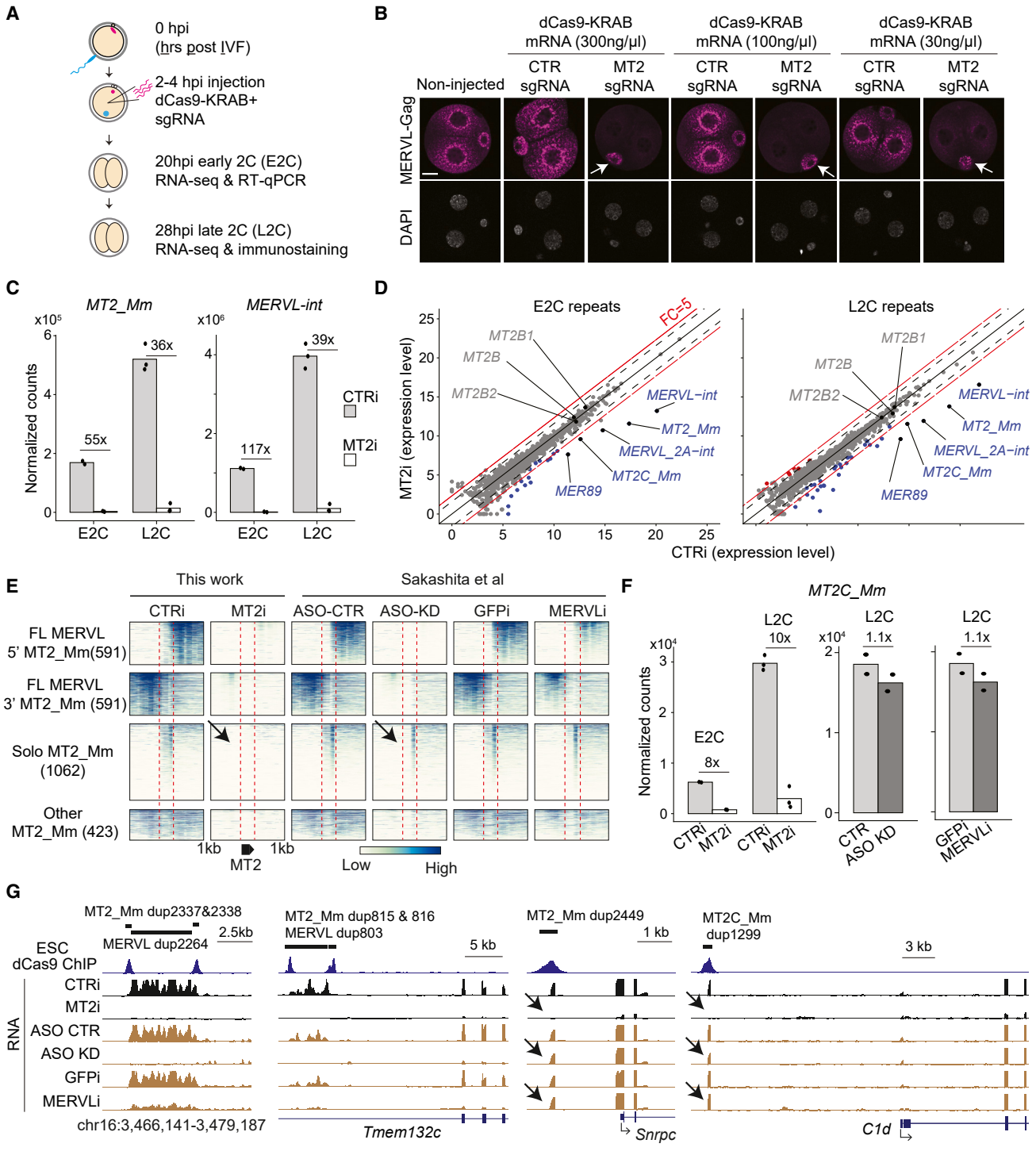


Figure 2. Rapid, robust, and specific perturbation of thousands of MT2 LTRs in embryos

(A) Experimental design. IVF: *in vitro* fertilization.

(B) Immunostaining of MERVL-Gag protein at the L2C stages. White arrows point to the second polar body. Scale bar: 20 µm.

(C) Normalized RNA-seq read counts for MT2_Mm and MERVL-int at E2C and L2C stages.

(D) Scatterplots comparing repeat expression at subfamily level between CTRi and MT2i.

(E) Heatmap illustrating the differences in repeats perturbation by different approaches. Arrows point to the solo MT2 LTRs that are perturbed by MT2i but not by MERVL ASO KD and MERVLi.

(F) Bar plots showing the effects of different perturbation methods on MT2C_Mm expression. Each dot represents one RNA-seq library.

(G) Genome browser views of dCas9 ChIP and RNA signals at two full-length MERVL and two solo MT2 LTRs. See also Figure S2.

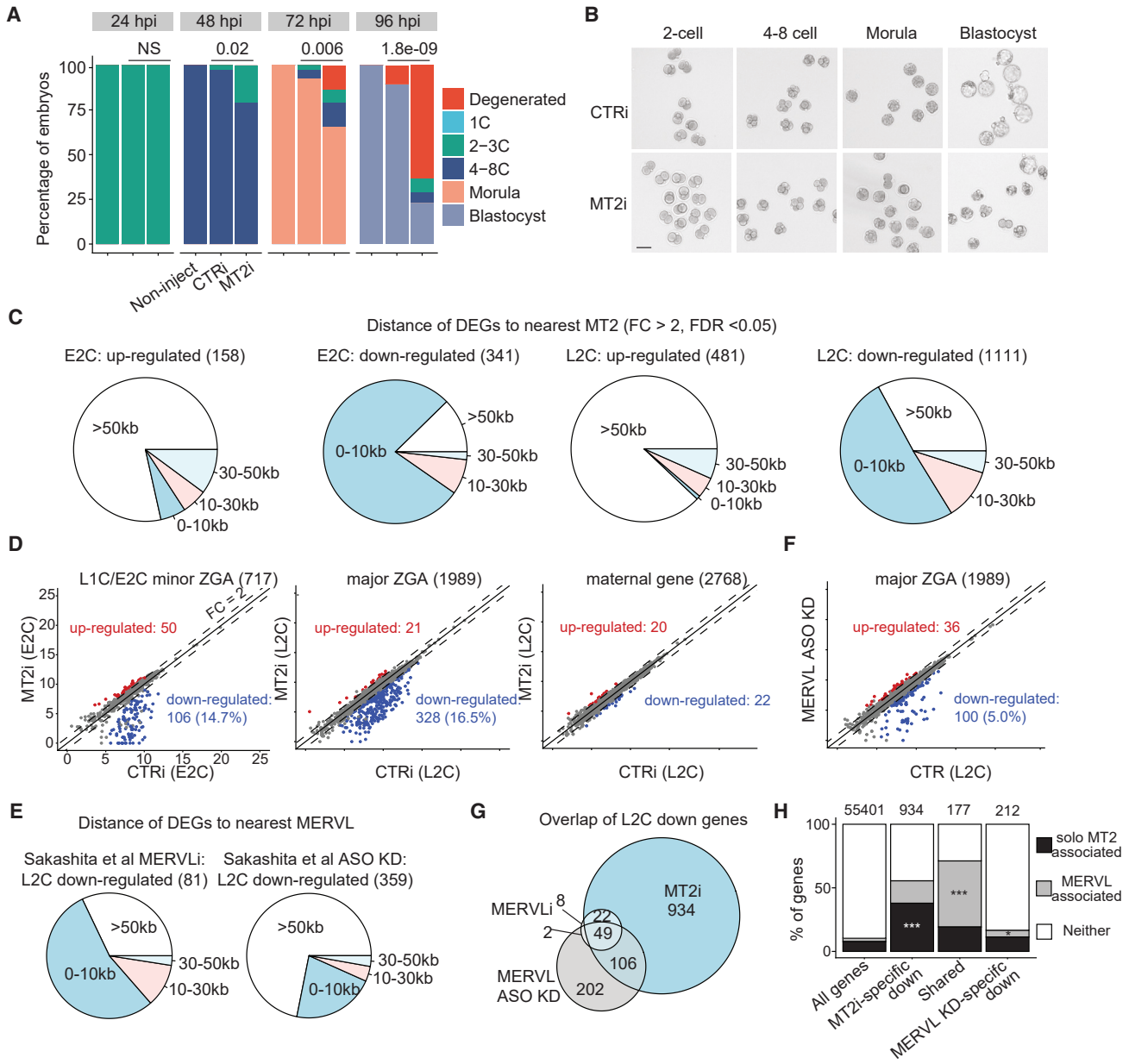


Figure 3. MT2 LTRs regulate ZGA and preimplantation development

(A) Percentage of embryos that reach to the indicated developmental stages at different time points. In total, 18, 38, and 52 embryos were analyzed for non-injected, CTRi, and MT2i groups, respectively. hpi: hours post insemination. p value: chi-square test.

(B) Bright-field images showing embryos that reach to different developmental stages at indicated time points. Scale bar: 70 μm.

(C) Pie charts showing the distances of DEGs to nearest MT2.

(D) Scatterplots showing the expression level changes of minor ZGA, major ZGA, and maternal decay genes in the MT2i group.

(E) Pie charts showing the distances of downregulated genes (FC > 2, p adj. < 0.05) to nearest MERVL.

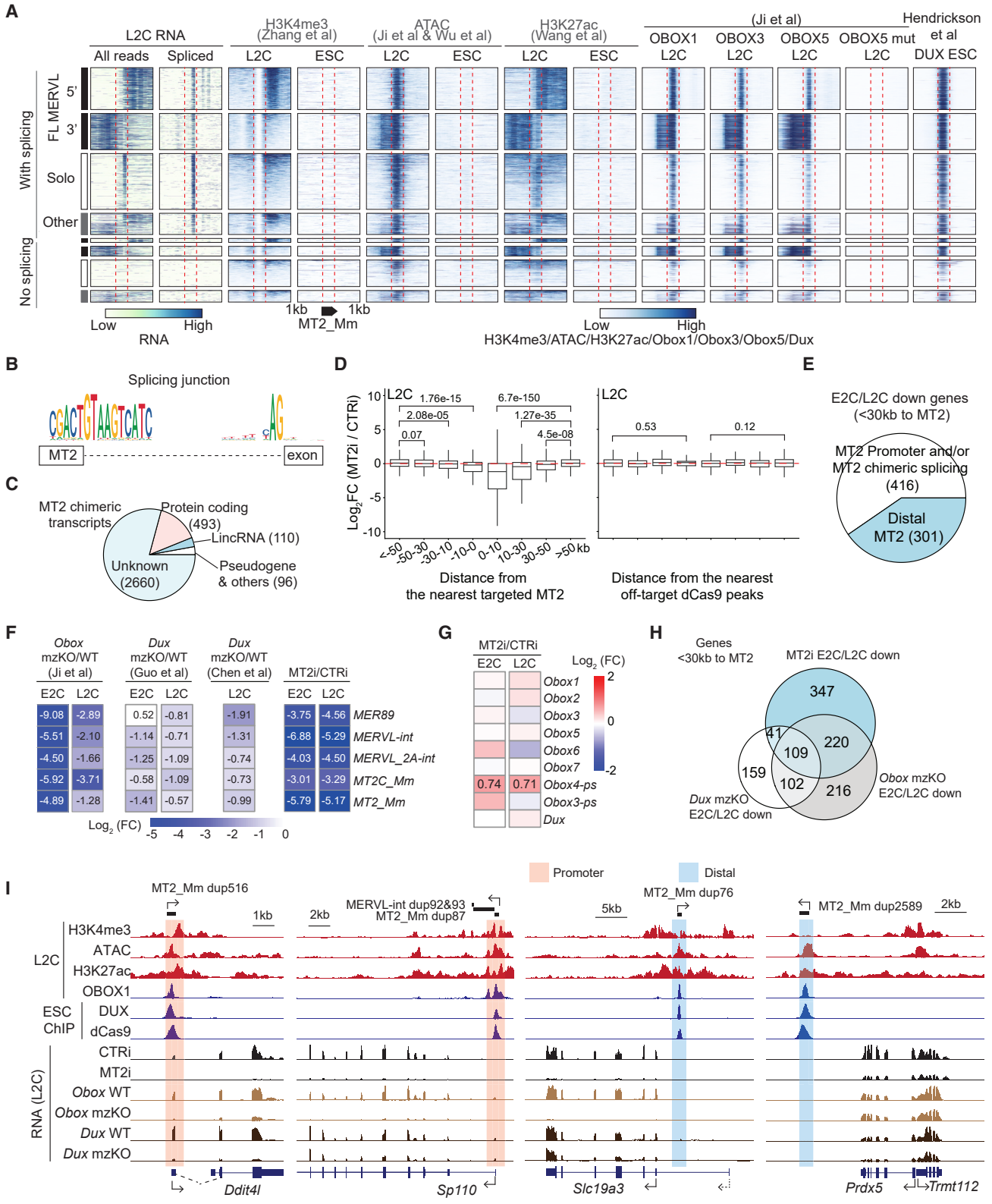
(F) Scatterplot showing the expression level changes of major ZGA genes in the MERVL ASK KD group.

(G) Venn diagram showing the overlap of downregulated genes between MT2i, MERVLi, and MERVL ASO KD at the late 2-cell (L2C) stage.

(H) Stacked plot showing the percentage of downregulated genes that are associated with solo MT2 or MERVL for different groups. Chi-square test: ***p value < 2.2e–16; *p value = 0.02. See also Figure S3.

the targeted repeats, suggesting an indirect effect. The downregulated genes were enriched for Gene Ontology (GO) terms such as "tRNA modification" and "ribonucleoprotein complex biogenesis," whereas the upregulated genes were enriched for "embryonic organ morphogenesis" and "cell fate commitment" (Figure S3A).

We next evaluated how MT2i may affect ZGA. There are two waves of ZGA in mouse embryos. Minor ZGA takes place from



(legend on next page)

L1C to E2C, whereas major ZGA occurs at the L2C stage.[23–25] Using a previously reported cutoff,[26] we identified 717 L1C/E2C minor ZGA genes, 1,989 L2C major ZGA genes, and 2,768 maternal genes using total RNA-seq datasets (Figure S3B and Table S2). A subset of minor (106, 14.7%) and major ZGA genes (328, 16.5%) were downregulated upon MT2i at E2C and L2C, respectively (Figure 3D). The ZGA defects should not be due to developmental delay because maternal decay at the L2C stage was not affected (Figure 3D). Since MT2i embryos can develop to the morula stage (Figures 3A and 3B), we also collected the healthiest 4C at 44 hpi and 8C at 54 hpi embryos for total RNA-seq analyses. The numbers of up- and downregulated genes were comparable at these two stages (Figure S3C and Table S2). The upregulated genes at the 4C stage were enriched for 2C genes (Figure S3D), likely due to developmental delay and/or a failed downregulation by MERVL depletion as previously suggested.[6] The GO terms enriched for the downregulated genes at 4C and 8C included "blastocyst development" and "stem cell population maintenance" (Figure S3C). These results suggest that MT2i-mediated ZGA defects at 2C manifest to cause gene dysregulation at later stages and underly the observed preimplantation phenotype.

We next compared to what extent the dysregulated genes are overlapped between our work and the previous study.[6] Consistent with the fact that our approach had a more comprehensive perturbation of MT2/MERVL (Figures 2E–2G), MT2i led to more downregulated genes (1,111) than MERVL KD (359) and MERVLi (81) at the L2C stage (FC > 2 and FDR < 0.05) (Figure 3E). Consistently, MT2i affected more major ZGA genes than MERVL KD (328 vs. 100) (Figure 3F). For both MT2i and MERVLi, most downregulated genes (67.0% and 67.9%) are within 50 kb of the targeted repeats (Figure 3E). For MERVL ASO KD, however, most downregulated genes (71.8%) are >50 kb from MERVL, which could be explained by the methodology differences (ASO vs. CRISPRi). We also found that 87.6% (71/81) and 43.1% (155/359) of downregulated genes by MERVLi and MERVL KD, respectively, were downregulated by MT2i (Figure 3G). The commonly downregulated genes are significantly associated with MERVL (e.g., *Tmem132c*), whereas MT2i-specific downregulated genes are significantly associated with solo MT2 (e.g., *Snrpc* and *C1d*) (Figures 3H and 2G).

Systematic evaluation of MT2 *cis*-regulatory function

To systematically evaluate *cis*-regulatory functions of MT2, we re-analyzed H3K4me3 (a marker of active promoter), ATAC (open chromatin), and H3K27ac (active enhancer) at the late 2C stage.[17,18,26,27] Strikingly, MT2_Mm are globally enriched for ATAC and H3K27ac signals (Figure 4A). In contrast, different MT2_Mm copies are differentially enriched for H3K4me3. For example, 5′ MT2_Mm LTRs associated with full-length MERVL generally show higher H3K4me3 signals than 3′ MT2_Mm, which is consistent with the notion that 5′ MT2_Mm LTRs initiate MERVL transcription. These observations are not artifacts caused by challenges of mapping reads at repetitive regions for three reasons. First, MT2_Mm LTRs are not enriched for H3K4me3/ATAC/H3K27ac in ESCs (Figure 4A). Second, the ATAC-seq signals at the E2C stage are greatly reduced upon CBP/p300 inhibition (Figure S4B). Third, the open chromatin signals are not enriched at a control repeat subfamily, L1_Mm, at the E2C stage (Figure S4B). Analyses of MT2C_Mm also reached a similar conclusion (Figure S4A). Thus, most MT2 copies act as active *cis*-regulatory elements at the 2C stage.

To better characterize how MT2 LTRs may function as promoters, we determined MT2 chimeric transcription by identifying spliced RNA-seq reads that are mapped to MT2/MERVL using an approach similar to previously reported.[28] We found that splicing occurred mostly at the same splice donor site at MT2 copies (Figures 4A and 4B). Furthermore, the splice acceptor sites are enriched for the typical motif, indicating that these splicing events are not artifacts. Most of these MT2 chimeric transcripts are unannotated (Figure 4C). Note that one MT2 splice donor may have multiple splice acceptor sites. These data also suggest that MT2/MERVL are at the 5′ end of these chimeric transcripts because (1) only splice donor motifs are enriched, and (2) there are no RNA signals immediately upstream of full-length MERVL and solo MT2.

Since the above-described analyses are mostly correlative, we next evaluated *cis*-regulatory function of MT2 by corroborating with MT2 perturbation. We first determined gene expression changes in relation to distances from the nearest targeted MT2 and found that only host genes in direct proximity of the targets (<30 kb) were strongly downregulated (Figure 4D). Note that such downregulation was not observed for genes adjacent to the off-targets identified by weak dCas9 binding in mESCs (Figures 1F and 4D). For the 717 MT2i-downregulated genes that are within 30 kb to nearest MT2, 416 of them are proximal to annotated transcription start sites (TSSs, −1,500 to 500 bp) and/or form MT2 chimeric splicing junctions (Figure 4E, *Ddit4l*, *Sp110* in Figure 4I). We consider that this group of MT2 LTRs mainly function as promoters (Table S3). For the rest of the 301

Figure 4. MT2 LTRs function as *cis*-regulatory elements downstream of OBOX/DUX

(A) Heatmaps illustrating RNA expression, open chromatin, H3K4me3, H3K27ac, OBOX1/3/5, mutant OBOX5, and DUX enrichment at MT2_Mm. MT2_Mm copies are classified into two groups based on whether they have spliced RNA-seq reads.

(B) Sequence logo of MT2 splice donor and acceptor sites.

(C) Percentage of MT2 splicing events that are annotated as protein coding, lincRNA, pseudogenes, or that are unannotated. Note that one MT2 splice donor site may have multiple corresponding splice acceptor sites.

(D) Gene expression changes in relation to distances from targeted repeats and off-target sites. p value: two-sided Wilcoxon rank-sum test.

(E) Pie chart showing the percentage of MT2 copies that function as promoters and distal elements.

(F) Heatmap illustrating expression changes of the indicated repeat subfamilies in *Obox* or *Dux* maternal zygotic KO (mzKO) and MT2i groups.

(G) Heatmap indicating expression changes of *Obox* and *Dux* upon MT2i.

(H) Venn diagram showing the overlap between MT2, OBOX, and DUX regulated genes.

(I) Genome browser views of the indicated genomic loci. Dashed lines at the *Ddit4l* locus represent a splicing junction supported by RNA-seq reads. See also Figure S4.

genes, the associated MT2 copies are far away from annotated TSSs and do not form MT2 chimeric splicing junctions. These MT2 copies are defined as distal elements (Figure 4E, *Slc19a3* and *Prdx5/Trmt112* in Figure 4I, Table S3). Furthermore, about half of the distal MT2 elements are located either in downstream or antisense direction of the associated genes, which is consistent with their putative enhancer functions (Figure S4C). Collectively, we defined contributions of MT2 as *cis*-regulatory elements by integrating MT2 perturbation and multi-omic datasets.

Role of OBOX/DUX in MT2 regulation

Of the transcription factors known to directly activate MT2, OBOX and DUX have been shown to regulate MT2 in embryos.[26,29–33] Thus, we sought to compare the roles of OBOX and DUX in regulating MT2 and MT2-associated genes. Remarkably, 2,278 (85.4%) and 2,464 (92.3%) MT2_Mm copies are bound by DUX and at least one OBOX protein, respectively (Figures 4A and S4D). As a control, MT2_Mm is not bound by the mutant OBOX5 (Figure 4A). For MT2C_Mm, 955 (48.1%) copies are bound by at least one OBOX protein, whereas 203 (10.2%) copies are bound by DUX (Figures S4A and S4D). Thus, OBOX and DUX binding are enriched at MT2 in a genome-wide manner.

We next analyzed expression of repeats in *Obox* and *Dux* maternal zygotic KO (mzKO) embryos.[26,32,33] The MT2i-perturbed repeats are all downregulated in both *Obox* and *Dux* mutants (Figure 4F). Notably, these repeats in general are more affected upon *Obox* depletion, especially at the E2C stage. This suggests that OBOX proteins play a major role in activating these repeats during minor ZGA. Interestingly, these repeats are only partially downregulated in *Obox* and *Dux* L2C mzKO embryos when comparing to MT2i, suggesting that they may compensate each other at this stage.[29] Since MT2i does not affect *Obox* nor *Dux* (Figure 4G), MT2 is unlikely to function upstream of OBOX/DUX in embryos.

We next characterized how MT2i, OBOX, and DUX regulate gene expression in 2C embryos. Of the 717 MT2i-downregulated genes that are within 30 kb of the nearest MT2, 109 (15.2%) are regulated by both OBOX and DUX (e.g., *Ddit4l*, *Gm8994*, *Pwp1*), and 220 (30.6%) and 41 (5.7%) are regulated only by OBOX (e.g., *Sp110*, *Slc19a3*) and DUX (e.g., *Tprkb*), respectively (Figures 4H, 4I, and S4E). This result is consistent with the observation that MT2 activation is more disrupted by depletion of OBOX than DUX (Figure 4F). In addition, a subset of genes (e.g., *Prdx5/Trmt112*) are not affected by OBOX nor DUX depletion, suggesting a redundancy between OBOX and DUX.[29] In summary, these analyses indicate that OBOX and DUX activate hundreds of genes at the 2C stage through regulating MT2.

DISCUSSION

MT2 LTRs have been known for initiating LTR chimeric transcripts in embryos.[5,14,16,34] Here, we extend this observation by discovering that most MT2_Mm copies contribute to chimeric splicing junctions for thousands of unannotated transcripts. We also defined MT2/MERVL copies that may function as putative enhancers. Some of these distal MT2 LTRs are not transcribed (i.e., *Slc19a3* and *Prdx5/Trmt112* in Figure 4I), whereas others are (i.e., *Zscan4c*, *Tprkb*, and *Pwp1* in Figure S4E). The RNAs transcribed from these copies do not have poly-A because they are detected by total RNA-seq (i.e., CTRi/MT2i datasets) but not poly-A RNA-seq (*Obox*/*Dux* datasets). It is possible that these MT2/MERVL LTRs function as dual promoter and enhancer elements.[35] Specifically, they function as promoters for the MT2/MERVL transcripts but enhancers for nearby protein coding genes. It will be interesting to tease apart the role of transcripts vs. enhancer activities of MT2 in future studies.

We also noted that some MT2-associated genes were downregulated by OBOX and/or DUX depletion but not affected by MT2i (Figure 4H). For example, *Zscan4c* is associated with a distal MT2/MERVL but remained unchanged upon MT2i (Figure S4E). Since CRISPR activation of MT2 induces *Zscan4c* in ESCs,[36] this distal MT2/MERVL may be sufficient but not essential for activating *Zscan4c*. It is likely that DUX compensates for *Zscan4c* activation because DUX directly binds to the *Zscan4c* promoter, and *Dux* mzKO causes downregulation of *Zscan4c* in embryos (Figure S4E). This example also highlights the importance of evaluating *cis*-regulatory function by functional perturbation because not all MT2 LTRs with promoter/enhancer signatures are essential for nearby gene expression in embryos.

A few groups have employed epigenome editing to perturb retrotransposons in mouse embryos. Jachowicz et al. reported moderate reduction of LINE1 levels using a TALEN-based approach.[9] Sakashita et al. achieved ~50% KD of MERVL mRNA by injecting a dCas9-KRAB-MECP2 plasmid into paternal pronuclei of zygotes.[6] Here, we found that injection of dCas9-KRAB mRNA and targeting sgRNA can achieve more robust perturbation (>35-fold), which is likely due to that mRNA injection produces more abundant dCas9-KRAB proteins. In addition, cytoplasmic injection used in our method is technically simpler than pronuclei injection. Since our approach does not require transgenic animals, we expect that it could be adapted to investigate retrotransposon function at the subfamily level in other mammalian early embryos.

Limitations of the study

Since MT2 CRISPRi disrupts both MT2-derived promoters/enhancers and MT2-driven MERVL RNA, understanding which mechanism(s) contribute to the severe preimplantation defects observed in MT2i condition remains not fully addressed. Notably, previous work indicated that MERVL LTRs repress 2C genes after ZGA.[6] However, our work suggests that MT2/MERVL mainly activate genes in *cis* during ZGA. How these antagonizing functions are differentially executed requires further investigations. Given that ASO and CRISPRi use different mechanisms to perturb targets, it is not ideal to directly compare transcriptome changes between MT2 CRISPRi (this study) vs. MERVL ASO KD.[6] In the future, the same methodology should be used to dissect how solo MT2 and full-length MERVL may similarly and/or differentially affect gene expression and chromatin dynamics in early development. Another outstanding question not addressed in this work is how MT2 chimeric transcripts regulate development. Future studies should focus on identifying and characterizing individual MT2 copies that are important for ZGA and embryo development.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Mouse embryonic stem cell culture
  - Mouse study
- METHOD DETAILS
  - MT2/MERVL classifications
  - sgRNAs design and synthesis
  - dCas9-KRAB mRNA synthesis
  - Generation of transgenic ESC lines
  - Animal maintenance
  - Collection of mouse embryos and micro-injection
  - Whole-mount immunostaining
  - RT-qPCR
  - ChIP-seq library preparation and sequencing
  - RNA-seq library preparation and sequencing
  - ChIP-seq data processing
  - RNA data processing
  - Define ZGA and maternal decay genes
  - MT2 proximal/distal elements analyses
  - Public datasets used in this study
  - Visualization and statistical analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.113775.

ACKNOWLEDGMENTS

We thank Drs. Yang Yu and Yueh-Chiang Hu for helping with mouse IVF and micro-injection system. We thank Drs. Makiko Iwafuchi, Azusa Inoue, and Paula Stein for helpful comments and editing on the paper. This project was supported by Z.C.'s start-up funding and the Trustee award from CCHMC. Z.C. is also supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R00HD104902).

AUTHOR CONTRIBUTIONS

Conceptualization, Z.C.; methodology, Z.C.; investigation, J.Y., L.C., and Z.C.; writing – original draft, Z.C.; writing – review & editing, J.Y., L.C., and Z.C.; funding acquisition, Z.C.; supervision, Z.C.; project administration, Z.C.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: September 12, 2023
Revised: November 27, 2023
Accepted: January 24, 2024

REFERENCES

1. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., 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. https://doi.org/10.1038/35057062.
2. Fueyo, R., Judd, J., Feschotte, C., and Wysocka, J. (2022). Roles of transposable elements in the regulation of mammalian transcription. Nat. Rev. Mol. Cell Biol. *23*, 481–497. https://doi.org/10.1038/s41580-022-00457-y.
3. Modzelewski, A.J., Gan Chong, J., Wang, T., and He, L. (2022). Mammalian genome innovation through transposon domestication. Nat. Cell Biol. *24*, 1332–1340. https://doi.org/10.1038/s41556-022-00970-4.
4. Flemr, M., Malik, R., Franke, V., Nejepinska, J., Sedlacek, R., Vlahovicek, K., and Svoboda, P. (2013). A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. Cell *155*, 807–816. https://doi.org/10.1016/j.cell.2013.10.001.
5. Modzelewski, A.J., Shao, W., Chen, J., Lee, A., Qi, X., Noon, M., Tjokro, K., Sales, G., Biton, A., Anand, A., et al. (2021). A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development. Cell *184*, 5541–5558.e22. https://doi.org/10.1016/j.cell.2021.09.021.
6. Sakashita, A., Kitano, T., Ishizu, H., Guo, Y., Masuda, H., Ariura, M., Murano, K., and Siomi, H. (2023). Transcription of MERVL retrotransposons is required for preimplantation embryo development. Nat. Genet. *55*, 484–495. https://doi.org/10.1038/s41588-023-01324-y.
7. Percharde, M., Lin, C.J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell *174*, 391–405.e19. https://doi.org/10.1016/j.cell.2018.05.043.
8. Huang, Y., Kim, J.K., Do, D.V., Lee, C., Penfold, C.A., Zylicz, J.J., Marioni, J.C., Hackett, J.A., and Surani, M.A. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. Elife *6*, e22345. https://doi.org/10.7554/eLife.22345.
9. Jachowicz, J.W., Bing, X., Pontabry, J., Bošković, A., Rando, O.J., and Torres-Padilla, M.E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat. Genet. *49*, 1502–1510. https://doi.org/10.1038/ng.3945.
10. Fuentes, D.R., Swigut, T., and Wysocka, J. (2018). Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. Elife *7*, e35989. https://doi.org/10.7554/eLife.35989.
11. Todd, C.D., Deniz, Ö., Taylor, D., and Branco, M.R. (2019). Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. Elife *8*, e44344. https://doi.org/10.7554/eLife.44344.
12. Yu, H., Chen, M., Hu, Y., Ou, S., Yu, X., Liang, S., Li, N., Yang, M., Kong, X., Sun, C., et al. (2022). Dynamic reprogramming of H3K9me3 at hominoid-specific retrotransposons during human preimplantation development. Cell Stem Cell *29*, 1031–1050.e12. https://doi.org/10.1016/j.stem.2022.06.006.
13. Ribet, D., Louvet-Vallée, S., Harper, F., de Parseval, N., Dewannieux, M., Heidmann, O., Pierron, G., Maro, B., and Heidmann, T. (2008). Murine endogenous retrovirus MuERV-L is the progenitor of the "orphan" epsilon viruslike particles of the early mouse embryo. J. Virol. *82*, 1622–1625. https://doi.org/10.1128/JVI.02097-07.
14. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57–63. https://doi.org/10.1038/nature11244.
15. Genet, M., and Torres-Padilla, M.E. (2020). The molecular and cellular features of 2-cell-like cells: a reference guide. Development *147*. https://doi.org/10.1242/dev.189688.
16. Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host

genes in mouse oocytes and preimplantation embryos. Dev. Cell *7*, 597–606. https://doi.org/10.1016/j.devcel.2004.09.004.

17. Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., et al. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature *537*, 553–557. https://doi.org/10.1038/nature19361.
18. Wang, M., Chen, Z., and Zhang, Y. (2022). CBP/p300 and HDAC activities regulate H3K27 acetylation dynamics and zygotic genome activation in mouse preimplantation embryos. EMBO J. *41*, e112012. https://doi.org/10.15252/embj.2022112012.
19. Zhang, C., Wang, M., Li, Y., and Zhang, Y. (2022). Profiling and functional characterization of maternal mRNA translation during mouse maternal-to-zygotic transition. Sci. Adv. *8*, eabj3967. https://doi.org/10.1126/sciadv.abj3967.
20. Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., et al. (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat. Methods *10*, 973–976. https://doi.org/10.1038/nmeth.2600.
21. Gu, B., Swigut, T., Spencley, A., Bauer, M.R., Chung, M., Meyer, T., and Wysocka, J. (2018). Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science *359*, 1050–1055. https://doi.org/10.1126/science.aao3136.
22. Jin, H., Han, Y., Wang, H., Li, J.X.H., Shen, W., Zhang, L., Chen, L., Jia, S., Yuan, P., Chen, H., and Meng, A. (2022). The second polar body contributes to the fate asymmetry in the mouse embryo. Natl. Sci. Rev. *9*, nwac003. https://doi.org/10.1093/nsr/nwac003.
23. Aoki, F., Worrad, D.M., and Schultz, R.M. (1997). Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. Dev. Biol. *181*, 296–307. https://doi.org/10.1006/dbio.1996.8466.
24. Bouniol, C., Nguyen, E., and Debey, P. (1995). Endogenous transcription occurs at the 1-cell stage in the mouse embryo. Exp. Cell Res. *218*, 57–62. https://doi.org/10.1006/excr.1995.1130.
25. Schulz, K.N., and Harrison, M.M. (2019). Mechanisms regulating zygotic genome activation. Nat. Rev. Genet. *20*, 221–234. https://doi.org/10.1038/s41576-018-0087-x.
26. 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. https://doi.org/10.1038/s41586-023-06428-3.
27. 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. https://doi.org/10.1038/nature18606.
28. Göke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., and Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell *16*, 135–141. https://doi.org/10.1016/j.stem.2015.01.005.
29. Guo, Y., K.T., Murano, K., Li, T., Sakashita, A., Ishizu, H., Sato, M., and Siomi, H. (2022). Obox4 secures zygotic genome activation upon loss of Dux. Preprint at bioRxiv. https://doi.org/10.1101/2022.07.04.498763.
30. Bosnakovski, D., Gearhart, M.D., Ho Choi, S., and Kyba, M. (2021). Dux facilitates post-implantation development, but is not essential for zygotic genome activationdagger. Biol. Reprod. *104*, 83–93. https://doi.org/10.1093/biolre/ioaa179.
31. De Iaco, A., Verp, S., Offner, S., Grun, D., and Trono, D. (2020). DUX is a non-essential synchronizer of zygotic genome activation. Development *147*, dev177725. https://doi.org/10.1242/dev.177725.
32. Chen, Z., and Zhang, Y. (2019). Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. Nat. Genet. *51*, 947–951. https://doi.org/10.1038/s41588-019-0418-7.
33. 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. https://doi.org/10.1038/s41422-019-0238-4.
34. 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. https://doi.org/10.1101/gr.216150.116.
35. Andersson, R., and Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. Nat. Rev. Genet. *21*, 71–87. https://doi.org/10.1038/s41576-019-0173-8.
36. Yang, F., Huang, X., Zang, R., Chen, J., Fidalgo, M., Sanchez-Priego, C., Yang, J., Caichen, A., Ma, F., Macfarlan, T., et al. (2020). DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces a Totipotent 2C-like State. Cell Stem Cell *26*, 234–250.e7. https://doi.org/10.1016/j.stem.2020.01.004.
37. Tan, X., W.J., Turunen, M., Wong, K., Fernández, E., Paull, E., Jones, S., Wang, J., Noh, H., Salvatori, B., et al. (2022). Interrogation of genome-wide, experimentally dissected gene regulatory networks reveals mechanisms underlying dynamic cellular state control. Preprint at bioRxiv. https://doi.org/10.1101/2021.06.28.449297.
38. Hazelbaker, D.Z., Beccard, A., Angelini, G., Mazzucato, P., Messana, A., Lam, D., Eggan, K., and Barrett, L.E. (2020). A multiplexed gRNA piggyBac transposon system facilitates efficient induction of CRISPRi and CRISPRa in human pluripotent stem cells. Sci. Rep. *10*, 635. https://doi.org/10.1038/s41598-020-57500-1.
39. Concordet, J.P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. *46*, W242–W245. https://doi.org/10.1093/nar/gky354.
40. Bae, S., Park, J., and Kim, J.S. (2014). a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics *30*, 1473–1475. https://doi.org/10.1093/bioinformatics/btu048.
41. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
42. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.
43. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
44. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.
45. Jin, Y., Tam, O.H., Paniagua, E., and Hammell, M. (2015). TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics *31*, 3593–3599. https://doi.org/10.1093/bioinformatics/btv422.
46. Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinf. *11*, 237. https://doi.org/10.1186/1471-2105-11-237.
47. Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS Comput. Biol. *9*, e1003118. https://doi.org/10.1371/journal.pcbi.1003118.
48. Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383. https://doi.org/10.1093/bioinformatics/btv145.

49. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

50. Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). fast processing of NGS alignment formats. Bioinformatics *31*, 2032–2034. https://doi.org/10.1093/bioinformatics/btv098.

51. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

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

53. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

54. Pockrandt, C., Alzamel, M., Iliopoulos, C.S., and Reinert, K. (2020). GenMap: ultra-fast computation of genome mappability. Bioinformatics *36*, 3687–3692. https://doi.org/10.1093/bioinformatics/btaa222.

55. Cotto, K.C., Feng, Y.Y., Ramu, A., Richters, M., Freshour, S.L., Skidmore, Z.L., Xia, H., McMichael, J.F., Kunisaki, J., Campbell, K.M., et al. (2023). Integrated analysis of genomic and transcriptomic data for the discovery of splice-associated variants in cancer. Nat. Commun. *14*, 1589. https://doi.org/10.1038/s41467-023-37266-6.

56. Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics *33*, 3645–3647. https://doi.org/10.1093/bioinformatics/btx469.

57. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006. https://doi.org/10.1101/gr.229102.

58. Storer, J., Hubley, R., Rosen, J., Wheeler, T.J., and Smit, A.F. (2021). The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob. DNA *12*, 2. https://doi.org/10.1186/s13100-020-00230-y.

59. Inoue, A., Chen, Z., Yin, Q., and Zhang, Y. (2018). Maternal Eed knockout causes loss of H3K27me3 imprinting and random X inactivation in the extraembryonic cells. Genes Dev. *32*, 1525–1536. https://doi.org/10.1101/gad.318675.118.

60. Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., 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. https://doi.org/10.1038/ng.3844.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
MERVL-GAG Rabbit Polyclonal antibody	HUABIO	Cat# ER50102; RRID: AB_2636876
HA-Tag (C29F4)	Cell Signaling Technology	Cat# 3724S; RRID: AB_1549585
Cas9 antibody	Diagenode	Cat# C15310258; RRID: AB_2715516
Donkey anti-Rabbit IgG (H + L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488	Thermo Fisher Scientific	Cat# A-21206; RRID: AB_2535792
Donkey anti-Rabbit IgG (H + L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 568	Thermo Fisher Scientific	Cat# A-10042; RRID: AB_2534017
Bacterial and virus strains		
NEB® Stable Competent E. coli (High Efficiency)	New England Biolabs	Cat# C3040I
NEB® 5-alpha Competent E. coli (High Efficiency)	New England Biolabs	Cat# C2987I
Chemicals, peptides, and recombinant proteins		
PrimeSTAR Max DNA Polymerase	Takara	Cat# R045A
EmbryoMax® 0.1% Gelatin Solution	Millipore Sigma	Cat# ES-006
DMEM, high glucose, no glutamine	Thermo Fisher Scientific	Cat# 11960069
Fetal Bovine Serum	HyClone	Cat# SH30396.03; Lot# AE24573266
Penicillin-Streptomycin (5,000 U/mL)	Thermo Fisher Scientific	Cat# 15070063
GlutaMAX™ Supplement	Thermo Fisher Scientific	Cat# 35050061
Sodium Pyruvate (100 mM)	Thermo Fisher Scientific	Cat# 11360070
MEM Non-Essential Amino Acids Solution (100X)	Thermo Fisher Scientific	Cat# 11140076
2-Mercaptoethanol	Thermo Fisher Scientific	Cat# 21985023
ESGRO® Recombinant Mouse LIF Protein	Millipore Sigma	Cat# ESG1107
PD0325901	Millipore Sigma	Cat# 444966
CHIR99021	Millipore Sigma	Cat# SML1046
Gibco™ Geneticin™ Selective Antibiotic (G418 Sulfate) (50 mg/mL)	Gibco	Cat# 10131035
Doxycycline hyclate	Millipore Sigma	Cat# D9891
Blasticidin S HCl, powder	Thermo Fisher Scientific	Cat# R21001
Pregnant Mare Serum Gonadotropin	BioVendor	Cat# RP1782725000
Chorionic gonadotropin human, lyophilized powder, vial of ~2,500 IU	Millipore Sigma	Cat# C1063-1VL
EmbryoMax® Human Tubal Fluid (HTF) (1X), liquid, for Mouse IVF	Millipore Sigma	Cat# MR-070-D
EmbryoMax® KSOM Mouse Embryo Media	Millipore Sigma	Cat# MR-106-D
M2 medium	Millipore Sigma	Cat# M7167
Fast SYBR™ Green Master Mix	Thermo Fisher Scientific	Cat# 4385616
Formaldehyde solution	Millipore Sigma	Cat# Millipore Sigma
Glycine	Millipore Sigma	Cat# G7126
PBS, pH 7.4	Thermo Fisher Scientific	Cat# 10010031
UltraPure™ SDS Solution, 10%	Thermo Fisher Scientific	Cat# 24730020
NaCl (5 M), RNase-free	Thermo Fisher Scientific	Cat# AM9760G
Tris (1 M), pH 8.0, RNase-free	Thermo Fisher Scientific	Cat# AM9855G

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
UltraPure™ 0.5M EDTA, pH 8.0	Thermo Fisher Scientific	Cat# 15575020
cOmplete™, EDTA-free Protease Inhibitor Cocktail	Millipore Sigma	Cat# 11873580001
Triton™ X-100	Millipore Sigma	Cat# T9284
RNase A, DNase and protease-free (10 mg/mL)	Thermo Fisher Scientific	Cat# EN0531
Proteinase K from Tritirachium album	Millipore Sigma	Cat# P2308
UltraPure™ Phenol:Chloroform:Isoamyl Alcohol (25:24:1, v/v)	Thermo Fisher Scientific	Cat# 15593031
Dynabeads™ Protein A for Immunoprecipitation	Thermo Fisher Scientific	Cat# 10008D
ERCC RNA Spike-In Mix	Thermo Fisher Scientific	Cat# 4456740
Critical commercial assays		
mMESSAGE mMACHINE™ T7 ULTRA Transcription Kit	Thermo Fisher Scientific	Cat# AM1345
Lipofectamine™ 3000 Transfection Reagent	Thermo Fisher Scientific	Cat# L3000001
SuperScript™ IV CellsDirect™ cDNA Synthesis Kit	Thermo Fisher Scientific	Cat# 11750150
NEBNext® Ultra™ DNA Library Prep Kit for Illumina®	NEB	Cat# E7370L
SMART-Seq® Stranded Kit	Takara	Cat# 634443
Deposited data		
Total RNA-seq	This study	GSE242121
ChIP-seq	This study	GSE242122
Experimental models: Cell lines		
ES-E14TG2a	ATCC	Cat# CRL-1821
E14TG2a, Dox-inducible dCas9-KRAB-HA	This study	N/A
E14TG2a, Dox-inducible dCas9-KRAB-HA, 6xMT2 sgRNA	This study	N/A
Experimental models: Organisms/strains		
B6D2F1/J	Jax	Strain# 100006; RRID: IMSR_JAX:100006
Oligonucleotides		
Primers, sgRNAs	This study	Table S4
Recombinant DNA		
PB-TRE-dCas9-KRAB-MeCP2	Tan et al.[37]	RRID: addgene_122267
pT7077	Hazelbaker et al.[38]	RRID: addgene_137879
Super piggyBac Transposase expression vector	System Biosciences	Cat# PB210PA-1
pPN458	Hazelbaker et al.[38]	RRID: addgene_137876
pPN458-6xMT2 sgRNA	This study	N/A
Software and algorithms		
CRISPOR v5.0.1	Concordet and Haeussler[39]	http://crispor.tefor.net
Cas-OFFinder v2.4	Bae et al.[40]	http://www.rgenome.net/cas-offinder/
Trimgalore v0.6.6	N/A	https://github.com/FelixKrueger/TrimGalore
bowtie2 v2.4.2	Langmead and Salzberg[41]	https://bowtie-bio.sourceforge.net/bowtie2/manual.shtml
Picard v2.18.22	N/A	http://broadinstitute.github.io/picard/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Macs2 v2.1.4	Zhang et al.[42]	https://hbctraining.github.io/Intro-to-ChIPseq/lessons/05_peak_calling_macs.html
Bedtools v2.27.0	Quinlan and Hall[43]	https://bedtools.readthedocs.io
Deeptools v2.0.0	Ramirez et al.[44]	https://deeptools.readthedocs.io/en/latest/
TEtranscripts v2.2.3	Jin et al.[45]	https://github.com/mhammell-laboratory/TEtranscripts
ChIPpeakAnno v3.36.0	Zhu et al.[46]	https://bioconductor.org/packages/release/bioc/html/ChIPpeakAnno.html
GenomicFeatures v1.54.1	Lawrence et al.[47]	https://bioconductor.org/packages/release/bioc/html/GenomicFeatures.html
ChIPseeker v 1.38.0	Yu et al.[48]	https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html
STAR v2.7.9	Dobin et al.[49]	https://github.com/alexdobin/STAR/blob/master/doc/STARmanual.pdf
Sambamba v0.6.8	Tarasov et al.[50]	https://lomereiter.github.io/sambamba/
StringTie v2.1.7	Pertea et al.[51]	http://ccb.jhu.edu/software/stringtie/index.shtml?t=manual
DESeq2 v1.38.3	Love et al.[52]	https://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html
clusterProfiler v4.6.2	Yu et al.[53]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
GenMap v1.2.0	Pockrandt et al.[54]	https://github.com/cpockrandt/genmap
Regtools v0.5.2	Cotto et al.[55]	https://regtools.readthedocs.io
Ggseqlogo v0.1	Wagih[56]	https://omarwagih.github.io/ggseqlogo/
UCSC genome browser	Kent[57]	https://genome.ucsc.edu/cgi-bin/hgGateway

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Zhiyuan Chen (Zhiyuan.chen@cchmc.org).

Materials availability

Materials generated in this study are available upon request.

Data and code availability

- The sequencing data generated in this study were deposited in the Gene Expression Omnibus under accession number GSE242123.
- The code for this study is available at https://github.com/ZYChen-lab/MT2_CRISPRi. https://doi.org/10.5281/zenodo.10535037
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Mouse embryonic stem cell culture

Mouse ESCs (E14TG2a) were obtained from ATCC and cultured on feeder-free dishes pre-coated with 0.1% gelatin (Sigma-Aldrich). ESCs were cultured in DMEM (high glucose, no glutamine, Thermo Fisher Scientific) medium supplemented with 15% Fetal Bovine Serum, 100U/ml Pen-streptomycin, 2mM GlutaMax, 1mM sodium pyruvate, 0.1mM non-essential amino acids, 0.084mM 2-mercaptoethanol, 1000 units/ml LIF, 0.5μM PD0325901, and 3μM ChIR99021. All ESCs were courted in 5% CO_2, 37°C and humidified incubator. Cells were passaged 1:10-20. MT2 CRIPSRi ESC lines are listed in the key resources table.

Mouse study

Wild type B6D2F1/J (BDF1) (6- to 9-week-old) were purchased from Jax and maintained in the animal facility at Cincinnati Children's Hospital Medical Center (CCHMC). All animal experiments were performed in accordance with the protocols of the Institutional Animal Care and Use Committee (IACUC) of CCHMC (PI: Zhiyuan Chen, protocol #2022-0070).

METHOD DETAILS

MT2/MERVL classifications

The MERVL-int (>5kb) or MERVL_2A-int (>5kb) flanked by two MT2_Mm or MT2C_Mm were classified as full-length elements. In addition, MT2 and MERVL-int/MERVL_2A-int were required to have the same orientations, and the total length (2xMT2+int) should be < 10kb. MT2 copies that do not have any MERVL-int/MERVL_2A-int nearby (>5kb) were classified as solo MT2. The rest were defined as other MT2 copies.

sgRNAs design and synthesis

MT2_Mm consensus sequence was retrieved from the Dfam database,[58] and CRISPOR[39] was used to design sgRNAs with the option "20bp-NGG-SpCas9". The sgRNA target sites were predicted using Cas-OFFinder[40] allowing a maximum of three mismatches. The top six sgRNAs were picked based on the following criteria. First, sgRNAs were predicted to target most MT2_Mm insertions. Second, sgRNAs showed high on-target scores (i.e., "Doench '14-Score" and "Moreno-Mateos-Score"). Third, sgRNAs had minimal off-targeting to other MT2 subfamilies including MT2B, MT2B1, and MT2B2. Given the consensus sequences are highly similar between MT2C_Mm and MT2_Mm, sgRNAs predicted to target MT2C_Mm were not considered as off-targeting. The sgRNAs used for embryo micro-injection experiments were synthesized by Synthego with the following chemical modifications "2′-*O*-Methyl at 3 first and last bases, and 3′ phosphorothioate bonds between first 3 and last 2 bases". All sgRNA sequences are available in Table S4.

dCas9-KRAB mRNA synthesis

The dCas9-HA-NLS-NLS-KRAB-NLS sequence was amplified from the plasmid PB-TRE-dCas9-KRAB-MeCP2[37] (addgene 122267) using PrimeSTAR Max 2× premix (Takara). A T7 promoter sequence was included in forward primer for *in vitro* transcription using the mMESSAGE mMACHINE T7 Ultra Transcription kit (Thermo Fisher Scientific). After LiCl precipitation, mRNA concentration was determined using a NanoDrop spectrophotometer, and stored at −80°C. Primer sequences are available in Table S4.

Generation of transgenic ESC lines

Mouse ESCs (E14TG2a, ATCC) were cultured on feeder-free dishes pre-coated with 0.1% gelatin (Sigma-Aldrich). ESCs were cultured in DMEM (high glucose, no glutamine, Thermo Fisher Scientific) medium supplemented with 15% Fetal Bovine Serum, 100U/ml Pen-streptomycin, 2mM GlutaMax, 1mM sodium pyruvate, 0.1mM non-essential amino acids, 0.084mM 2-mercaptoethanol, 1000 units/ml LIF, 0.5μM PD0325901, and 3μM ChIR99021. Cells were passaged 1:10-20.

To generate ESC lines carrying Dox-inducible dCas9, plasmid pT7077[38] (addgene 137879) was co-transfected with Super piggyBac Transposase expression vector (System Biosciences) using Lipofectamine 3000 Transfection Reagent (Thermo Fisher Scientific). After transfection, cells were selected under 300μg/ml G418 (Fisher Scientific) for two days. After all cells in non-transfected negative control had died, G418 was supplemented at 100μg/ml throughout the rest of the experiments to avoid transgene silencing. Single colonies were picked 4–6 days later, and positive clones were selected based on GFP signal at 24 h post Dox (1 μg/ml)(Sigma Aldrich) induction.

The six MT2 sgRNAs were assembled using the CARGO method[21] and inserted into pPN458[38] BamH-I site (addgene 137876). After sequencing the whole plasmid using Plasmidsaurus, pPN458-6×-MT2_Mm was co-transfected with Super piggyBac Transposase expression vector (System Biosciences) into an ESC line with Dox-inducible dCas9. After transfection, cells were selected under 10μg/ml Blasticidin S HCl (Thermo Fisher Scientific) for two days. Positive clones were picked based on RFP signal. GFP, RFP, and bright field images were acquired using EVOSII cell imaging system (Thermo Fisher Scientific).

Animal maintenance

Wild type B6D2F1/J (BDF1) (6- to 9-week-old) were purchased from Jax and maintained in the animal facility at Cincinnati Children's Hospital Medical Center (CCHMC). All animal experiments were performed in accordance with the protocols of the Institutional Animal Care and Use Committee (IACUC) of CCHMC.

Collection of mouse embryos and micro-injection

Superovulation, metaphase II (MII) eggs collection, *in vitro* fertilization (IVF), embryo micro-injection, and embryo culture were described previously.[32] The time when sperm were added to cumulus oocyte complexes (COCs) was considered as 0 h post IVF (hpi). About 6 pL dCas9-KRAB mRNA and sgRNA mixes were injected into cytoplasm of zygotes between 2 and 4 hpi using a Piezo impact-driven micromanipulator (Eppendorf). For all micro-injection experiments, sgRNAs were injected at 60 ng/μL: 10 ng/μL for

each MT2 sgRNA or 60 ng/μL for the non-targeting control sgRNA. For micro-injection followed by immunostaining, dCas9-KRAB mRNA was injected at 300, 100, or 30 ng/μL. For embryos used for analyses of developmental phenotypes, RT-qPCR, and total RNA-seq, dCas9-KRAB mRNA was injected at 100 ng/μL. Embryos were cultured in KSOM (Millipore) at 37 °C under 5% CO2 with air. Embryo bright field images were acquired using EVOSII cell imaging system (Thermo Fisher Scientific).

Whole-mount immunostaining

Immunostaining was described previously.[59] The primary antibody was rabbit anti-MERVL-Gag (1:200, HUABIO, ER50102) and rabbit anti-HA (1:100, Cell Signaling, #3724). The secondary antibody was Alexa Fluor 568 donkey anti-rabbit IgG (1:200) and Alexa Fluor 488 donkey anti-rabbit IgG (1:200) (Thermo Fisher Scientific). Fluorescence was detected under a laser-scanning confocal microscope (Nikon A1R inverted), and images were analyzed using ImageJ (NIH).

RT-qPCR

Twenty E2C (20hpi) embryos were collected and used for cDNA synthesis using SuperScript IV CellsDirect Lysis Reagents (Thermo Fisher Scientific). *MT2_Mm*, *MERVL-int*, *Obox3*, and *Zscan4* transcript abundance were determined using SYBR green gene expression assay (Thermo Fisher Scientific) in a QuantStudio 6 Flex system (Thermo Fisher Scientific). The Ct cycles were normalized to the *Actb* gene. The primers used are included in Table S4.

ChIP-seq library preparation and sequencing

About 10 million ESCs at 24 h post Dox (1 μg/ml) induction were collected for ChIP-seq library preparation. Cells were fixed in culture medium at room temperature for 10 min with final 1% formaldehyde (Sigma Aldrich), and fixation was quenched by final 0.14M glycine (Sigma Aldrich) at room temperature for 5 min. After cells were washed twice with ice-cold PBS (Thermo Fisher Scientific), cells were collected with scrapers. Following centrifugation (500g, 5 min, and 4°C) and the removal of supernatants, cell pellets were suspended in 1mL SDS lysis buffer (0.5% SDS, 100mM NaCl, 50mM Tris HCl-pH8.0, 5mM EDTA-pH8.0, 1× cOmplete EDTA free proteinase inhibitor cocktail), and incubated on ice for 10 min with a slight vortex every 2 min. After centrifugation (1000g, 5min, and room temperature), cell pellets were resuspended in 0.5mL IP buffer (0.3% SDS, 100mM NaCl, 50mM Tris HCl-pH8.0, 5mM EDTA-pH8.0, 1.6% Triton X-100, 1× cOmplete EDTA free proteinase inhibitor cocktail). Sonication was performed using Fisher Scientific Sonic Dismembrator with the following settings: 30% amplitude, ON = 10s, OFF = 30s for 8 cycles. Following sonication, supernatants containing chromatin were collected after centrifugation (20,000g, 4°C, 10 min). To determine chromatin concentration, about 10 μL chromatin was diluted using 190 μL elution buffer (1% SDS, 50mM Tris HCl-pH8.0, 1mM EDTA-pH8.0). Rest chromatin solution was aliquoted and stored at −80°C. Following reverse crosslinking (65°C, 600 rpm on a thermomixer, overnight), RNase A treatment (0.2 mg/ml, 37°C, 1 h), and Proteinase K (0.2 mg/ml, 55°C, 1hr), DNA was isolated using UltraPure Phenol:Chloroform:Isoamyl Alcohol (Thermo Fisher Scientific) and quantified by a NanoDrop spectrophotometer. Chromatin containing 25μg DNA was used for each ChIP.

For each ChIP, 30μL protein A magnetic dynabeads (Thermo Scientific) were washed with IP buffer 3 times and incubated with primary antibody by rotation at 4°C for 4 h. Primary antibodies were rabbit anti-Cas9 (Diagenode C15310258, 1:100) and rabbit anti-HA (Cell Signaling 3724S, 1:50). After pre-incubation, the beads/antibody were suspended and incubated with 100μL chromatin containing 25μg DNA at 4°C overnight on a rotator. The following day, beads/chromatin were washed once with low salt buffer (0.1% SDS, 150mM NaCl, 20mM Tris-HCl-pH8.0, 2mM EDTA, 1% Triton X-100), twice with high salt buffer (0.1% SDS, 500mM NaCl, 20mM Tris-HCl-pH8.0, 2mM EDTA-pH8.0, 1% Triton X-100), and twice with TE buffer (10mM Tris-HCl-pH8.0, 1mM EDTA-pH8.0). Chromatin was eluted by incubation with 200μL elution buffer at room temperature for 15 min and 37°C for 10 min on a 500rpm thermomixer. Eluted chromatin was reverse-crosslinked and treated with RNaseA and Proteinase K as described above. DNA was purified and quantified by Qubit (Thermo Fisher Scientific). About 1ng DNA was used for library construction using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB). Sequencing (2 × 150bp) was performed on Illumina NovaSeq 6000 platform at Novogene. Library information is available in Table S4.

RNA-seq library preparation and sequencing

The total RNA-seq libraries were constructed using the SMART-Seq Stranded Kit (Takara) as previously described.[18] For each library, 1μL 1/10000 ERCC RNA spike-in Mix was included prior to RNA shearing (Thermo Fisher Scientific). Sequencing (2 × 150bp) was performed on the Illumina NovaSeq 6000 platform at Novogene. Library information is available in Table S4.

ChIP-seq data processing

For ChIP-seq data, the raw read pairs were trimmed by Trimgalore (v0.6.6)(https://github.com/FelixKrueger/TrimGalore) with the option "–paired", and aligned to mm10 by bowtie2 (v2.4.2)[41] with the options "-p 6 –no-unal –no-mixed –no-discordant -I 0, -X 1000". Under this setting, one alignment was randomly selected for read pairs that were aligned to multiple genomic locations. PCR duplicates were removed using Picard (v2.18.22)(http://broadinstitute.github.io/picard/), and Macs2 (v2.1.4)[42] was used to call peaks with the option "-f BAMPE -g mm -B -q 0.0001 –nolambda –nomodel". For each antibody (i.e., anti-Cas9 and anti-HA), only peaks that were present in the "dCas9+sgRNA" group, but not the "dCas9-only", were kept for downstream analyses. Overlaps between dCas9-ChIP and HA-ChIP were identified using bedtools (v2.27.0)[43] intersect function and considered as high confidence dCas9

target sites (n = 4040). Bigwig files were generated using deeptools (v2.0.0)[44] bamCoverage function with the options "–binSize 25 –normalizeUsing RPKM –scalFactor 1".

The dCas9/HA ChIP peaks were intersected to mm10 repeats annotations to identify peaks that overlapped with MT2_Mm and MT2C_Mm insertions. The pre-generated transposable elements (TE) GTF file "mm10_rmsk_TE.gtf.gz" was retrieved from the TEtranscripts[45] FTP site. The peaks were annotated using "makeTxDbFromGFF" and "annotatePeak" function from the "ChIPpeakAnno", "GenomicFeatures", and "ChIPseeker" Bioconductor R packages.[46–48] The gene annotation file was downloaded from GENCODE M25 (https://www.gencodegenes.org). The deeptools (v2.0.0)[44] "compueMatrix" and "plotHeatmap" functions were used to generate heatmaps illustrating dCas9/HA enrichment at MT2_Mm and MT2C_Mm insertions.

RNA data processing

Raw RNA-seq read pairs were trimmed by Trimgalore (v0.6.6)(https://github.com/FelixKrueger/TrimGalore) with the options "–paired –clip_R2 5", and aligned to mm10 by STAR (v2.7.9)[49] with the options "–outFilterMultimapNmax 100 –winAnchorMultimapNmax 100 –outSAMattributes NH HI NM MD XS AS –alignIntronMax 1000000 –alignMatesGapMax 1000000 –limitBAMsortRAM 30000000000 –outFilterType BySJout". Following STAR mapping, TEtranscripts (v2.2.3)[45] was used to count uniquely and multiple aligned read pairs for each repeat subfamily. For genes, only uniquely mapped read pairs were counted. Uniquely aligned read pairs were further identified by sambamba (v0.6.8)[50] with the option "-F mapping_quality $\geq$255", and then supplied for StringTie (v2.1.7)[51] to generate FPKM values for each gene. Bigwig files were generated using deeptools (v2.0.0)[44] bamCoverage function with the options "–binSize 25 –normalizeUsing RPKM –scalFactor 1" using uniquely mapped read pairs. RNA-seq data were also aligned to a custom dCas9-KRAB and ERCC reference by bowtie2 (v2.4.2)[41] to identify read pairs from dCas9-KRAB mRNA and ERCC RNA mixes.

The DESeq2 (v1.38.3)[52] was used for differential expression analyses between different groups. For comparison between the non-injected group and CTRi, library size factors were estimated using the "estimateSizeFactors" function with ERCC read counts as the "controlGenes". This comparison is to ensure that global transcriptional level was not affected by overexpression of dCas9-KRAB mRNA alone. For other comparisons, default DESeq2 settings were used. In addition to "FDR <0.05", fold change >2 were used to define differentially expressed genes and repeat subfamilies. GO terms were analyzed using "enrichGO" function from Bioconductor R package "clusterProfiler" (v4.6.2). The nearest MT2, solo MT2 and MERVL to each gene were identified using bedtools (v2.27.0)[43] closest function. For the analyses in Figure 3H, a gene is defined to be associated with solo MT2 or MERVL if they are within 30kb to each other. The deeptools (v2.0.0)[44] computeMatrix and plotHeatmap functions were used to generate heatmaps describing MT2 single copy expression dynamics. Mappability was computed using GenMap (v1.2.0)[54] with the option "-K 50 -E 3".

Define ZGA and maternal decay genes

The cutoffs used to define minor/major ZGA genes and maternal decay genes were described previously.[26] The total RNA-seq data of MII eggs, L1C, E2C, and L2C were from two previous reports.[18,19] The reason why gene numbers for each category were different from the previous report[26] could be due to different reference genomes (mm9 vs. mm10) and gene annotations (UCSC vs. Gencode) being used.

MT2 proximal/distal elements analyses

Regtools (v0.5.2)[55] was used to identify and annotate spliced RNA-seq reads that overlap MT2/MERVL. To be comprehensive, polyA RNA-seq[17] and total RNA-seq of the L2C (this study) were combined and supplied for Regtools analyses. The sequence motifs enriched at the splicing junctions were identified using ggseqlogo (v0.1).[56] Of the genes that were downregulated in E2C and/or L2C by MT2i (n = 1155), 717 genes are <30kb to nearest MT2. Of the 717 genes, 416 genes are proximal to annotated transcription start sites (−1.5kb to +0.5kb) and/or have chimeric splicing junctions detected by Regtools. Other genes are far away from known promoters and do not form chimeric splicing junctions with MT2/MERVL.

Public datasets used in this study

The poly-A and total RNA-seq datasets used in Figures 1A, S1A, and S1B were from.[17–19] The MERVL ASO KD and MERVLi RNA-seq datasets analyzed in Figures 2, 3, and S2 were from.[6] The *Obox/Dux* RNAseq datasets (Figures 4 and S4) were from.[26,32,33] All RNA-seq datasets were processed similarly as described in the "RNA data processing" section. The H3K4me3, ATAC-seq, H3K27ac, OBOX1/3/5, and DUX DNA binding datasets (Figures 4 and S4) were from.[17,18,26,27,60] All these datasets were processed similarly as described in the "ChIP-seq data processing" section.

Visualization and statistical analysis

All statistical analyses were performed using the R programming language. All genomic browser tracks were viewed using the UCSC genome browser.[57]

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analyses details such as statistical tests and sample sizes are indicated in the figure legends. Details of statistical tests of next generation sequencing datasets are described in the method details sections including differential gene expression analyses, GO term analyses, and ChIP-seq/ATAC-seq enrichment analyses. All statistical analyses were performed with R (v4.2.2) scripts on the R-studio platform (v1.4.1).