7314–7329 *Nucleic Acids Research, 2023, Vol. 51, No. 14* *Published online 3 July 2023*
https://doi.org/10.1093/nar/gkad540

ZMYM2 is essential for methylation of germline genes and active transposons in embryonic development

Adda-Lee Graham-Paquin[1,2,†], Deepak Saini[1,†], Jacinthe Sirois[1,2], Ishtiaque Hossain[1], Megan S. Katz[1,2], Qinwei Kim-Wee Zhuang[3,4], Sin Young Kwon[1], Yojiro Yamanaka[2,3], Guillaume Bourque[3,4,5], Maxime Bouchard[1,2,‡] and William A. Pastor[1,2,*]

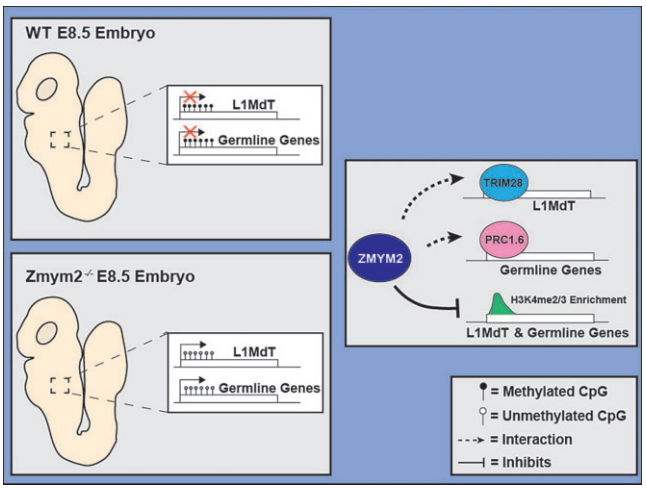
[1]Department of Biochemistry, McGill University, Montreal, Quebec, Canada, [2]The Rosalind & Morris Goodman Cancer Institute, McGill University, Montreal, Quebec, Canada, [3]Department of Human Genetics, McGill University, Montreal, Quebec, Canada, [4]Institute for the Advanced Study of Human Biology (WPI-ASHBi), Kyoto, Japan and [5]Canadian Center for Computational Genomics,McGill University, Montreal, Quebec, Canada

Received December 06, 2022; Revised May 24, 2023; Editorial Decision June 06, 2023; Accepted June 09, 2023

ABSTRACT

ZMYM2 is a transcriptional repressor whose role in development is largely unexplored. We found that *Zmym2*$^{-/-}$ mice show embryonic lethality by E10.5. Molecular characterization of *Zmym2*$^{-/-}$ embryos revealed two distinct defects. First, they fail to undergo DNA methylation and silencing of germline gene promoters, resulting in widespread upregulation of germline genes. Second, they fail to methylate and silence the evolutionarily youngest and most active LINE element subclasses in mice. *Zmym2*$^{-/-}$ embryos show ubiquitous overexpression of LINE-1 protein as well as aberrant expression of transposon-gene fusion transcripts. ZMYM2 homes to sites of PRC1.6 and TRIM28 complex binding, mediating repression of germline genes and transposons respectively. In the absence of ZMYM2, hypermethylation of histone 3 lysine 4 occurs at target sites, creating a chromatin landscape unfavourable for establishment of DNA methylation. *ZMYM2*$^{-/-}$ human embryonic stem cells also show aberrant upregulation and demethylation of young LINE elements, indicating a conserved role in repression of active transposons. ZMYM2 is thus an important new factor in DNA methylation patterning in early embryonic development.

GRAPHICAL ABSTRACT



INTRODUCTION

Methylation of the 5-position of the DNA base cytosine is the most common epigenetic modification of DNA in mammals, typically occurring at the dinucleotide CpG. DNA methylation promotes heterochromatinization, and a high density of methylated CpGs at a gene promoter has a strong silencing effect (1). DNA methylation is largely lost from the genome during pre-implantation mammalian development and is then re-established globally during the peri-implantation period (2). Although methylation patterns continue to change during development, this global reprogramming in the first days of life is far more extensive than anything that occurs subsequently in somatic tissue (3).

Most CpG-rich transcript promoters remain unmethylated even after global methylation establishment, with two categorical exceptions: germline genes and transposons.

*To whom correspondence should be addressed. Tel: +1 514 398 8350; Email: william.pastor@mcgill.ca
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
‡Posthumous.

Genes selectively expressed in germ cells (germline genes) are silenced by promoter methylation and only reactivated in the developing germline, where DNA methylation is lost globally (4,5). These genes are targeted for silencing and subsequent methylation by the non-canonical polycomb repressor complex PRC1.6, which recognizes E2F and E-Box motifs at their promoters (6,7). Transposons are also targeted for extensive methylation. During the pre- and peri-implantation period, many transposons are silenced by the TRIM28 complex, which is recruited by hundreds of distinct KRAB-Zinc finger proteins that recognize distinct sequences present on transposons (8). TRIM28 silences target transposons by recruiting histone deacetylases and the histone 3 lysine 9 (H3K9) methyltransferase SETDB1 (8). The TRIM28 complex is essential during development, as *Trim28*$^{-/-}$ mice show severe abnormality by embryonic day (E)5.5 (9). Additionally, TRIM28 promotes transposon DNA methylation (10,11), and eventually DNA methylation replaces TRIM28 as the critical factor for silencing transposons (9,12,13). Mice deficient for the *de novo* methyltransferases DNMT3A and DNMT3B show aberrant expression of germline genes and transposons along with severe growth restriction and abnormality by E8.5 (13).

The transcription factor Zinc-finger MYM-type protein 2 (ZMYM2) has been implicated in transcriptional repression in a variety of cell types (14–16) and recruitment of ZMYM2 silences reporter constructs (14,15). Its MYM-type Zinc fingers are reported not to bind DNA and they lack most of the DNA interaction surface present on other Zinc fingers (17). Instead, ZMYM2 can bind the post-translational modification SUMO2 via distinct SUMO Interaction motifs (SIMs) (14) and the MYM-Zinc fingers themselves are reported to facilitate protein-protein interaction (17,18). Many transcription factors and chromatin-bound complexes are SUMOylated, with SUMOylation often functioning to reduce transcriptional activity (19,20), and ectopically expressed ZMYM2 homes to regions of SUMOylation (14). Additionally, ZMYM2 is known to interact with the CoREST complex (15–17,21,22). The CoREST complex includes the histone deacetylases HDAC1 and HDAC2 and the histone demethylase LSD1 (21), and loss of ZMYM2 results in a global reduction of association of CoREST with chromatin (16,17). ZMYM2 can thus home to SUMOylated complexes on chromatin and recruit CoREST to mediate transcriptional silencing.

There is evidence that ZMYM2 mediates transcriptional repression in developmentally relevant cell types. ZMYM2 was identified as a factor important for suppressing totipotency in murine embryonic stem cells (mESCs) and zygotic depletion of *Zmym2* RNA reduces efficiency of blastocyst formation (16). ZMYM2 was identified in a CRISPR screen as a gene that facilitates exit from pre-implantation-like 'naïve' pluripotency in mESCs (23) and has been implicated in transposon silencing in mESCs (16,24). In subsequent CRISPR screens, ZMYM2 was found to restrict growth in human embryonic stem cells (hESCs) (25) and its deletion results in reactivation of the silenced copy of imprinted genes in both mESCs and hESCs (26,27). Heterozygous mutations of ZMYM2 in humans cause craniofacial abnormalities and congenital anomalies of the kidney and urinary tract (CAKUT) and *Zmym2*$^{+/-}$ mice show elevated rates of CAKUT (15). Nonetheless, its importance in mammalian embryonic development, and the molecular mechanisms underlying its role in development, remain largely unknown. To determine the developmental role of ZMYM2, we generated *Zmym2*$^{-/-}$ mice and found a striking role for ZMYM2 in facilitating the methylation of germline genes and young LINE transposons.

MATERIALS AND METHODS

Mice

Parental *Zmym2*$^{+/-}$ mice were generated by the Transgenic Core Facility of the Goodman Cancer Institute in a C57BL/6 background using a CRISPR-Cas9 targeting approach as described(15). Timed mating was performed, and presence of a copulatory plug was identified at E0.5. Animals and experiments were kept in accordance with the standards of the animal ethics committee of McGill University, and the guidelines of the Canadian Council on Animal Care.

Embryo collection and processing

All embryos were dissected in cold PBS and fixed for 20 minutes in 4% paraformaldehyde at 25°C. Following imaging of embryos with a Zeiss Lumar V12 stereomicroscope, samples used for tissue analysis were processed for either cryosection or paraffin embedding. Cryo-embedded samples were flash frozen in O.C.T. compound and sectioned to obtain 8 to 10μm thick sections as described (28). Paraffin-embedded samples were processed and embedded by the GCI Histology Core Facility and 6μm serial sections were obtained.

Immunofluorescence staining of tissue sections

Immunofluorescence analyses were performed as described (29). Anti-LINE-1 ORFp (Abcam ab216324) was used at 1:100 dilution in PBS.

Embryonic stem cell culture conditions

Both wild-type control and *Zmym2*$^{-/-}$ mESCs were generated from the V6.5 ESC line (Novus Biologicals NBP1-41162) derived from C57BL/6 × 129/sv cross. mESCs were cultured on 0.1% gelatin (EmbryoMax, ES-006-B) pre-coated plates and media was refreshed every day. Cells were passaged every 2–3 days using 0.05% Trypsin-EDTA (Gibco, 25300062). These cells were cultured in Knockout™ D-MEM (Gibco, 10829018) media supplemented with 1000 U/ml ESGRO Recombinant Mouse LIF (Millipore Sigma, ESG1107), 2 mM GlutaMAX (Gibco, 35050061), 100 μg/ml Primocin (InvivoGen, ant-pm-2), 0.1 mM 2-mercaptoethanol and 15% ES-qualified fetal bovine serum (Gibco, 10439024). All mESCs were mycoplasma-free and were periodically tested for mycoplasma contamination.

Immunofluorescence staining of mESCs

mESCs were cultured on glass coverslips coated with 0.1% gelatin. Cells were then fixed with 4% paraformaldehyde for

15 min at room temperature and washed twice with 1× PBS. The mESCs were then permeabilized and blocked using 0.1% Triton-X diluted in a 5% solution of donkey serum for 15 min at room temperature and washed twice with 1× PBS-T (0.1% Tween20 in 1× PBS). Primary antibody (ZMYM2, Invitrogen PA5-83208 at 1:1000 dilution) was added to the cells and incubated overnight at 4°C. After overnight incubation, cells were washed twice with 1× PBS-T and secondary antibody and DAPI fluorophores were added (Donkey anti-Rabbit Alexa Fluor™ 488, A-21206 as a 1:500 dilution) for 45 min at room temperature. Cells were then washed twice with 1× PBS-T and coverslips were mounted on microscopy slides using ProLong Gold (ThermoFisher, P36930). Cells were imaged on the LSM710 confocal microscope.

RNA isolation, quantitative PCR and RNA-sequencing of embryos

qPCR: Total mRNA was extracted from dissected embryos using a RNeasy micro kit (Qiagen, 74084) or AllPrep DNA/RNA Micro Kit (Qiagen, 82084). mRNA was reverse transcribed with MMLV (Invitrogen) according to manufacturer's procedures. Real-time quantitative PCR was performed using TransStart® Tip Green qPCR SuperMix (AQ141-01) on Realplex2 Mastercycler (Eppendorf).

RNA sequencing: Total RNA was isolated and sequenced from whole E8.5 embryos. Sequencing libraries were prepared by Genome Quebec Innovation Centre (Montreal, Canada), using the NEB mRNA stranded library preparation. cDNA libraries were sequenced using the Illumina NovaSeq 6000 S2 sequencer, 100 nucleotide paired-end reads, generating 40–75 million reads per sample.

RNA-sequencing analysis

RNA-seq fastq reads were trimmed using Trimmomatic(v0.34) to remove low-quality bases and remove adapters. Trimmed reads were then aligned to the mm10 reference genome using STAR (v2.7.8a) with default parameters to generate .bam files. Picard (v2.9.0) was then used to sort and mark duplicates from the aligned bam files. Raw and normalized reads were then quantified using HTSeq-count and the StringTie suite (v1.3.5). The raw read counts were then used for differential expression analysis using DESeq2 and genes showing a minimum fold change of 4 and FDR (q-value) of <0.05 were selected as significant differentially expressed genes. Gene ontology over-representation was determined using clusterProfiler for genes displaying a minimum fold change of 4. Redundant GO-terms were removed using the simplify function in clusterProfiler. Gene set enrichment analysis (GSEA) was then performed using the normalized read counts measured from the DESeq2 software. GSEA was performed using default parameters (1000 permutations) with the mouse hallmarks and biological process gene sets. To measure differences in *Zmym2*, *Dnmt3a* and *Dnmt3b* expression during murine embryonic development, RNA-seq from Smith 2017 was processed as above. Normalized RPKM counts were generated using the StringTie suite. RPKM counts (8-cell to E6.5 from Smith 2017 and E8.5 from this study) of *Zmym2*, *Dnmt3a* and *Dnmt3b* were plotted as line graphs across developmental timepoints.

To assess expression of *ZMYM2*, *DNMT3A* and *DNMT3B* in cynomolgus macaques, we averaged processed RPM values from GSE74767 for each indicated cell type, using cell identities indicated in Nakamura 2016.

The TEtranscript pipeline was used to determine differential expression of transposon classes in *Zmym2*$^{-/-}$ E8.5 embryos. RNA-seq reads were first trimmed using Trim Galore! (v0.6.6) with default parameters. Trimmed reads were then mapped to the mm10 genome using STAR aligner supporting multi-alignments per read (–winAnchorMultimapNmax 200, –outFilterMultimapNmax 100 and a curated GTF file from the TEtranscript website). The resulting BAM files were used as inputs for the TEtranscript pipeline using default parameters and curated annotation refGene and repeatMasker files from the TEtranscript website. DESeq2 was then used to calculate differentially expressed transposon classes in *Zmym2*$^{-/-}$ embryos.

Genetic ablation of Zmym2 by CRISPR-cas9 nucleofections

Zmym2$^{-/-}$ mESCs on a V6.5 background were generated using the same sgRNA sequence used to generate the *Zmym2*$^{-}$ murine allele: AATGTTACAACCTTAGAAAC. CRISPR-Cas9 and sgRNA were delivered using the Lonza Biosciences 4D-Nucleofector to electroporate ribonuclear particles into cells. 300K cells, prepared for nucleofection according to manufacturer's instructions, were nucleofected using the CA-137 protocol, and were hastily passaged to pre-warmed serum + LIF media after nucleofection. Clonal lines were generated by picking and expanding individual mESC colonies. Clonal mESC lines were then validated using PCR and Sanger sequencing, western blotting and immunofluorescence. Control lines were generated by nucleofection with a non-targeting sgRNA and were otherwise treated identically to *Zmym2*$^{-/-}$ lines.

Zmym2$^{-/-}$ mESCs in a 129/Sv background (CCE line) were donated by the lab of Jianlong Wang (16).

Reverse-transcriptase quantitative PCR

According to MIQE guidelines, total mRNA was derived from 3 clonal non-target and 3 clonal *Zmym2*$^{-/-}$ mESC lines cultured in serum + LIF and Day 3/6 embryoid bodies (EBs). Additional validating experiment setup were done with another J1 WT and *Zmym2*$^{-/-}$ mESC line (Supplemental Figure S9D) and a C2 (C57BI6) WT and *Zmym2*$^{-/-}$ lines (data not shown). All mESC samples were harvested using 0.05% Trypsin-EDTA and were neutralized with Trypsin inhibitor. Cell pellets were washed with PBS, flash frozen in liquid nitrogen and stored at –80°C for one week. Total RNA was extracted using RNAzol RT (Molecular Research Center Inc, RN 190). mRNA was precipitated with 75% ethanol and resuspended in DNase/RNase-free H_2O. All RNA work was done on a designated RNA bench cleaned with RNase Away (Fisher, 1437535), with designated pipettes and filtered tips. RNA samples were then stored at –20°C for 3 weeks before downstream analysis.

RNA was quantified using the Nanodrop 1000 UV instrument using 1.5 μl of RNA sample and DNase/RNase-free H_2O as a blank. All $A_{260/280}$ ratio were between 1.8–2.1 and the RNA yield ranged from 86 ng/μl to 1658 ng/μl. Isolated RNA samples were then used to generate cDNA with the SensiFast cDNA Synthesis Kit (FroggaBio, DD-BIO-65053). 200 ng (12 μl) of RNA from each sample was mixed with 4 μl of 5× TransAmp buffer, 1 μl of reverse transcriptase enzyme, and 3 μl of DNase/RNase-free H_2O resulting in a final concentration of 10 ng/μl of cDNA for all samples. The cDNA mastermix was then added to a thermocycler with the following conditions: 25°C (10 min), 45°C (15 min), 85°C (5 min), 4°C overnight.

Quantitative PCR was then performed using 5 ng of cDNA for each individual sample, 2× PowerUP SYBR Green PCR (Invitrogen, A25742), and primers at concentrations of 0.5 μM. PCR reactions totaled to a volume of 7 μl. Each sample was set-up in triplicates in 384-well plates from Life technologies (MicroAmp EnduraPlate Optical 384-well plate, Cat. 4483285) using an electronic multichannel pipette. The following cycling conditions were used on the Quantstudio5: 50°C 2 min, 95°C 2 min, 45× (95°C 15 s, 60°C 1 min), 95°C 15 s, 60°C 1 min).

Quantitative PCR data analysis was performed on Excel and the QuantStudio5 software (applied BioSystems/ThermoFisher Scientific). $2^{-\Delta\Delta Ct}$ was used to calculate fold change using GAPDH as a reference housekeeping gene. Values were calculated relative to mESC WT (Day 0 of differentiation). Outliers were identified manually (Cq variation over 0.3 from average, divergence of the melting curve, irregular amplification).

qPCR primers were validated using primer efficiencies tested from gradient concentration of cDNA (cDNA mix of WT, Zmym2 KO cDNA from both mESC and EBs). Specificity of qPCR reactions were verified using melting curve analysis (verifying single narrow peaks) during efficiency testing. List of primer sequences, efficiencies, LODs, accession numbers, primer location, amplicon length and targeted variants are located in Supplementary Table S3.

Western blot

Protein was extracted using ice cold 1× RIPA lysis buffer supplemented with fresh protease inhibitors (1 mM phenylmethylsulfonyl fluoride, 10 mM sodium fluoride and 1 mM sodium orthovanadate). Cell pellets were then subjected to 5 cycles of freeze-thawing using liquid nitrogen to ensure complete breakage of both the cell and nuclear membranes. Lysate protein concentration was measured using a Bradford Assay and 30–40 μg of protein was run on a 6–12.5% gradient SDS-PAGE. The resolved proteins were transferred to a polyvinylidene fluoride (PVDF) membrane, which was blocked using 5 mL of LI-COR Odyssey Blocking Buffer for 1h at room temperature. The membrane was then incubated overnight with primary antibody in 1× Odyssey Blocking Buffer 0.15% Tween-20 at 4°C. Anti-ZMYM2 (ThermoFisher PA5-83208) was used at a 1:1000 dilution and anti-H3 (Abcam ab1791) was used at a 1:10 000 dilution. The membrane was then washed 3 × 5 min in PBS supplemented 0.1% Tween-20 and then incubated with secondary antibodies (LI-COR IRDye 680RD, 1:20 000 dilution) diluted in Odyssey Blocking Buffer and 0.15% Tween-20 for 1 h at room temperature. Membranes were then washed for 5 min twice in PBS 0.1% Tween-20 and kept in PBS before imaging on the LI-COR imaging system.

Whole-genome bisulfite library preparation

WGBS libraries were prepared from 500 ng of genomic DNA. Isolated DNA (500 ng) from E8.5 mouse embryos (QIAamp DNA mini kit, 51304 or AllPrep DNA/RNA Micro Kit, Qiagen, 82084) and spiked-in lambda DNA (1.25 ng) (New England Biolabs) was sheared using the M220 ultrasound sonicator (Covaris) to an average genomic size of 350 bp. 200 ng of sheared DNA was bisulfite-converted using the EZ DNA Methylation-Gold Kit (Zymo Research, D5005) using manufacturer's instructions. Up to 100 ng of genomic DNA was then used to generate WGBS libraries using the Accel-NGS Methyl-Seq DNA Library prep kit (Swift Biosciences, 30024) according to manufacturer's instructions and 13 cycles of PCR for final library amplification. Final concentration of libraries were determined using the Qubit 1X dsDNA High Sensitivity Assay Kit (Invitrogen, Q33230) and final library sizes were assessed by agarose gel electrophoresis. Libraries were then sequenced in paired-end using the NovaSeq 6000 at the Center for Applied Genomics operated by the SickKids Research Institute.

Whole-genome bisulfite sequencing analysis

WGBS reads were trimmed to remove low-quality bases and the first five bases of R1 and ten bases of R2 were removed using Trim Galore! v(0.6.6) software (parameters: -q 20, –clip_R1 5, –clip_R2 10). Reads were then aligned to the mm10 reference genome using Bismark (v0.22.3) with default parameters. Aligned reads were then deduplicated and filtered for incomplete bisulfite conversion. Methylation calling over cytosines was done using Bismark and overall DNA methylation was calculated as a mean of all cytosine bases in CpG context. Metaplots analysis of CpG methylation over gene bodies or promoter regions were generated by calculating the percentage of CpG methylation of each RefSeq gene and 3 kb flanking regions. Similarly, CpG methylation was also measured across 22 loci of genomics imprints(30). The percentage of CpG methylation was visualized as a heatmap. To determine statistical significance of methylation changes across imprint loci between wild-type and *Zmym2*$^{-/-}$ KO embryos, a two-tailed homoscedastic student's T-test was performed, using each individual mouse as a replicate and was corrected for multiple hypothesis testing using the Bonferonni correction. No imprints showed statistically significant hypomethylation.

Differentially methylated regions across the whole genome were identified using the methylKit R package of 1000 bp tiled regions with a minimum of 10 read coverage, a difference of >25% methylation and a q-value of <0.05. To determine statistically significant DMRs using methylKit, a logistic regression model was used to determine P-value and was corrected using a sliding linear model to determine the false discovery rate.

ChIP of histone modifications

Both wild-type and *Zmym2*$^{-/-}$ mESCs were crosslinked in 1% paraformaldehyde for 10 min and quenched with the addition of 1 mM of glycine for 10 min to stop crosslinking. Cells were then lysed using lysis buffer solutions and sonicated with the M220 ultrasonicator (Covaris) in 1 ml tube with an AFA Fiber and the following conditions (cycles/burst = 200, duty factor = 20%, peak intensity power = 75, time = 10 min and temperature = 7°C). The sheared DNA was then pre-cleared using magnetic beads (Sera-Mag Protein A/G SpeedBeads, VWR 17152104010150) for 1 h to remove non-specific binding. Using a magnetic rack, the sonicated lysate was then separated to a new tube and primary antibodies were added (anti-H2AK119Ub Cell Signaling D27C, anti-H3K4me3 Cell Signaling CD42D8, anti-H3K4me2 Cell Signaling C64G9) and incubated overnight. Magnetic beads were then washed and added to the sonicated lysate and incubated for 2 h at 4°C to allow for the beads to bind to the antibody/protein/DNA complex. Beads were then washed with buffer of increasing salt concentration (Wash Buffer A: 50 mM HEPES, 1% Triton X-100, 0.1% deoxycholate, 1 mM EDTA, 140 mM NaCl and Wash Buffer B: 50 mM HEPES, 0.1% SDS, 1% Triton X-100, 0.1% deoxycholate, 1 mM EDTA, 500 mM NaCl) and TE buffer. Protein/DNA complexes of interest were finally eluted using elution buffer (50 mM Tris–HCl, 1 mM EDTA, 1% SDS) incubated at 65°C for 10 min and separated from the beads using a magnetic rack. Samples were then de-crosslinked by incubating overnight at 65°C. Residual RNA and protein were removed with the addition of RNaseA and proteinase K. DNA was then purified using the Qiagen MinElute PCR Purification Kit using manufacturer's instructions.

CUT&RUN of L3MBTL2

To determine the localization of L3MBTL2 in Zmym2$^{+/+}$ and Zmym2$^{-/-}$ mESCs (CCE line from Yang and colleagues, see above), CUT&RUN was used using manufacturer's instructions (EpiCypher CUTANA ChIC/CUT&RUN Kit, 14–1048). The eluted DNA from the CUT&RUN or ChIP was then used to construct multiplex libraries with the NEBNext Ultra II DNA library Kit using manufacturer's instructions.

ChIP-seq and CUT&RUN data analysis

Reads were trimmed using Trimmomatic (v0.6.6) using default parameters and then aligned to the mm10 genome using BWA (v0.7.17). Removal of PCR duplicates was then done using Picard (v2.0.1). To account for repetitive sequences found in transposable elements, BAMs were marked for duplication but were not filtered for unique reads. MACS2 was then used for peak identification using default parameters and inputs as controls. To determine the significance of detected peaks, MACS2 used a Poisson distribution model to identify peaks and were corrected for false discovery rate using Benjamini-Hochberg correction. Aligned BAM files were then used to generate the bigwigs using DeepTools. The resulting bigwigs were used for metaplot and heatmaps using the computeMatrix and plotHeatmap functions in DeepTools. To normalize metaplots across samples, ChIP signals were normalized by (i) measuring the median signal across the whole genome, (ii) measuring the max height of signal across regions of interest genome-wide (H3K4me2/3 signal across all gene promoters and H3K9me3 signal across all H3K9me3 peaks derived from Barral *et al.* 2022). (iii) The scaling factor per sample was then calculated by subtracting the max height (from (ii))—median signal (from (i)). (iv) Sample specific ChIP signals were then normalized across the regions of interests by the following equation: (signal height – median)/scaling factor.

RESULTS

Zmym2$^{-/-}$ mouse embryos show developmental abnormality and early lethality

The *Zmym2*$^{-}$ murine allele corresponds to a mutation in the first coding exon, resulting in an early frameshift (Supplementary Figure S1A). *Zmym2*$^{-/-}$ embryos were found at Mendelian ratios until E9.5, but by E10.5 resorptions were observed and no *Zmym2*$^{-/-}$ embryos were recovered, demonstrating embryonic lethality (Figure 1A). A variable *Zmym2*$^{-/-}$ phenotype was observed at E9.5, with most embryos showing gross phenotypic abnormality, including a failure to undergo turning of the embryonic trunk (Figure 1B, C). A statistically significant reduction in length, size and somite count was also observed (Figure 1D, Supplementary Figure S1B, C).

Most E8.5 *Zmym2*$^{-/-}$ embryos appeared normal (Supplementary Figure S1D), so the molecular phenotype of *Zmym2*$^{-/-}$ was analyzed using E8.5 embryos with no visible abnormalities.

Zmym2 suppresses expression of germline genes and young LINE elements

To determine the transcriptional changes that occur in *Zmym2*$^{-/-}$ embryos prior to embryonic lethality, we performed RNA-sequencing of six *Zmym2*$^{+/+}$ and six *Zmym2*$^{-/-}$ whole E8.5 embryos, three males and three females of each genotype (Supplementary Tables S1 and S2). Upon the loss of ZMYM2, knockout embryos exhibited an upregulation of 165 genes (fold change ≥ 4, q-value < 0.05) (Figure 2A), and downregulation of only three genes, consistent with a repressive role for ZMYM2. GO term analysis showed that the upregulated genes were enriched for terms related to germ cell development (Figure 2B), and GSEA analysis confirmed upregulation of genes involved in meiosis and spermatogenesis (Supplementary Figure S2A, B). Upregulation of select genes was confirmed through RT-qPCR of additional *Zmym2*$^{-/-}$ and *Zmym2*$^{+/+}$ E8.5 embryos (Supplementary Figure S2C, Supplementary Table S3).

We observed two classes of upregulated genes in *Zmym2*$^{-/-}$ embryos. Ninety-five genes showed upregulation of transcription from an annotated promoter (Figure 2C). By contrast, 60 genes showed clear evidence of splicing from an internal or upstream transposon (Figure 2D, annotations in Supplementary Table S4). In 46 of the 60 fusion transcripts, the transposon in question was a LINE

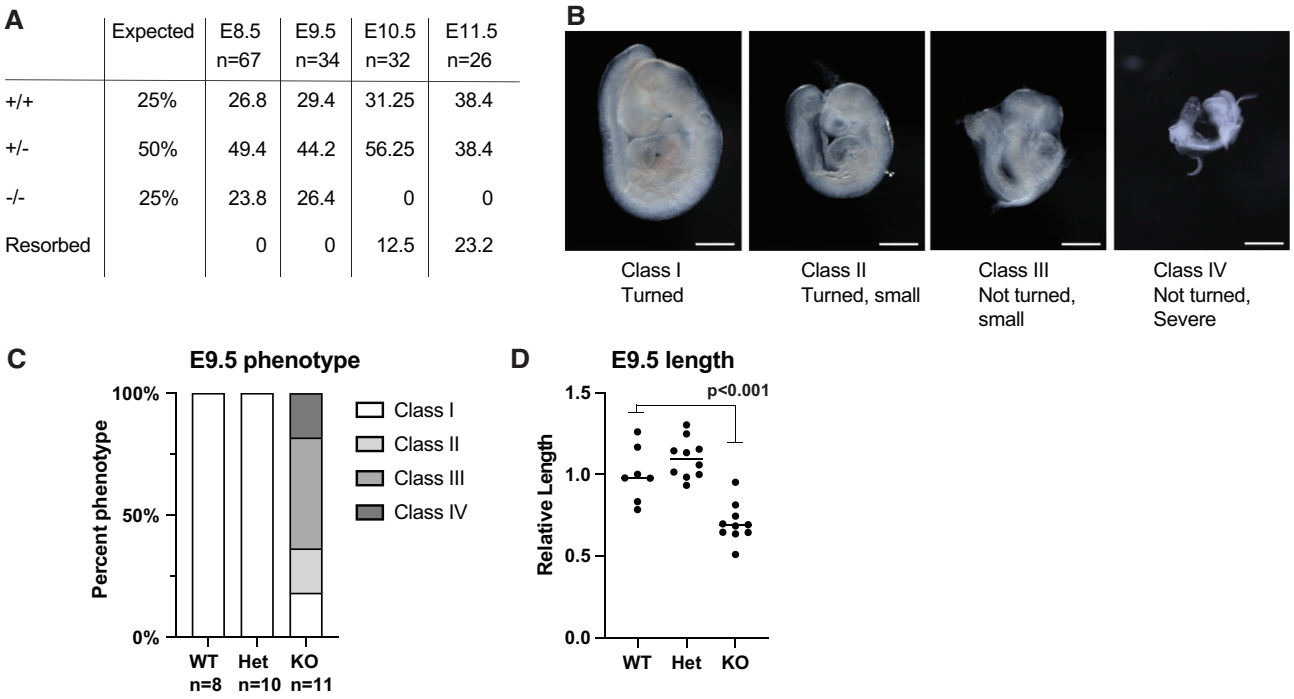


Figure 1. *Zmym2*$^{-/-}$ embryos show lethality by E10.5 and variable developmental delays at E9.5. (**A**) Embryonic viability by genotype from embryonic day E8.5 to E11.5 compared to the expected Mendelian ratios from *Zmym2*$^{+/-}$; *Zmym2*$^{+/-}$ crosses. Each embryonic age includes pooled data from 3+ litters. (**B**) Images taken of E9.5 *Zmym2*$^{-/-}$ embryos representing four different classifications of phenotype severity. Scale bars = 500 μm. (**C**) Distribution of E9.5 phenotype classes by genotype. (**D**) Relative length of embryos (crown to tail) normalized by litter as a ratio to average *Zmym2*$^{+/+}$ length. Significance from one sided t-test is indicated.

element of the L1MdT subclass. Both sense and antisense transcription from LINE promoters seeded transcription of fusion genes (Figure 2D, Supplementary Figures S3A–C). Interestingly, published data(31) indicates that some of these LINE-gene fusion transcripts are expressed at relatively high levels in early development and suppressed between E6.5 and E8.5 in wild-type mice (Supplementary Figure S3D–F).

We analyzed transposon expression in *Zmym2*$^{-/-}$ using the TEtranscript pipeline (32) and observed strong upregulation of L1MdT and the less abundant L1MdGf subclasses (Figure 2E). In any given species, the youngest transposons classes are typically the most active and capable of transposition, and it is notable that L1MdT and L1MdGf are the youngest and third youngest LINEs in *Mus musculus* respectively (33). Immunofluorescence staining of cross-sectional slices of E8.5 and E9.5 *Zmym2*$^{-/-}$ embryos showed high expression of the LINE-1 ORF1 protein (L1ORF1p) in all cells, in striking contrast with *Zmym2*$^{+/+}$ controls (Figure 2F, Supplementary Figure S3G). L1ORF1p is also highly expressed in *Zmym2*$^{-/-}$ trophoblast giant cells, indicating that ZMYM2 is also important for suppressing transposons in the placental lineage (Supplementary Figure S3H).

Zmym2 is required for DNA methylation of germline gene promoters and LINE elements

As germline genes and transposons are known to be silenced by DNA methylation in the somatic lineage, we conducted whole genome bisulfite sequencing (WGBS) of nine *Zmym2*$^{+/+}$ and nine *Zmym2*$^{-/-}$ embryos. We targeted > 30-fold coverage for male and female, *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ (Supplementary Table S1). Two embryos, one of each genotype and originating from the same litter, were excluded from subsequent analysis because they showed abnormally low levels of global DNA methylation for E8.5 (Figure 3A).

We observed no defect in global DNA methylation establishment in *Zmym2*$^{-/-}$ embryos (Figure 3A). However, we observed a substantial reduction in average DNA methylation over the promoters of the 95 upregulated genes expressed from annotated TSS (Figures 2C and 3B, C). Hypomethylation was also observed globally over the promoters of full length L1MdT elements (Figures 2D and 3D, Supplementary Figures S3A–C). We compared gene expression changes in the *Zmym2*$^{-/-}$ embryos to published data(13) of E8.5 *Dnmt3a*$^{-/-}$/*3b*$^{-/-}$ mice, in which all *de novo* methylation is impaired. Genes upregulated in *Zmym2*$^{-/-}$ and expressed from annotated TSS showed a globally similar degree of upregulation in *Dnmt3a*$^{-/-}$/*3b*$^{-/-}$embryos, consistent with their normally being silenced by DNA methylation (Figure 3E, Supplementary Figure S4). When this subgroup was subdivided into genes with methylated (>20% CpG methylation at TSS) and unmethylated promoters, higher upregulation in *Dnmt3a*$^{-/-}$/*3b*$^{-/-}$ was observed in the methylated set (Figure 3E, Supplementary Figure S4A, B). L1MdT-fusion transcripts were also upregulated in *Dnmt3a*$^{-/-}$/*3b*$^{-/-}$, although not to the same extent as in *Zmym2*$^{-/-}$ (Figure 3E, Supplementary Figure S4C). Likewise, more overall trans-

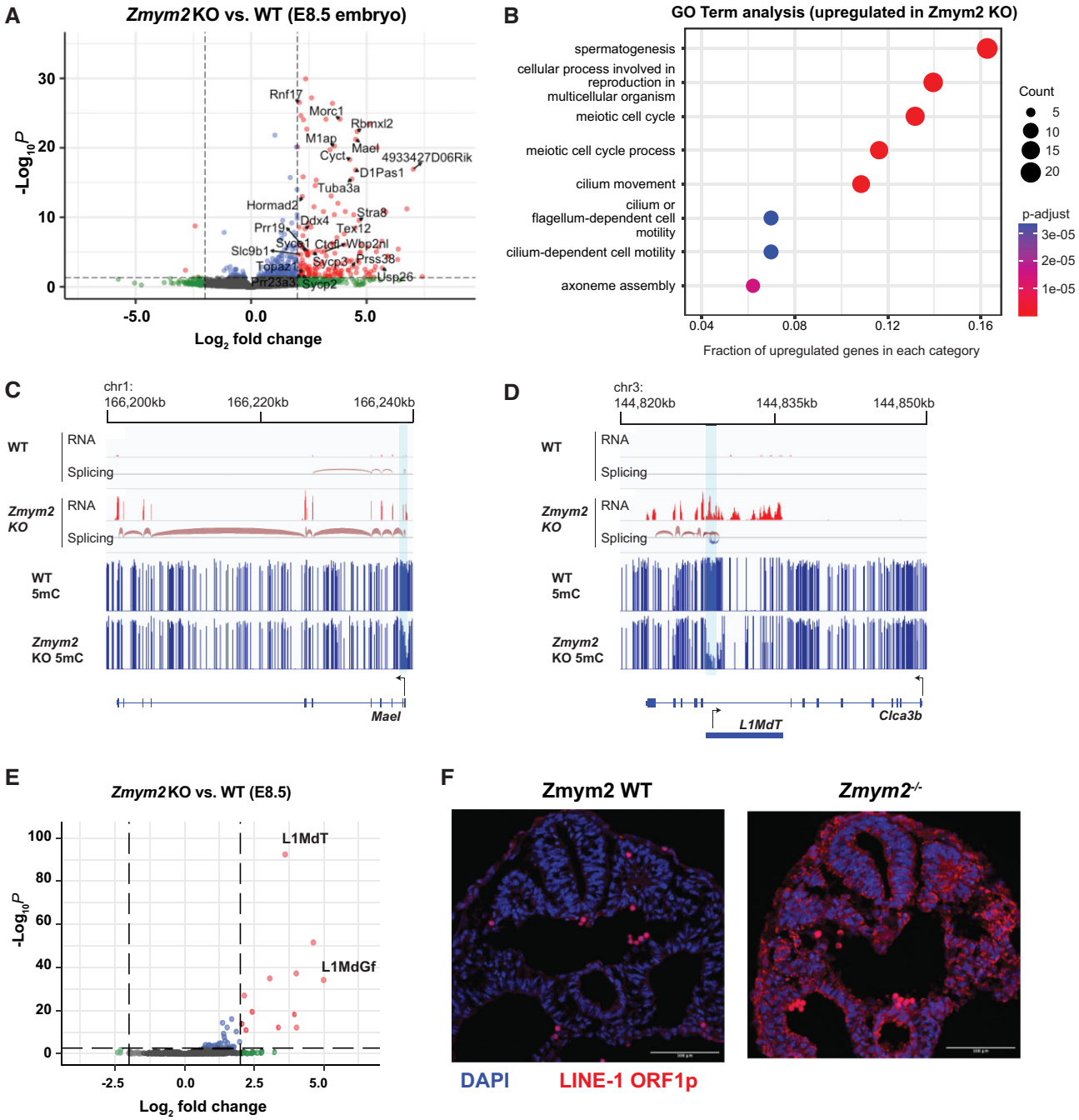


Figure 2. Germline genes and young LINE elements are upregulated in *Zmym2*$^{-/-}$ embryos. (**A**) Volcano plot of differentially expressed genes comparing *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ E8.5 embryos. Significant differentially expressed genes (fold change ≥ 4, q-value < 0.05) are expressed in red. Germline genes are labeled. The upregulated gene *Pla2g1b* is off-axis and not plotted. (**B**) Gene ontology (GO) analysis of significantly upregulated genes (fold change ≥ 4, q-value < 0.05). (**C**) Transcription and DNA methylation over the germline gene *Mael*. Note upregulation in *Zmym2*$^{-/-}$ and promoter hypomethylation. Splicing event coloured by direction of transcript (red = plus strand). A hypomethylated DMR is indicated in light blue. (**D**) Transcription and DNA methylation over the gene *Clca3b*. Note splicing from the L1MdT element to the *Clca3b* transcript, as well as hypomethylation of the L1MdT promoter. Splicing event coloured by direction of transcript (red = plus strand, blue = minus strand). A hypomethylated DMR is indicated in light blue. (**E**) Volcano plot of differentially expressed transposable elements comparing *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ E8.5 embryos. Significant differentially expressed transposons (fold change ≥ 4, q-value < 0.05) are coloured red. (**F**) Immunofluorescence staining of LINE-1 ORF1p (red) and DAPI (blue) in E9.5 *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ embryonic tissue. Note that the few, bright red cells in *Zmym2*$^{+/+}$ are autofluorescent red blood cells. Scale bars = 500μm.

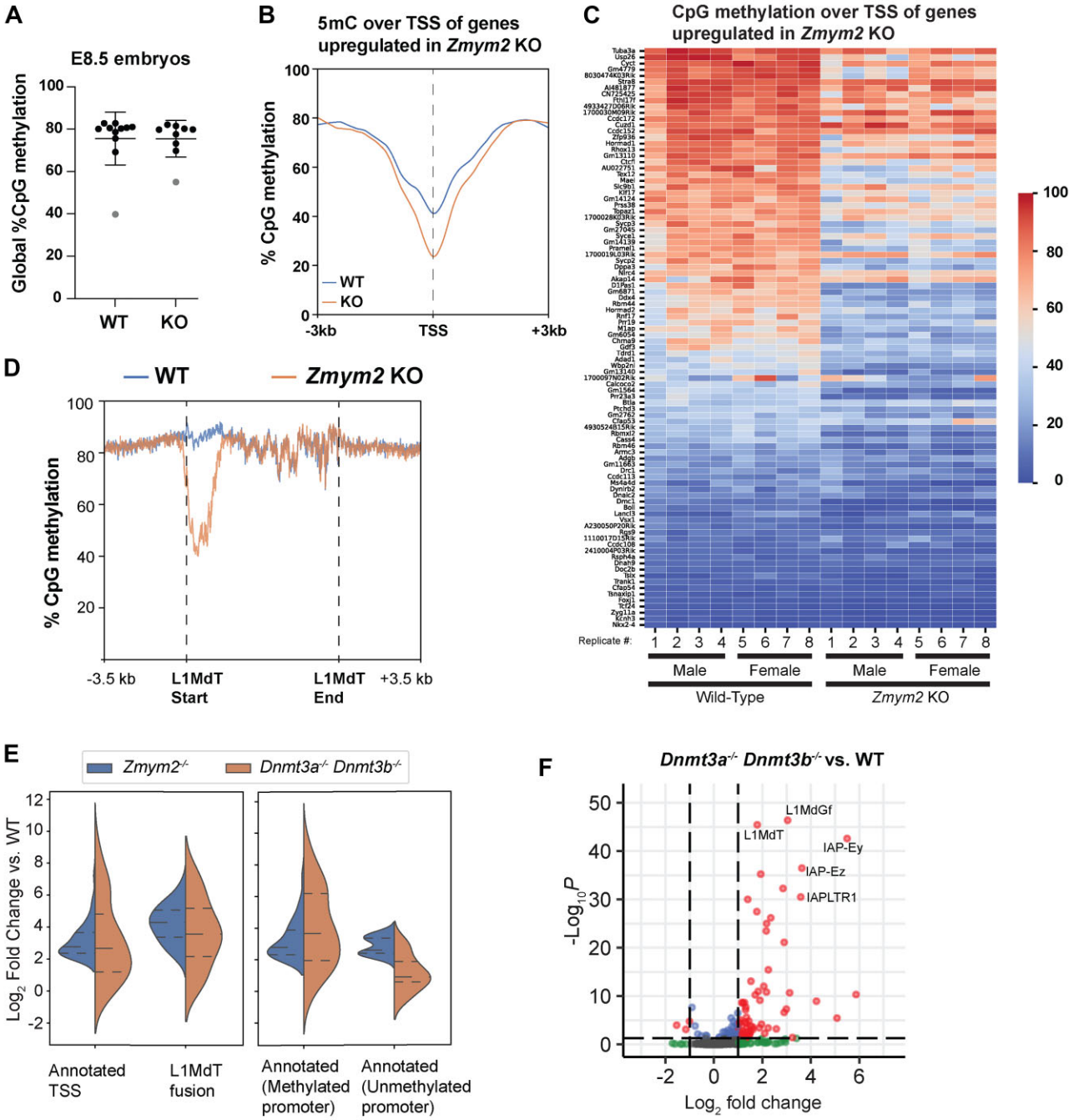


Figure 3. Impaired DNA methylation in *Zmym2*$^{-/-}$ over germline genes and LINE elements. (**A**) Swarm plot of global CpG methylation levels of individual *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ E8.5 embryos. Two samples coloured gray were excluded from subsequent analysis. (**B**) Metaplot of CpG methylation over TSS of 95 upregulated genes in *Zmym2*$^{-/-}$ embryos expressed from annotated TSS. (**C**) Heatmap of DNA methylation within 500 bp of the TSS of each of upregulated gene is indicated for each E8.5 embryo. Transposon-fusion genes and the Y-chromosomal gene *Uba1y* are excluded. (**D**) Metaplot of DNA methylation over all full-length (>6 kb) L1MdT elements in the genome. (**E**) Expression of each gene in the given category is shown (KO/WT) in *Zmym2*$^{-/-}$ and *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$. Note that most genes upregulated in *Zmym2*$^{-/-}$ show elevated expression in *Dnmt* double knockout. (**F**) Volcano plot of differentially expressed transposable elements comparing *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$ and control E8.5 embryos.

poson classes were upregulated in *Dnmt3a*$^{-/-}$/*3b*$^{-/-}$ than *Zmym2*$^{-/-}$ but upregulation of L1MdT and L1MdGf elements was somewhat weaker, suggesting that ZMYM2 may silence young LINEs by both methylation-dependent and -independent mechanisms (Figure 3F).

Further consistent with a role for ZMYM2 in promoting DNA methylation, ZMYM2 expression peaks during post-implantation development in both mouse(31) and primate(34), coincident with expression of *de novo* methyltransferases (Supplementary Figure S5A, B). As with L1MdT-fusion transcripts, ZMYM2 target genes generally show downregulation between E6.5 and E8.5, consistent with their being silenced and methylated during this time (Supplementary Figure S5C).

We then identified differentially methylated regions (DMRs) using a 1,000bp tiling approach. Three thousand three hundred forty such regions were hypomethylated in *Zmym2*$^{-/-}$, and only 224 regions were hypermethylated (Supplementary Table S5). 40.9% of these hypomethylated DMRs overlap with L1MdT elements, compared with 2.6% overlap expected by chance (Supplementary Figure S6A). Other LINE elements, most notably L1MdGf, also overlapped with DMRs more than expected by chance (Supplementary Figure S6B–D). This approach confirms ZMYM2's role in directing DNA methylation.

We then examined other types of loci regulated by DNA methylation. Methylation of gene promoters on the X-chromosome proceeded normally in female *Zmym2*$^{-/-}$ mice, indicating that ZMYM2 was not essential for methylation after X-inactivation (Supplementary Figure S7A). Canonical, stably imprinted loci, which inherit methylation from parental gametes, do not show hypomethylation (Supplementary Figure S7B), consistent with ZMYM2's role being establishment rather than maintenance of methylation. Intriguingly though, we observed hypomethylation and transcription of the transient imprint *Liz* (Long isoform of *Zdbf2*) (Supplementary Figure S7C), a locus which is imprinted during pre-implantation development but becomes biallelically methylated and fully silenced during the peri-implantation period (35).

ZMYM2 methylates young LINE retrotransposons in human ESCs

We used published RNA-seq and reduced representation bisulfite-seq (RRBS) data from *ZMYM2*$^{-/-}$ hESCs (25,26) to determine if dysregulation of transposons was observed. Indeed, we observed highly significant upregulation of the L1Hs and L1PA2 sub-families, the two youngest LINE element families in humans (Figure 4A). This was not a result of the modest shift toward a pre-implantation 'naïve' gene expression signature observed in *ZMYM2*$^{-/-}$ hESCs (25) because naïve cells do not show substantial increases in L1Hs and L1PA2 (36) (Supplementary Figure S8A). Individual L1Hs elements are too similar to each other to assign individual RRBS reads to individual elements efficiently, but mapping to a common consensus sequence showed that loss of *ZMYM2* or its known protein interactor *ATF7IP*(15,24) results in hypomethylation of the L1Hs promoter region in hESCs (Figure 4B, Supplementary Figure S8B, C). Strong hypomethylation of L1PA2 elements is also observed in *ZMYM2*$^{-/-}$ and *ATF7IP*$^{-/-}$ (Figure 4C, Supplementary Figure S8D,E). Thus, ZMYM2's role in methylating young transposons is conserved in humans.

ZMYM2 silences target loci in embryonic stem cells and embryoid bodies

We generated *Zmym2*$^{-/-}$ and non-target control mESC lines from the V6.5 mESC line (Supplementary Figure S9A, B). *Zmym2*$^{-/-}$ mESCs could be differentiated to embryoid bodies (EBs), showing similar upregulation of differentiation markers as *Zmym2*$^{+/+}$ mESCs (Figure 5A). By day 6 of differentiation reduced EB size and viability was observed, mimicking the lethality observed in mice (Supplementary Figure S9C).

Zmym2$^{-/-}$ mESCs showed upregulation of germline and transposon fusion genes relative to control *Zmym2*$^{+/+}$ mESC lines (Figure 5A, Supplementary Figure S10A–C, Supplementary Table S2). Over the course of EB differentiation, expression of the ZMYM2-target transcripts dropped substantially in *Zmym2*$^{+/+}$ cells but these genes remained expressed in *Zmym2*$^{-/-}$ (Figure 5A), akin to the phenomenon by which ZMYM2 targets are suppressed in *Zmym2*$^{+/+}$ but not *Zmym2*$^{-/-}$ mouse embryo development (Supplementary Figures S3F, S5C). Similar results were observed from a *Zmym2*$^{-/-}$ CCE mESC line (16) (Supplementary Figure S9D). *Zmym2*$^{-/-}$ mESCs showed upregulation of a wide range of transposons, but d3 EBs showed more selective enrichment of young LINE elements (Supplementary Figure S10D,E). Analysis of published data(27,37) indicates that *Zmym2*$^{-/-}$ mESCs showed reduced DNA methylation over DMRs hypomethylated in *Zmym2*$^{-/-}$ embryos and over genes upregulated in *Zmym2*$^{-/-}$ embryos (Supplementary Figure S11). Finally, using published ChIP-seq data from mESCs (16,38), we see enrichment of ZMYM2 and other CoREST components over these same hypomethylated DMRs (Supplementary Figure S12).

Thus, *Zmym2*$^{-/-}$ mESCs and EBs recapitulate dysregulation of germline genes and young LINE transposons and can be used as an *in vitro* model with which to study ZMYM2 targeting and activity.

ZMYM2 silences genes and transposons downstream of PRC1.6 and TRIM28 respectively

ZMYM2 lacks autonomous capacity for sequence-specific DNA binding but can bind SUMOylated transcription factors or complexes on chromatin and act as a corepressor (24,39). As discussed above, PRC1.6 is essential for silencing of germline genes in early embryonic development while TRIM28 silences transposons. Also, we noted that a previously reported ZMYM2-motif(16) is almost identical to the binding motif of the transcription factor REST. This suggests that ZMYM2 may also home to chromatin by interaction with REST, a transcription factor which primarily functions to repress neuronal genes in non-neuronal lineages (40). We thus performed cluster analysis of all high-enrichment ZMYM2 ChIP-seq peaks, incorporating published mESC ChIP-seq data for ZMYM2(16), PRC1.6 components (MGA and L3MBTL2(7)), TRIM28 (41),

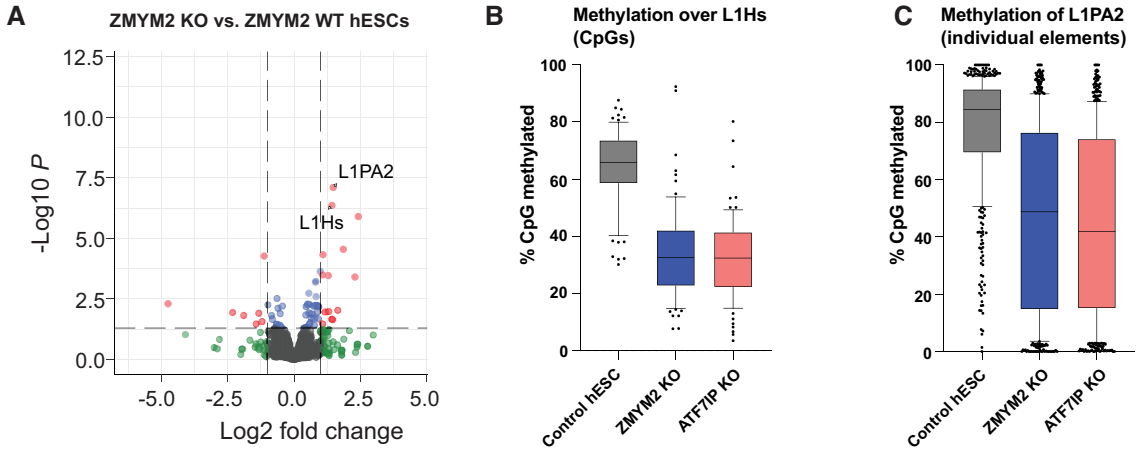


Figure 4. Upregulation and hypomethylation of young LINE elements in $ZMYM2^{-/-}$ human embryonic stem cells. (**A**) Volcano plot of differentially expressed transposons comparing $ZMYM2^{+/+}$ and $ZMYM2^{-/-}$ hESCs. Significant differentially expressed transposons (fold change ≥ 2, q-value < 0.05) are coloured red. (**B**) Boxplot of CpG methylation across 1 kb start regions of full length L1Hs transposable elements. Data was mapped to an L1Hs consensus element, and individual CpG sites are represented as dots. (**C**) Boxplot of average CpG methylation across L1PA2 individual transposable elements. Each element is represented as a dot.

REST (42) and SUMO2 (41). We observed strongly distinct PRC1.6^{+}, TRIM28^{+} and REST^{+} clusters, indicating that ZMYM2 homes to each of these complexes independently (Figure 5B).

ZMYM2 is known to interact with the PRC1.6 complex (15,16), which is heavily SUMOylated in mESCs (43). ZMYM2 tightly colocalizes with PRC1.6 at promoters of germline genes in mESCs (Figure 5C,D, Supplementary Figure S13A–C) and PRC1.6 is strongly enriched at genes upregulated in *Zmym2*$^{-/-}$ mouse embryos relative to other genes (Figure 5E, Supplementary Figure S13D). ZMYM2 is globally enriched over PRC1.6 ChIP-seq peaks (Figure 5F), with virtually all PRC1.6 sites showing enriched binding of ZMYM2 (Figure 5G). We conducted CUT&RUN to detect distribution of the PRC1.6 component L3MBTL2 in control and *Zmym2*$^{-/-}$ mESCs and found no difference in binding to genes upregulated in *Zmym2*$^{-/-}$ embryos (Figure 5H). Likewise, there is no loss of the histone modification H2AK119Ub (deposited by the RING1 component of PRC1.6) in *Zmym2*$^{-/-}$ mESCs or day 3 EBs (Supplementary Figure S13E, F). Thus, ZMYM2 is not essential for PRC1.6 binding or activity but is nonetheless essential as a corepressor to silence some PRC1.6 target genes.

The TRIM28 complex is heavily SUMOylated in mESCs (43), and ZMYM2 could also home with the TRIM28 complex via their known mutual interactor ATF7IP (24). ZMYM2 shows colocalization with TRIM28 at L1MdT elements, and ZMYM2 is strongly enriched over TRIM28 binding sites genome wide (Figure 5G, I–K, Supplementary Figure S14A–D). As with PRC1.6, ZMYM2 homes to virtually all TRIM28 targets and is essential for silencing of a distinct subset of them. H3K9me3 enrichment is still observed over TRIM28 sites and hypomethylated L1MdT elements in *Zmym2*$^{-/-}$ mESCs, indicating that recruitment of TRIM28 and SETDB1 still occurs in *Zmym2*$^{-/-}$ mESCs. A localized drop in H3K9me3 observed immediately over the hypomethylated L1MdTs may be a consequence of overall nucleosome depletion (Supplementary Figure S14E–G).

How ZMYM2 homes to REST bindings sites is unclear. ZMYM2 and REST both bind the CoREST complex but in a mutually exclusive manner (17) and REST is not known to be SUMOylated. Nonetheless, ZMYM2 is very strongly enriched at REST sites (Figure 5B, Supplementary Figure S15A–D). We do not observe hypomethylation of REST sites (Supplementary Figure S15E) or upregulation of REST target genes (40), so the biological significance of this interaction remains unclear.

Finally, Cluster 5 of ZMYM2 peaks showed no strong co-enrichment with PRC1.6, TRIM28 and REST (Figure 5B). Motif analysis of Cluster 5 peaks showed enrichment for the DR0 Nuclear receptor motif (Figure 5L) suggesting that an as yet unidentified nuclear receptor may recruit ZMYM2 to chromatin.

Loss of ZMYM2 results in H3K4 hypermethylation at target loci

The histone modifications H3K4me2 and H3K4me3 are ubiquitous at sites of transcriptional initiation in mammals (44,45). The *de novo* methyltransferases DNMT3A, DNMT3B and DNMT3L in turn bind to the H3 N-terminus via their ADD domains, but this binding is strongly antagonized by the presence of the modification H3K4me2 or H3K4me3 (46–48). H3K4me2/3-marked chromatin thus escapes *de novo* DNA methylation during development (49–51), and actively transcribed gene promoters remain unmethylated in the post-implantation embryo (52–54). PRC1.6 and TRIM28 both silence target loci prior to deposition of methylation (9,55). We therefore theorized that ZMYM2 may promote DNA methylation by causing transcriptional silencing and loss of H3K4me2/3 at target sites.

Consistent with the gene expression pattern of ZMYM2-targets in development, published data (56,57) reveals high H3K4me3 in pre-implantation embryos at ZMYM2-hypomethylated DMRs and genic targets, with a dramatic decline during post-implantation development as the tar-

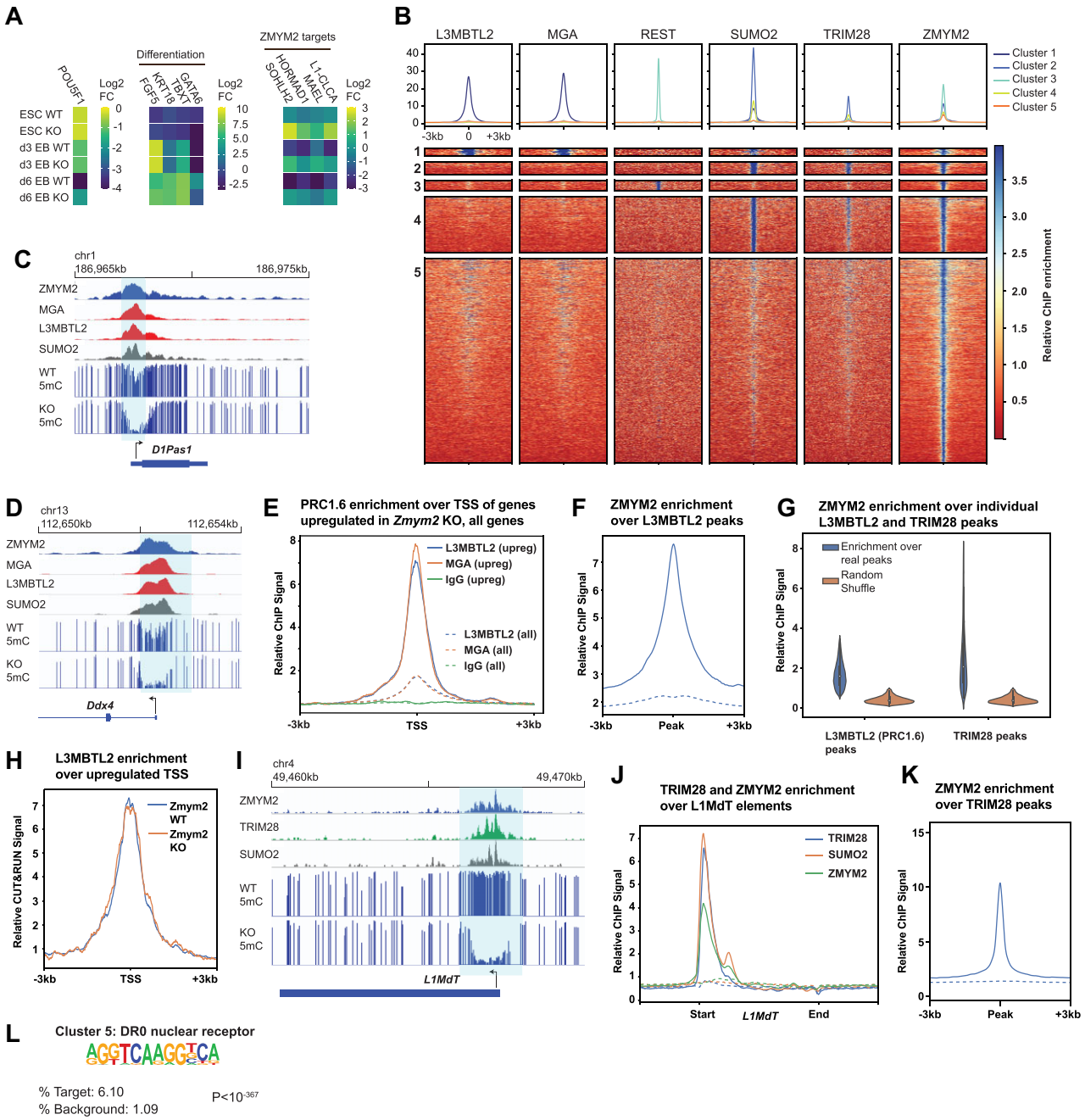


Figure 5. ZMYM2 silences PRC1.6 and TRIM28 targets and promotes their methylation. (**A**) Relative expression of the pluripotency factor *Pou5f1*, differentiation genes, and ZMYM2-target genes in three *Zmym2*$^{-/-}$ lines as compared with three control lines, during two separate embryoid body differentiation experiments. Average results from these $n = 6$ replicates total are shown. (**B**) Heatmaps of ZMYM2, L3MBTL2, MGA, REST, SUMO2 and TRIM28 enrichment over ZMYM2 ChIP-seq peak summits and the flanking 3 kb regions ($n = 21\,549$ peaks). Peaks were partitioned into five clusters using a *k*-means algorithm. Peak numbers: Cluster 1, 290 peaks; Cluster 2, 931 peaks; Cluster 3, 796 peaks; Cluster 4, 4080 peaks; Cluster 5, 14 969 peaks. Cluster 1 peaks are shown at twice the width of other peaks to enhance visibility. (**C, D**) ChIP-seq and DNA methylation data over the *D1Pas* (**C**) and *Ddx4* (**D**) loci. Note colocalization of ZMYM2 with PRC1.6 components and hypomethylation in *Zmym2*$^{-/-}$. (**E**) Metaplot of L3MBTL2, MGA and IgG negative control ChIP-seq signals over the annotated transcription start site of upregulated genes in *Zmym2*$^{-/-}$ E8.5 embryos (solid lines) with transposon-fusions excluded and the annotated transcription start site of all genes (dashed lines). (**F**) Metaplot of ZMYM2 ChIP-seq signal over L3MBTL2 peaks. Dashed line indicates ChIP input. (**G**) ZMYM2 ChIP-seq enrichment over L3MBTL2 (13 936 peaks) and TRIM28 (6830 peaks) peak sets. Enrichment over random peaks of the same size is shown as a comparison. (**H**) Metaplot of L3MBTL2 CUT&RUN ChIP-seq signal over the annotated TSS of upregulated genes in both *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ mESCs. (**I**) ChIP-seq and DNA methylation data over a representative LINE element. Note overlap of TRIM28 and ZMYM2 peaks and hypomethylation over the peak site. (**J**) Metaplot of ZMYM2, TRIM28, SUMO2 ChIP-seq over all full length L1MdT elements. Dashed line indicates ChIP input. (**K**) Metaplot of ZMYM2 over all uniquely mapped TRIM28 peaks. Dashed line indicates ChIP input. (**L**) HOMER motif analysis of Cluster 5 ZMYM2 peaks show enrichment of the DR0 nuclear receptor motif relative to random (background) signal.

gets are silenced (Supplementary Figure S16). To determine the effects of *Zmym2* deficiency on H3K4 methylation, we performed ChIP-seq of H3K4me2 and H3K4me3 in both $Zmym2^{+/+}$ and $Zmym2^{-/-}$ mESCs and day 3 EBs. Each replicate was normalized to control for ChIP quality (Supplementary Figures S17A–D). We observe elevated levels of H3K4me2 and H3K4me3 in $Zmym2^{-/-}$ mESCs over the hypomethylated DMRs found in E8.5 $Zmym2^{-/-}$ mice (Figure 6A, B). H3K4 hypermethylation was also observed at the promoters of full-length L1MdT elements and the genes upregulated in $Zmym2^{-/-}$ mice (Figure 6C–H, Supplementary Figure S17E–J). These trends are accentuated further upon differentiation to EBs, as H3K4 levels over ZMYM2-regulated loci drop substantially in $Zmym2^{+/+}$ cells but much less so in $Zmym2^{-/-}$ cells (Figure 6, Supplementary Figure S17E–J). Hence, ZMYM2 functions to reduce H3K4me2/3 at target genes and transposons and thus creates a chromatin environment conducive to *de novo* methylation.

DISCUSSION

ZMYM2 is clearly essential for silencing and DNA methylation of germline genes and young transposons in development. A combined model of its activity runs as follows. ZMYM2 binds to PRC1.6 and TRIM28 sites, where it functions as a corepressor. By suppressing transcription, it reduces levels of H3K4 methylation at these sites, facilitating DNA methylation upon implantation and establishing stable silencing (Figure 7). As discussed above, ZMYM2 is known to recruit the CoREST complex, whose constituent protein LSD1 can directly mediate H3K4me1 and H3K4me2 demethylation, providing a second potential mechanism by which H3K4 methylation can be removed (58).

PRC1.6 and TRIM28 are heavily SUMOylated in mESCs (43) and ZMYM2 could home to them via these SUMO2 marks (Figure 5). A recent report indicates association of ZMYM2 with TRIM28 in U2OS cells, with homing to TRIM28 sites lost if a SIM domain in ZMYM2 is mutated (59). At the same time, we cannot rule out SUMO-independent recruitment of ZMYM2 to these complexes. ZMYM2 interacts with the Fibronectin III domain of ATF7IP in an apparently SUMO-independent manner, and ATF7IP has been reported to interact with both TRIM28 and the PRC1.6 component MGA(24). It is striking that loss of either ATF7IP or ZMYM2 results in hypomethylation of young LINE elements in hESCs (Figure 4B, C, Supplementary Figure S8B–E), and both genes have been hits in CRISPR screens for factors that maintain imprint methylation in stem cells (26,27). One possibility is that ATF7IP's primary function at these loci is to recruit ZMYM2 to such elements. However, ATF7IP performs other function such as promoting nuclear localization of the H3K9 methyltransferase SETDB1 (60), so the similarity of the $ATF7IP^{-/-}$ and $ZMYM2^{-/-}$ phenotypes could reflect genuine interaction or just that both proteins are important for silencing.

The $Zmym2^{-/-}$ mice show embryonic lethality amidst aberrant upregulation of LINE protein. It is difficult to determine what precisely halts the development of the $Zmym2^{-/-}$ embryos, but it is well established that LINE elements can induce cellular lethality, both by inducing anti-viral responses and by the endonuclease activity of L1ORF2p(61). The phenomenon of LINE-gene fusion transcripts warrants further note. There are numerous examples of LTR-retrotransposons, most frequently 'solo LTRs' detached from larger transposons, serving as alternative promoters or enhancers for protein-coding transcripts (62,63). LINE elements integrate into the genome in a 3' to 5' manner, so a promoter is only present if complete integration has occurred. Thus, 'orphan' LINE promoters are expected to be a rarer phenomenon. Nonetheless, LINE-gene fusions apparently occur, both as a result of bidirectional and forward transcription, and are expressed at significant levels in wild-type pre-implantation embryos. We do not know how many of the 46 LINE-transcript fusions we observe code for functional protein or whether they have any biological function, but it is intriguing that L1MdT expression during pre-implantation murine development is essential for developmental progression(64,65).

It is important to consider the phenotypes we do not see in $Zmym2^{-/-}$ but which might have been expected based on literature. We do not see pre-implantation developmental arrest (16), but this can be explained by maternal deposition of intact *Zmym2* RNA in the oocytes of $Zmym2^{+/-}$ mothers. A number of published observations in embryonic stem cells were not observed in the mice. We also do not observe failure to exit pluripotency (23,25), with E8.5 $Zmym2^{-/-}$ having undergone gastrulation normally. Embryonic stem cells adapted to continuous pluripotent culture may be more resistant to differentiation than developing mice. We also do not observe loss of canonical imprints in $Zmym2^{-/-}$ embryos (26,27) (Supplementary Figure S7B). It may be that the dynamic state of mESCs, in which both *de novo* methylation and demethylation occur in continuous culture, does not reflect methylation dynamics in the rapidly developing mouse embryo. Alternatively, ZMYM2 may be essential to maintain imprints during the pre-implantation period when DNA methylation is lost, and maternally-inherited *Zmym2* masks this phenotype in the $Zmym2^{-/-}$ mice.

Unlike the $Zmym2^{-/-}$ mice, we also do not observe demethylation of germline genes in $ZMYM2^{-/-}$ hESCs (data not shown). While this could reflect species difference, it is important to note here that conventionally cultured primed hESCs correspond to a developmentally advanced state that features high global DNA methylation, including over germline genes (36). Even if ZMYM2 is critical for DNA methylation establishment over germline genes in development, its loss will not necessarily cause loss of existing DNA methylation. Young LINE elements by contrast are targets of continuous TET activity in hESCs and thus depend on continual *de novo* methyltransferase activity to remain methylated (66), potentially explaining how they become demethylated upon loss of ZMYM2.

ZMYM2 shows co-association with a variety of chromatin-bound complexes. Zygotic depletion of *Zmym2* RNA blocks blastocyst formation (16). The $Zmym2^{-/-}$ mouse we have generated, which is to our knowledge the first report of a ZMYM-family transcription factor knockout in literature, shows lethality and severe epigenetic abnormality. The cranial, cardiac, musculoskeletal,

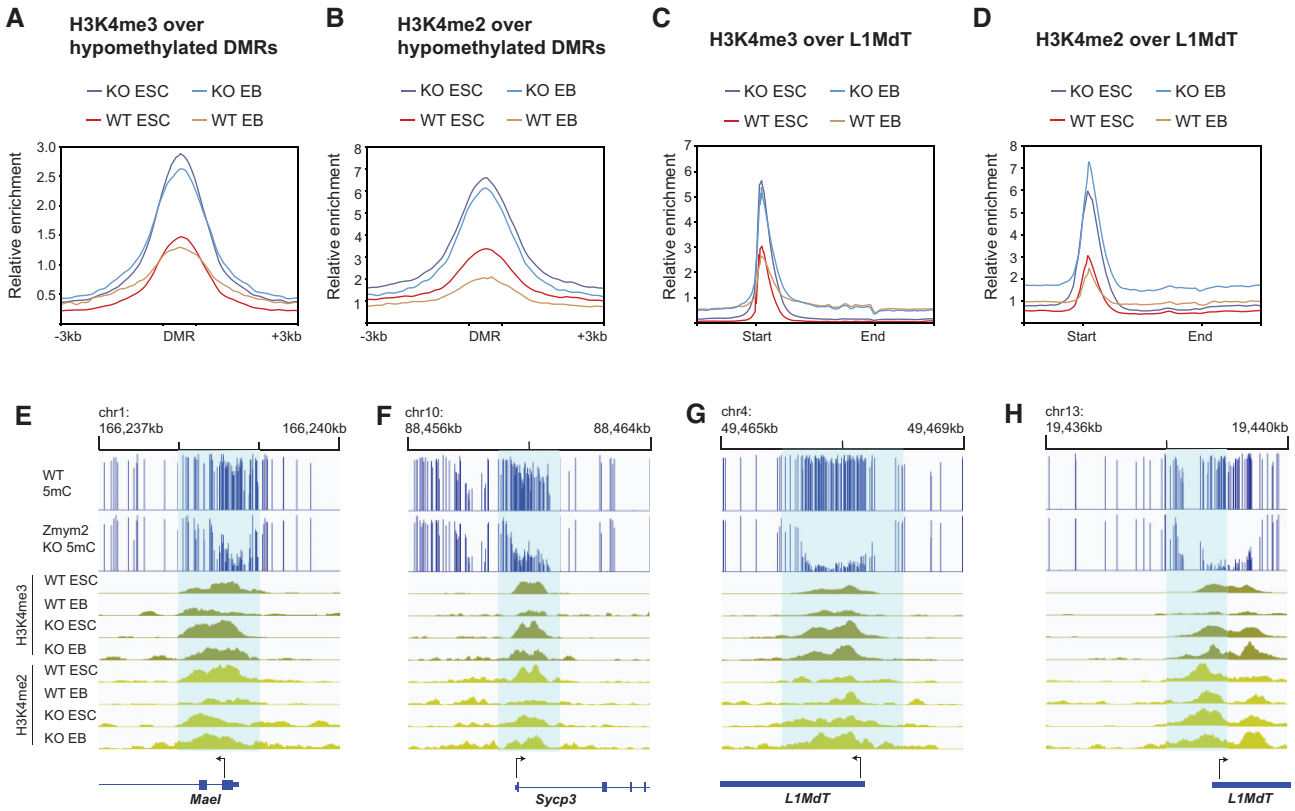


Figure 6. Aberrant H3K4 hypermethylation at ZMYM2-target sites in *Zmym2*$^{-/-}$ mESCs and EBs. (**A**) Metaplot of H3K4me3 over DMRs hypomethylated in *Zmym2*$^{-/-}$ embryos in *Zmym2*$^{-/-}$ and control ESCs and EBs. (**B**) Metaplot of H3K4me2 over hypomethylated DMRs. (**C**) Metaplot of H3K4me3 over full length L1MdT elements. (**D**) Metaplot of H3K4me2 over full length L1MdT elements. For (A–D) averaged data for two *Zmym2*$^{-/-}$ and two control mESC lines are shown. To normalize for ChIP quality, enrichment is normalized based on peak height for all annotated genes. (**E–H**) Examples of increased H3K4me2/3 in *Zmym2*$^{-/-}$ mESCs over hypomethylated DMRs including genes (E, F) and transposons (G, H).

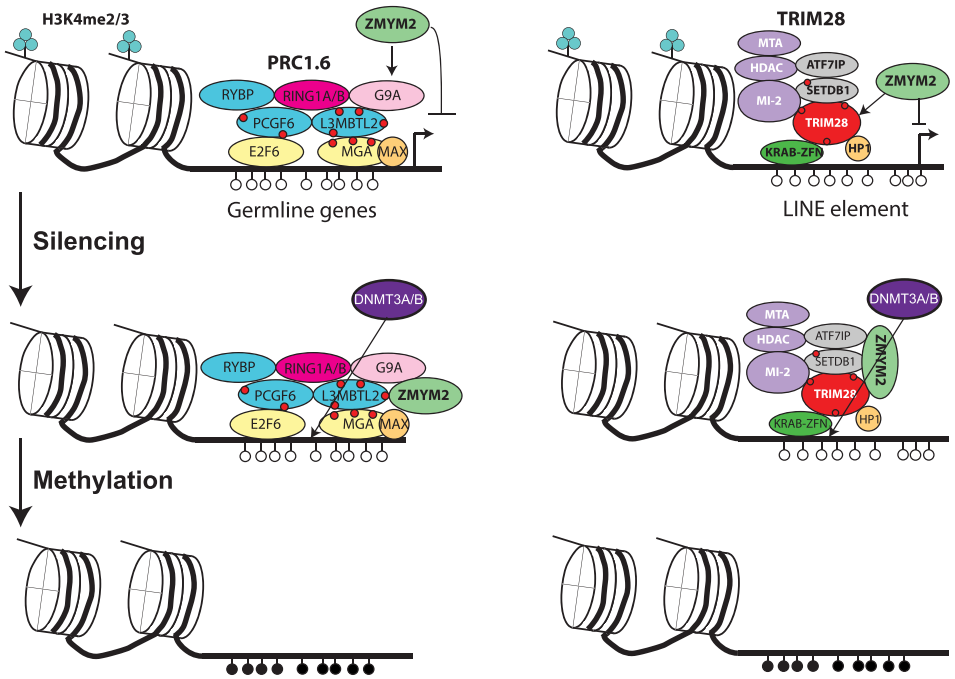


Figure 7. Proposed mechanism for ZMYM2-mediated DNA methylation. ZMYM2 is recruited to PRC1.6 and TRIM28 binding sites respectively via binding to SUMOylation sites (indicated as red circles) or interaction with complex components such as ATF7IP. ZMYM2 mediates silencing of target loci, resulting in loss of H3K4 methylation and acquisition of DNA methylation.

CAKUT and possible infertility phenotype caused by ZMYM2 heterozygosity in humans suggests extensive further roles in development (15). Mutations of ZMYM3 has been implicated in mental retardation (67,68), and all six ZMYM-family transcription factors are widely expressed in human tissues (69). The role of ZMYM-family transcription factors as cofactors for repression may be widespread and underappreciated.

DATA AVAILABILITY

All sequencing data (RNA-seq, ChIP-seq, CUT&RUN, WGBS) has been uploaded to the GEO repository under the accession number GSE214233.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the McGill Genome Centre, Génome Québec, the Centre for Applied Genomics at SickKids Hospital, the McGill University Advanced BioImaging Facility (ABIF), and the Rosalind and Morris Goodman Cancer Institute Histology Innovation Platform for assistance. We thank Mitra Cowan and the McGill Integrated Core for Animal Modeling for assistance generating and maintaining mice. We thank Jianlong Wang (Columbia University) for donating *Zmym2*$^{-/-}$ mESCs.
Author contributions: A.G., D.S., J.S. and M.K. conducted experiments. D.S., I.H., Q.Z. and S.K. conducted bioinformatic analysis. Y.Y. provided guidance on phenotypic analysis of mice. G.B. provided guidance on bioinformatic analysis. M.B. and W.P. supervised the research project. A.G., D.S. and W.P. wrote the manuscript.

FUNDING

CIHR Project Grant [PJT-159768 to M.B.]; CIHR Project Grant [PJT-166169 to W.P.]; CIHR Priority Announcement [PJG-183984 to W.P., M.B.]; New Frontiers in Research Fund Grant [NFRFE-2018-00883 to W.P.]; NSERC Grant [RGPIN-2018-04856 to W.P.]; A.G. was supported by a CIHR Canada Graduate Scholarships Doctoral Award; Q.Z. was supported by doctoral training awards by the FRQS and JST; G.B. was supported by a Canada Research Chair and an FRQS Chercheurs-mérite; W.P. was supported by an FRQS Chercheurs-boursier. Funding for open access charge: CIHR [PJT-166169].
Conflict of interest statement. None declared.

REFERENCES

1. Baubec,T., Ivanek,R., Lienert,F. and Schubeler,D. (2013) Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell*, **153**, 480–492.
2. Smith,Z.D., Chan,M.M., Mikkelsen,T.S., Gu,H., Gnirke,A., Regev,A. and Meissner,A. (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature*, **484**, 339–344.
3. Bestor,T.H., Edwards,J.R. and Boulard,M. (2014) Notes on the role of dynamic DNA methylation in mammalian development. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 6796–6799
4. Borgel,J., Guibert,S., Li,Y., Chiba,H., Schubeler,D., Sasaki,H., Forne,T. and Weber,M. (2010) Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet*, **42**, 1093–1100.
5. Hargan-Calvopina,J., Taylor,S., Cook,H., Hu,Z., Lee,S.A., Yen,M.R., Chiang,Y.S., Chen,P.Y. and Clark,A.T. (2016) Stage-specific demethylation in primordial germ cells safeguards against precocious differentiation. *Dev. Cell*, **39**, 75–86.
6. Pohlers,M., Truss,M., Frede,U., Scholz,A., Strehle,M., Kuban,R.J., Hoffmann,B., Morkel,M., Birchmeier,C. and Hagemeier,C. (2005) A role for E2F6 in the restriction of male-germ-cell-specific gene expression. *Curr. Biol.*, **15**, 1051–1057.
7. Stielow,B., Finkernagel,F., Stiewe,T., Nist,A. and Suske,G. (2018) MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical polycomb repressive complex PRC1.6. *PLoS Genet.*, **14**, e1007193.
8. Iyengar,S. and Farnham,P.J. (2011) KAP1 protein: an enigmatic master regulator of the genome. *J. Biol. Chem.*, **286**, 26267–26276.
9. Rowe,H.M., Jakobsson,J., Mesnard,D., Rougemont,J., Reynard,S., Aktas,T., Maillard,P.V., Layard-Liesching,H., Verp,S., Marquis,J. *et al.* (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, **463**, 237–240.
10. Rowe,H.M., Friedli,M., Offner,S., Verp,S., Mesnard,D., Marquis,J., Aktas,T. and Trono,D. (2013) De novo DNA methylation of endogenous retroviruses is shaped by KRAB-zfps/KAP1 and ESET. *Development*, **140**, 519–529.
11. Tao,Y., Yen,M.R., Chitiashvili,T., Nakano,H., Kim,R., Hosohama,L., Tan,Y.C., Nakano,A., Chen,P.Y. and Clark,A.T. (2018) TRIM28-Regulated transposon repression is required for Human germline competency and not primed or naive Human pluripotency. *Stem Cell Rep.*, **10**, 243–256.
12. Rousseaux,M.W., Revelli,J.P., Vazquez-Velez,G.E., Kim,J.Y., Craigen,E., Gonzales,K., Beckinghausen,J. and Zoghbi,H.Y. (2018) Depleting Trim28 in adult mice is well tolerated and reduces levels of alpha-synuclein and tau. *Elife*, **7**, e36768.
13. Dahlet,T., Argueso Lleida,A., Al Adhami,H., Dumas,M., Bender,A., Ngondo,R.P., Tanguy,M., Vallet,J., Auclair,G., Bardet,A.F. *et al.* (2020) Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. *Nat. Commun.*, **11**, 3153.
14. Aguilar-Martinez,E., Chen,X., Webber,A., Mould,A.P., Seifert,A., Hay,R.T. and Sharrocks,A.D. (2015) Screen for multi-SUMO-binding proteins reveals a multi-SIM-binding mechanism for recruitment of the transcriptional regulator ZMYM2 to chromatin. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E4854–E4863.
15. Connaughton,D.M., Dai,R., Owen,D.J., Marquez,J., Mann,N., Graham-Paquin,A.L., Nakayama,M., Coyaud,E., Laurent,E.M.N., St-Germain,J.R. *et al.* (2020) Mutations of the transcriptional corepressor ZMYM2 cause syndromic urinary tract malformations. *Am J. Hum. Genet*, **107**, 727–742.
16. 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.e237.
17. Gocke,C.B. and Yu,H. (2008) ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. *PLoS One*, **3**, e3255.
18. Guzzo,C.M., Ringel,A., Cox,E., Uzoma,I., Zhu,H., Blackshaw,S., Wolberger,C. and Matunis,M.J. (2014) Characterization of the SUMO-binding activity of the myeloproliferative and mental retardation (MYM)-type zinc fingers in ZNF261 and ZNF198. *PLoS One*, **9**, e105271.
19. Gill,G. (2005) Something about SUMO inhibits transcription. *Curr. Opin. Genet Dev.*, **15**, 536–541.
20. Hendriks,I.A. and Vertegaal,A.C. (2016) A comprehensive compilation of SUMO proteomics. *Nat. Rev. Mol. Cell Biol.*, **17**, 581–595.
21. Shi,Y.J., Matson,C., Lan,F., Iwase,S., Baba,T. and Shi,Y. (2005) Regulation of LSD1 histone demethylase activity by its associated factors. *Mol. Cell*, **19**, 857–864.
22. Hakimi,M.A., Dong,Y., Lane,W.S., Speicher,D.W. and Shiekhattar,R. (2003) A candidate X-linked mental retardation gene

is a component of a new family of histone deacetylase-containing complexes. *J. Biol. Chem.*, **278**, 7234–7239.
23. Hackett,J.A., Huang,Y., Gunesdogan,U., Gretarsson,K.A., Kobayashi,T. and Surani,M.A. (2018) Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nat. Commun.*, **9**, 4292.
24. Tsusaka,T., Fukuda,K., Shimura,C., Kato,M. and Shinkai,Y. (2020) The fibronectin type-III (FNIII) domain of ATF7IP contributes to efficient transcriptional silencing mediated by the SETDB1 complex. *Epigenetics Chromatin*, **13**, 52.
25. Lezmi,E., Weissbein,U., Golan-Lev,T., Nissim-Rafinia,M., Meshorer,E. and Benvenisty,N. (2020) The chromatin regulator ZMYM2 restricts Human pluripotent stem cell growth and is essential for teratoma formation. *Stem Cell Rep.*, **15**, 1275–1286.
26. Bar,S., Vershkov,D., Keshet,G., Lezmi,E., Meller,N., Yilmaz,A., Yanuka,O., Nissim-Rafinia,M., Meshorer,E., Eldar-Geva,T. *et al.* (2021) Identifying regulators of parental imprinting by CRISPR/Cas9 screening in haploid human embryonic stem cells. *Nat. Commun.*, **12**, 6718.
27. Butz,S., Schmolka,N., Karemaker,I.D., Villasenor,R., Schwarz,I., Domcke,S., Uijttewaal,E.C.H., Jude,J., Lienert,F., Krebs,A.R. *et al.* (2022) DNA sequence and chromatin modifiers cooperate to confer epigenetic bistability at imprinting control regions. *Nat. Genet*, **54**, 1702–1710
28. Nguyen,A.H., Tremblay,M., Haigh,K., Koumakpayi,I.H., Paquet,M., Pandolfi,P.P., Mes-Masson,A.M., Saad,F., Haigh,J.J. and Bouchard,M. (2013) Gata3 antagonizes cancer progression in Pten-deficient prostates. *Hum. Mol. Genet*, **22**, 2400–2410.
29. Narlis,M., Grote,D., Gaitan,Y., Boualia,S.K. and Bouchard,M. (2007) Pax2 and pax8 regulate branching morphogenesis and nephron differentiation in the developing kidney. *J. Am Soc. Nephrol.*, **18**, 1121–1129.
30. Wang,L., Zhang,J., Duan,J., Gao,X., Zhu,W., Lu,X., Yang,L., Zhang,J., Li,G., Ci,W. *et al.* (2014) Programming and inheritance of parental DNA methylomes in mammals. *Cell*, **157**, 979–991.
31. Smith,Z.D., Shi,J., Gu,H., Donaghey,J., Clement,K., Cacchiarelli,D., Gnirke,A., Michor,F. and Meissner,A. (2017) Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature*, **549**, 543–547.
32. 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.
33. Sookdeo,A., Hepp,C.M., McClure,M.A. and Boissinot,S. (2013) Revisiting the evolution of mouse LINE-1 in the genomic era. *Mobile DNA*, **4**, 3.
34. Nakamura,T., Okamoto,I., Sasaki,K., Yabuta,Y., Iwatani,C., Tsuchiya,H., Seita,Y., Nakamura,S., Yamamoto,T. and Saitou,M. (2016) A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature*, **537**, 57–62.
35. Greenberg,M.V., Glaser,J., Borsos,M., Marjou,F.E., Walter,M., Teissandier,A. and Bourc'his,D. (2017) Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. *Nat. Genet*, **49**, 110–118.
36. Pastor,W.A., Chen,D., Liu,W., Kim,R., Sahakyan,A., Lukianchikov,A., Plath,K., Jacobsen,S.E. and Clark,A.T. (2016) Naive Human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. *Cell Stem Cell*, **18**, 323–329.
37. Li,Z., Dai,H., Martos,S.N., Xu,B., Gao,Y., Li,T., Zhu,G., Schones,D.E. and Wang,Z. (2015) Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. *Genome Biol.*, **16**, 115.
38. Whyte,W.A., Bilodeau,S., Orlando,D.A., Hoke,H.A., Frampton,G.M., Foster,C.T., Cowley,S.M. and Young,R.A. (2012) Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature*, **482**, 221–225.
39. Ludtke,T.H., Kleppa,M.J., Rivera-Reyes,R., Qasrawi,F., Connaughton,D.M., Shril,S., Hildebrandt,F. and Kispert,A. (2022) Proteomic analysis identifies ZMYM2 as endogenous binding partner of TBX18 protein in 293 and A549 cells. *Biochem. J.*, **479**, 91–109.
40. Otto,S.J., McCorkle,S.R., Hover,J., Conaco,C., Han,J.J., Impey,S., Yochum,G.S., Dunn,J.J., Goodman,R.H. and Mandel,G. (2007) A new binding motif for the transcriptional repressor REST uncovers large gene networks devoted to neuronal functions. *J. Neurosci.*, **27**, 6729–6739.
41. Yang,B.X., El Farran,C.A., Guo,H.C., Yu,T., Fang,H.T., Wang,H.F., Schlesinger,S., Seah,Y.F., Goh,G.Y., Neo,S.P. *et al.* (2015) Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell*, **163**, 230–245.
42. McGann,J.C., Oyer,J.A., Garg,S., Yao,H., Liu,J., Feng,X., Liao,L., Yates,J.R. 3rd and Mandel,G. (2014) Polycomb- and REST-associated histone deacetylases are independent pathways toward a mature neuronal phenotype. *Elife*, **3**, e04235.
43. Cossec,J.C., Theurillat,I., Chica,C., Bua Aguin,S., Gaume,X., Andrieux,A., Iturbide,A., Jouvion,G., Li,H., Bossis,G. *et al.* (2018) SUMO safeguards somatic and pluripotent cell identities by enforcing distinct chromatin states. *Cell Stem Cell*, **23**, 742–757.
44. Krogan,N.J., Dover,J., Wood,A., Schneider,J., Heidt,J., Boateng,M.A., Dean,K., Ryan,O.W., Golshani,A., Johnston,M. *et al.* (2003) The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol. Cell*, **11**, 721–729.
45. Lee,J.H. and Skalnik,D.G. (2008) Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol. Cell. Biol.*, **28**, 609–618.
46. Noh,K.M., Wang,H., Kim,H.R., Wenderski,W., Fang,F., Li,C.H., Dewell,S., Hughes,S.H., Melnick,A.M., Patel,D.J. *et al.* (2015) Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse escs. *Mol. Cell*, **59**, 89–103.
47. Otani,J., Nankumo,T., Arita,K., Inamoto,S., Ariyoshi,M. and Shirakawa,M. (2009) Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep.*, **10**, 1235–1241.
48. Zhang,Y., Jurkowska,R., Soeroes,S., Rajavelu,A., Dhayalan,A., Bock,I., Rathert,P., Brandt,O., Reinhardt,R., Fischle,W. *et al.* (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.*, **38**, 4246–4253.
49. Singh,P., Li,A.X., Tran,D.A., Oates,N., Kang,E.R., Wu,X. and Szabo,P.E. (2013) De novo DNA methylation in the male germ line occurs by default but is excluded at sites of H3K4 methylation. *Cell Rep.*, **4**, 205–219.
50. Dahl,J.A., Jung,I., Aanes,H., Greggains,G.D., Manaf,A., Lerdrup,M., Li,G., Kuan,S., Li,B., Lee,A.Y. *et al.* (2016) Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature*, **537**, 548–552.
51. Morselli,M., Pastor,W.A., Montanini,B., Nee,K., Ferrari,R., Fu,K., Bonora,G., Rubbi,L., Clark,A.T., Ottonello,S. *et al.* (2015) In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *Elife*, **4**, e06205.
52. Weber,M., Hellmann,I., Stadler,M.B., Ramos,L., Paabo,S., Rebhan,M. and Schubeler,D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet*, **39**, 457–466.
53. Greenfield,R., Tabib,A., Keshet,I., Moss,J., Sabag,O., Goren,A. and Cedar,H. (2018) Role of transcription complexes in the formation of the basal methylation pattern in early development. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, 10387–10391.
54. Zhou,F., Wang,R., Yuan,P., Ren,Y., Mao,Y., Li,R., Lian,Y., Li,J., Wen,L., Yan,L. *et al.* (2019) Reconstituting the transcriptome and DNA methylome landscapes of human implantation. *Nature*, **572**, 660–664.
55. Mochizuki,K., Sharif,J., Shirane,K., Uranishi,K., Bogutz,A.B., Janssen,S.M., Suzuki,A., Okuda,A., Koseki,H. and Lorincz,M.C. (2021) Repression of germline genes by PRC1.6 and SETDB1 in the early embryo precedes DNA methylation-mediated silencing. *Nat. Commun.*, **12**, 7020.
56. Liu,X., Wang,C., Liu,W., Li,J., Li,C., Kou,X., Chen,J., Zhao,Y., Gao,H., Wang,H. *et al.* (2016) Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature*, **537**, 558–562.
57. Wang,C., Liu,X., Gao,Y., Yang,L., Li,C., Liu,W., Chen,C., Kou,X., Zhao,Y., Chen,J. *et al.* (2018) Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol.*, **20**, 620–631.

58. Shi,Y., Lan,F., Matson,C., Mulligan,P., Whetstine,J.R., Cole,P.A., Casero,R.A. and Shi,Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, **119**, 941–953.
59. Owen,D., Aguilar-Martinez,E., Ji,Z., Li,Y. and Sharrocks,A.D. (2023)ZMYM2 controls transposable element transcription through distinct co-regulatory complexes. bioRxiv doi: https://doi.org/10.1101/2023.01.24.525372, 24 January 2023, preprint: not peer-reviewed.
60. Tsusaka,T., Shimura,C. and Shinkai,Y. (2019) ATF7IP regulates SETDB1 nuclear localization and increases its ubiquitination. *EMBO Rep.*, **20**, e48297.
61. Tharp,M.E., Malki,S. and Bortvin,A. (2020) Maximizing the ovarian reserve in mice by evading LINE-1 genotoxicity. *Nat. Commun.*, **11**, 330.
62. Brind'Amour,J., Kobayashi,H., Richard Albert,J., Shirane,K., Sakashita,A., Kamio,A., Bogutz,A., Koike,T., Karimi,M.M., Lefebvre,L. *et al.* (2018) LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. *Nat. Commun.*, **9**, 3331.
63. Pontis,J., Planet,E., Offner,S., Turelli,P., Duc,J., Coudray,A., Theunissen,T.W., Jaenisch,R. and Trono,D. (2019) Hominoid-specific transposable elements and kzfps facilitate Human embryonic genome activation and control transcription in naive Human escs. *Cell Stem Cell*, **24**, 724–735.
64. Jachowicz,J.W., Bing,X., Pontabry,J., Boskovic,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.
65. 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.
66. Charlton,J., Jung,E.J., Mattei,A.L., Bailly,N., Liao,J., Martin,E.J., Giesselmann,P., Brandl,B., Stamenova,E.K., Muller,F.J. *et al.* (2020) TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nat. Genet*, **52**, 819–827.
67. van der Maarel,S.M., Scholten,I.H., Huber,I., Philippe,C., Suijkerbuijk,R.F., Gilgenkrantz,S., Kere,J., Cremers,F.P. and Ropers,H.H. (1996) Cloning and characterization of DXS6673E, a candidate gene for X-linked mental retardation in Xq13.1. *Hum. Mol. Genet*, **5**, 887–897.
68. Philips,A.K., Siren,A., Avela,K., Somer,M., Peippo,M., Ahvenainen,M., Doagu,F., Arvio,M., Kaariainen,H., Van Esch,H. *et al.* (2014) X-exome sequencing in Finnish families with intellectual disability–four novel mutations and two novel syndromic phenotypes. *Orphanet J. Rare. Dis.*, **9**, 49.
69. Uhlen,M., Fagerberg,L., Hallstrom,B.M., Lindskog,C., Oksvold,P., Mardinoglu,A., Sivertsson,A., Kampf,C., Sjostedt,E., Asplund,A. *et al.* (2015) Proteomics. Tissue-based map of the human proteome. *Science*, **347**, 1260419.