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

Adda-Lee Graham-Paquin[1,2,5], Deepak Saini[1,5], Jacinthe Sirois[1,2], Ishtiaque Hossain[1], Megan S. Katz[1], Qinwei Kim-Wee Zhuang[3,4], Sin Young Kwon[1], Yojiro Yamanaka[2,3], Guillaume Bourque[3,4], Maxime Bouchard[1,2,6,7], William A. Pastor[1,2,6]

[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
[5]Equal contribution
[6]Joint supervision
[7]Posthumous
e-mail: william.pastor@mcgill.ca

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. Interaction and colocalization data indicate that ZMYM2 homes to germline genes via binding to the non-canonical polycomb complex PRC1.6 and to transposons via the TRIM28 complex. *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 embryogenesis.

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 heterochromatization, 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, 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. 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 H3K9 methyltransferase SETDB1[8]. The TRIM28 complex is essential during development, as *Trim28*$^{-/-}$ mice show severe abnormality by E5.5[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]. Its MYM-type Zinc fingers are reported not to bind DNA[17], but ZMYM2 can bind to the post-translational modification SUMO2/3 both via a MYM-Zinc finger and via a distinct SUMO Interaction Motif (SIM)[17,18]. Many transcription factors and chromatin-bound complexes are SUMOylated, with SUMOylation often functioning to reduce transcriptional activity[19,20]. ZMYM2's best-characterized interaction is with the CoREST complex[15-17,21,22] which includes the histone deacetylases HDAC1 and HDAC2 and the histone demethylase LSD1[21]. Loss of ZMYM2 in HeLa cells results in reduced global association of CoREST with chromatin[17]. A plausible model of ZMYM2 activity is that it binds SUMOylated complexes on chromatin and recruits CoREST.

There is increasing evidence for an important role for ZMYM2 in transcriptional repression during embryonic development. ZMYM2 was identified as a factor important for suppressing totipotency in murine embryonic stem cells (mESCs), and zygotic depletion of ZMYM2 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 resulted in reactivation of the silenced copy of the imprinted gene PEG10[26]. 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, it's role and importance in mammalian embryogenesis remained unknown. Here we sought to determine the developmental role of ZMYM2 using *Zmym2*$^{-/-}$ mice and found a striking role for ZMYM2 in facilitating the methylation of germline genes and young LINE transposons.

RESULTS

Zmym2$^{-/-}$ mice show developmental abnormality and embryonic lethality

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

Most E8.5 *Zmym2*$^{-/-}$ embryos appeared normal (Extended Data Fig. 1d), 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 male and three female of each genotype (Supplementary Tables 1,2). Upon the loss of ZMYM2, knockout embryos exhibited an upregulation of 165 genes (fold change ≥ 4, q-value < 0.05) (Fig. 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 (Fig. 2b). Upregulation of select genes was confirmed through RT-qPCR of additional *Zmym2*$^{-/-}$ and *Zmym2*$^{+/+}$ E8.5 embryos (Extended Data Fig. 1e, Supplementary Table 3).

We observed two classes of upregulated genes in *Zmym2*$^{-/-}$ embryos. Ninety-five genes showed upregulation of transcription from an annotated promoter (Fig. 2c). By contrast, 60 genes showed clear evidence of splicing from an internal or upstream transposon (Fig. 2d, annotations in Supplementary Table 4). In 46 of these 60 fusion transcripts, the transposon in question was a LINE element of the L1MdT subclass. Both sense and antisense transcription from LINE promoters seeded transcription of fusion genes (Fig. 2d, Extended Data Fig. 2a-c). Interestingly, analysis of published data[27] 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 (Extended Data Fig. 2d-f).

We analyzed transposon expression in *Zmym2*$^{-/-}$ using the TEtranscript pipeline[28] and observed strong upregulation of L1MdT and the less abundant L1MdGf subclasses (Fig. 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[29]. Immunofluorescent 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 (Fig. 2f, Extended Data Fig. 2g). L1ORF1p is also highly expressed in *Zmym2*$^{-/-}$ trophoblast giant cells, indicating that ZMYM2 is also important for suppressing transposons in the placental lineage (Extended Data Fig. 2h).

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

As germline genes and transposons elements 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 1). 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 (Fig. 3a).

We observed no defect in global DNA methylation establishment in *Zmym2*$^{-/-}$ embryos (Fig. 3a). However, we observed a substantial reduction in DNA methylation over the promoters of the 95 upregulated genes expressed from annotated TSS (Fig. 2c, Fig. 3b,c). Hypomethylation was also observed globally over the promoters of full length L1MdT elements (Fig. 2d, Fig. 3d, Extended Data Fig. 2a-c). We compared gene expression changes in the *Zmym2*$^{-/-}$ mice 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*$^{-/-}$, consistent with their normally being silenced by DNA

methylation (Extended Data Fig. 3a). 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 (Extended Data Fig. 3a). L1MdT-fusion transcripts were also upregulated in *Dnmt3a*$^{-/-}$/*3b*$^{-/-}$, although not to quite the same extent as in *Zmym2*$^{-/-}$ (Extended Data Fig. 3a). Likewise, more overall transposon 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 (Extended Data Fig. 3b).

In further support of a role for ZMYM2 in promoting DNA methylation, ZMYM2 expression peaks during post-implantation development, coincident with *de novo* methylation (Extended Data Fig. 3c). 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 (Extended Data Fig. 3d).

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 5). 40.9% of these hypomethylated DMRs overlap with L1MdT elements, compared with 2.6% overlap expected by chance (Extended Data Fig. 4a). Other LINE elements, most notably L1MdGf, also overlapped with DMRs more than expected by chance (Extended Data Fig. 4b-d). This approach confirms ZMYM2's role in directing DNA methylation.

We then examined other regulatory mechanisms that involve 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 (Extended Data Fig. 5a). Canonical imprinted loci, which inherit methylation from parental gametes, did not show hypomethylation (Extended Data Fig. 5b), 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*) (Extended Data Fig. 5c), which is normally methylated and fully silenced during the peri-implantation period[30].

ZMYM2 methylates young LINE retrotransposons in human ESCs

We mined 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 (Fig. 3e). 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[31] (Extended Data Fig. 6a). Individual L1Hs elements are so similar to each other that we cannot assign individual RRBS reads to individual elements efficiently, but mapping to common consensus sequence showed that loss of *ZMYM2* or its interactor *ATF7IP* results in hypomethylation of the L1Hs promoter region in hESCs (Fig. 3f, Extended Data Fig. 6b,c). Strong hypomethylation of L1PA2 elements is also observed in *ZMYM2*$^{-/-}$ and *ATF7IP*$^{-/-}$ (Fig. 3g, Extended Data Fig. 6d). Thus, ZMYM2's role in methylating young transposons is conserved in humans.

ZMYM2 suppresses H3K4 methylation at target loci

We generated *Zmym2*$^{-/-}$ mESC lines on a V6.5 background (mixed C57BL6/129) (Extended Data Fig. 7a-c). These *Zmym2*$^{-/-}$ mESCs showed upregulation of germline genes (*Mael, Hormad1, Sohlh2*) and a transposon fusion (*L1MdT-Clca3b*) (Extended Data Fig. 7d). Upon differentiation to embryoid bodies (EBs), expression of the ZMYM2-target genes dropped substantially in *Zmym2*$^{+/+}$ mESCs but to a lesser degree in *Zmym2*$^{-/-}$, accentuating the relative upregulation (Extended Data Fig. 7d). A similar pattern was observed in a *Zmym2*$^{-/-}$ line from a pure 129/Sv genetic background donated by the lab of Dr. Jianlong Wang[16] (Extended Data Fig. 7e). Thus, *Zmym2*$^{-/-}$ mESCs and EBs are a plausible *in vitro* model with which to study ZMYM2 targeting and activity.

There is no evidence that ZMYM2 directly interacts with *de novo* methyltransferases, so we hypothesized an indirect mechanism by which ZMYM2 could promote DNA methylation. Transcriptional initiation promotes H3K4 methylation[32,33], which in turn is well established to antagonize binding of the *de novo* methyltransferases DNMT3A, DNMT3B and DNMT3L[34,35]. Impaired silencing in *Zmym2*$^{-/-}$ could thus lead to increased H3K4 methylation and impaired *de novo* DNA methylation. ChIP-seq of H3K4me3 and H3K4me2 in mESCs (d0) and day 3 EBs indeed indicated increased H3K4 methylation at ZMYM2 target genes as well as L1MdT elements in *Zmym2*$^{-/-}$ (Figure 3h – k, Extended Data Figure 8), suggesting a physical basis for the observed DNA hypomethylation phenotype.

ZMYM2 colocalizes with PRC1.6, TRIM28 and other repressive complexes

Because ZMYM2 lacks known capacity for sequence-specific binding and is thus likely to interact with chromatin via intermediary complexes, we analyzed published ChIP-seq data in mESCs to determine how ZMYM2 homes to germline genes and transposons. ZMYM2 is known to interact with the PRC1.6 complex[15,16], a non-canonical PRC1 complex which is critical for repression and methylation of germline genes during early embryonic development. TRIM28 is crucial for silencing transposons during pre- and peri-implantation development. While no direct interaction of ZMYM2 and TRIM28 has been observed, TRIM28 is heavily SUMOylated[20,36] and ZMYM2 could potentially interact with TRIM28 indirectly via ATF7IP[24]. Finally, we noted that a reported motif enriched at ZMYM2 binding sites[16] was similar to the motif for the transcription factor REST. We thus performed cluster analysis of all high-enrichment ZMYM2 peaks, incorporating ChIP-seq data for PRC1.6 components (MGA and L3MBTL2[7]), TRIM28[37], REST[38] and SUMO2[37].

Cluster 1 of ZMYM2 peaks corresponded to PRC1.6 peaks (Fig. 4a). Motif analysis of this cluster displayed strong enrichment for the E-Box and E2F motifs corresponding to MGA and E2F6 (Fig. 4b), the two PRC1.6 components which bind DNA in a sequence-specific manner. Cluster 1 was globally hypomethylated in *Zmym2*$^{-/-}$ (Fig. 4c) and includes promoters of germline genes upregulated in *Zmym2*$^{-/-}$ (Fig. 4d, e, Extended Data Fig. 9a-d). PRC1.6 components are heavily enriched over the promoters of genes upregulated in *Zmym2*$^{-/-}$ (Fig. 4f) and ZMYM2 is globally enriched over all PRC1.6 binding sites (Fig. 4g).

We performed ChIP-seq of the PRC1.6 component L3MBTL2 in control and *Zmym2*$^{-/-}$ mESCs and observed no defect in PRC1.6 recruitment upon loss of ZMYM2 at upregulated gene promoters (Fig. 4h). Likewise, there is no loss of H2AK119Ub (deposited by all PRC1 complexes) in mESCs or day3 EBs deficient for ZMYM2 (Extended Data Fig. 9e, f). Thus, ZMYM2 is not necessary for PRC1.6 binding or activity, but is essential at many target sites for repression and DNA methylation acquisition.

Cluster 2 and Cluster 4 displayed co-enrichment of ZMYM2 with TRIM28. These clusters included a diverse array of transposons. Globally, hypomethylation over these clusters is observed but modest (Extended Data Fig. 10a,b) and many target transposons in this cluster are not upregulated, indicating that while ZMYM2 strongly colocalizes with TRIM28, it is not always essential for target silencing. At L1MdT elements, where ZMYM2 is apparently essential, we observe strong ZMYM2, TRIM28 and SUMO2 binding (Fig. 4i, 4j, Extended Data Fig. 10c-f).

Cluster 3 consisted of REST binding sites and most peaks in the cluster showed canonical REST motifs (Fig. 4k). How ZMYM2 homes to REST bindings sites is unclear. ZMYM2 and REST both bind the CoREST complex but in a mutually exclusive manner[17], REST is not known to be SUMOylated, and physical interaction between the proteins has not been reported. Nonetheless, ZMYM2 is very strongly enriched at REST sites (Fig. 4a, Extended Data Fig. 11a-c). We observed no hypomethylation of Cluster 3 sites in *Zmym2*$^{-/-}$ (Extended Data Fig. 11d), so the biological consequences of ZMYM2's recruitment to REST sites is unclear.

Cluster 5 included weaker PRC1.6 and TRIM28 sites as well as sites that contained neither complex. Intriguingly, the top motif enriched in Cluster 5 corresponds to the DR0 nuclear receptor motif (Fig. 4l) suggesting that ZMYM2 may home to chromatin in some cases by binding a nuclear receptor. While a number of nuclear receptors can bind to the DR0 motif[39], it is noteworthy that Retinoic Acid Receptors (RARs) bind DR0 disproportionately in mESCs[40,41]. RARs and their dimerization partners RXRs can be SUMOylated[42,43] and mESCs treated with retinoic acid have an increased propensity to revert to the totipotent "2-cell" like state[44-46], a property shared with *Zmym2*$^{-/-}$ mESCs[16]. Likewise, RA signaling is critical during the totipotent 2-cell stage[46], and ZMYM2 has been reported to be essential for exit from totipotency[16]. It is possible that ZMYM2 suppresses totipotency, in mESCs and *in vivo*, partially by attenuating the activity of RAR transcription factors.

Finally, at SINE elements, we noticed largely TRIM28-independent recruitment of ZMYM2. Here, comparison with published data indicated colocalization with ADNP[47], a component of the repressive ChAHP complex and an interactor identified by BioID (Extended Data Figure 12). In total, we observe evidence that ZMYM2 is recruited to chromatin by at least five distinct complexes (PRC1.6, TRIM28, REST, ChAHP, a nuclear receptor). For the first two complexes, there is clear evidence of failure to silence and methylate target loci in *Zmym2*$^{-/-}$ mice.

DISCUSSION

ZMYM2 is clearly essential for silencing and DNA methylation of germline genes and young transposons in development. A potential model of ZMYM2 activity runs as follows. ZMYM2 binds a variety of chromatin-bound complexes including PRC1.6 and TRIM28. Where it binds, it helps mediate target silencing, potentially by its established ability to recruit the CoREST complex[16,17]. By blocking transcription, and potentially by the ability of the CoREST component LSD1 to mediate H3K4me1/2 demethylation, ZMYM2 reduces levels of H3K4 methylation, facilitating DNA methylation during the peri-implantation period and establishing stable silencing (Figure 5).

PRC1.6 and TRIM28 are heavily SUMOylated in mESCs[36] and are key ZMYM2 targets. While SUMOylation can perform many functions depending on context, there is a general tendency toward a silencing effect for SUMO[19,20]. ZMYM2 and possibly other ZMYM-family members may function as general readers of transcriptional silencing. At the same time, we

cannot rule out SUMO-independent recruitment of ZMYM2 to targets. 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 3f,g). One possibility is that ATF7IP's primary function at these loci is to recruit ZMYM2 to such elements. However, ATF7IP performs other functions such as promoting nuclear localization of the H3K9 methyltransferase SETDB1[48], 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 showed embryonic lethality amidst aberrant upregulation of LINE protein. It is difficult to determine what halts the development of the *Zmym2*$^{-/-}$ embryos, but it is well established that LINE elements can induce cell death, both by inducing anti-viral responses and by the endonuclease activity of L1ORF2p[49].

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[50,51]. 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 pre-implantation embryos. We do not know how many of the 46 LINE-transcript fusions we observe code for functional protein or whether any have biological function, but it is intriguing that L1MdT expression during pre-implantation murine development is essential for developmental progression[52,53].

It is important to consider the phenotypes we do not see in *Zmym2*$^{-/-}$ but might have expected based on literature. We do not see pre-implantation developmental arrest[16], but this can be explained by the presence of intact *Zmym2* RNA from *Zmym2*$^{+/-}$ mothers. We also do not observe failure to exit pluripotency[23,25], with E8.5 *Zmym2*$^{-/-}$ having undergone gastrulation normally. ESCs adapted to continuous pluripotent culture may be more resistant to differentiation than developing 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[31]. 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[54], 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, 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[55,56], and all six ZMYM-family transcription factors are widely

expressed in human tissues[57]. The role of ZMYM-family transcription factors as cofactors for repression may be widespread and underappreciated.

FIGURE CAPTIONS

Figure 1. *Zmym2*$^{-/-}$ embryos are embryonic lethal by E10.5 and show 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= 500µm. **c,** Distribution of E9.5 phenotype class 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.

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 ≥ 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. 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. 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 ≥ 4, q-value < 0.05) are coloured red. **f,** Immunofluorescence staining of LINE-1 ORFp (red) and DAPI (blue) in E9.5 *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ embryonic tissue. Note that the few, bright red cells in *Zmym2*$^{+/+}$ are red blood cells, which autofluoresce. Scale bars= 500µm.

Figure 3. Impaired DNA methylation in *Zmym2*$^{-/-}$ a, Dot 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 genes upregulated in *Zmym2*$^{-/-}$ embryos. Transposon fusion genes are excluded. **c,** Heatmap of DNA methylation of the TSS of 94 upregulated genes in *Zmym2*$^{-/-}$ expressed from annotated TSS. Average methylation within 500bp of transcription start site of the 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 (>6kb) L1MdT elements in the genome. **e.** Volcano plot of differentially expressed transposons comparing *ZMYM2*$^{+/+}$ and *ZMYM2*$^{-/-}$ hESCs. Significant differentially expressed transposons (fold ≥ 2, q-value < 0.05) are coloured red. **f,** A boxplot of average CpG methylation across 1Kb start regions of full length L1Hs transposable elements. **g,** A boxplot of average CpG methylation across L1PA2 transposable elements. **h – k,** Metaplot of H3K4me3 (**h, j**) and H3K4me2 (**i, k)** over the TSS of 94 upregulated genes in *Zmym2*$^{-/-}$ expressed from annotated TSS (**h,i**) or L1MdT elements **(j,k)** in *Zmym2*$^{-/-}$ (KO) and control (Ctrl) mESCs. To normalize for ChIP quality, enrichment is normalized based on peak height for all annotated genes, with two biological replicates merged.

Figure 4. ZMYM2 is recruited to chromatin by a variety of distinct complexes. **a,** Heatmaps of ZMYM2, L3MBTL2, MGA, REST, SUMO2, and TRIM28 enrichment over ZMYM2 ChIP-seq peak summits and the flanking 3Kb regions (n = 21,549 peaks). Peaks were partitioned into five clusters using a k-means algorithm. Peak numbers: Cluster 1, 290 peaks; Cluste 2, 931 peaks; Cluster 3, 796 peaks; Cluster 4, 4080 peaks; Cluster 5, 14969 peaks. Cluster 1 peaks are shown at twice the width of other peaks to enhance visibility. **b,** HOMER motif analysis of Cluster 1 ZMYM2 peaks show enrichment of E-Box and E2F motifs. **c,** Metaplot of CpG methylation spanning Cluster 1 ZMYM2 peaks and the flanking 3Kb region. **d,e,** ChIP-seq and DNA methylation data over the *Ddx4* (d) and *D1Pas* (e) loci. Note colocalization of ZMYM2 with PRC1.6 components and hypomethylation in *Zmym2*$^{-/-}$. **f,** A 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). **g,** Metaplot of ZMYM2 ChIP-seq signal over L3MBTL2 peaks. Dashed line indicates ChIP input. **h,** Metaplot of L3MBTL2 CUT&RUN ChIP-seq signal over the annotated TSS of upregulated genes in both *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ mESCs. **i,** Metaplot of ZMYM2, TRIM28, SUMO2 ChIP-seq over all full length L1MdT elements. Dashed line indicates ChIP input. **j,** ChIP-seq and DNA methylation data over a representative LINE element. Note overlap of TRIM28 and ZMYM2 peaks, and hypomethylation over the peak site. **k**, HOMER motif analysis of Cluster 3 ZMYM2 peaks show enrichment of REST-binding motif. **l,** HOMER motif analysis of Cluster 5 ZMYM peaks show enrichment of DR0 nucleic acid motif.

Figure 5. Proposed mechanism for ZMYM2-mediated DNA methylation. ZMYM2 is recruited to PRC1.6 and TRIM28 binding sites respectively, potentially 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.

Extended Data Figure 1. Reduced size and developmental delays in *Zmym2*$^{-/-}$ mice and confirmation of gene dysregulation. a, Schematic representation of point mutation and truncation point in *Zmym2*$^{-}$ allele. **b,** Relative size of embryos measured as 2D area in embryos in sagittal position normalized by litter as a ratio to average *Zmym2*$^{+/+}$ area. p-value calculated on basis of one-tailed t-test. **c,** Somite counts of E9.5 embryos per genotype. Each litter is represented in a different color to account for litter variability. Note that *Zmym2*$^{-/-}$ mice show reduced somite count within each litter. p-value calculated on basis of one-tailed t-test. **d,** Images of E8.5 *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ mice. **e,** Quantitative reverse-transcription PCR (qRT-PCR) of four upregulated genes expressed from promoter (*Ddx4*, *Tdrd1*, *Sohlh2*, *Hormad1*) and LINE-fusions (*Gc*, *Reg2*). Each dot reflects the ratio of expression in KO/WT littermate embryos, with expression normalized to *RPS16.*

Extended Data Figure 2. Expression of L1MdT-gene fusion transcripts in *Zmym2*$^{-/-}$ embryos and early embryogenesis. a-c, Representative examples of L1MdT-gene fusion transcripts. Note examples of bidirectional transcription of an internal L1MdT element (**a**), splicing from an upstream L1MdT in a forward direction into a gene (**b**) and splicing from an internal L1MdT promoter in a forward direction into a gene (**c**). Hypomethylated DMRs are indicated in light blue. **d, e,** Representative examples of L1MdT-gene fusion transcript

expression in early embryogenesis. **f,** Box plot of L1MdT-gene fusion transcripts during early mouse embryonic development. In **d-f,** 8-cell, E3.5 and E6.5 RNA-seq data is from Smith 2017. **g,** Immunofluorescent staining of LINE-1 ORFp (red) in E8.5 *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ embryos. Regions of mesoderm and ectoderm in inset are indicated. Scale bar= 20µm **h,** Immunofluorescent staining of LINE-1 ORFp (red) and DAPI (blue) of E9.5 *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ implantation sites. TGC= trophoblast giant cell, MD= maternal decidua. Small red cells visible in both samples are maternal blood cells. Scale bars= 100µm.

Extended Data Figure 3. Hypomethylation of germline genes in *Zmym2*$^{-/-}$. a, Expression of each gene in the given category is shown (KO/WT) in *Zmym2*$^{-/-}$ and *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$. Note that most genes upregulated in *Zmym2*$^{-/-}$ are also upregulated in the *Dnmt* double knockout. **b,** Volcano plot of differentially expressed transposable elements comparing *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$ and control E8.5 embryos. Significant differentially expressed transposable elements (fold ≥ 2, q-value < 0.05) are coloured red. **c,** Expression of ZMYM2 and *de novo* methyltransferases DNMT3A and DNMT3B during early embryogenesis. **d,** Expression of the 95 upregulated genes expressed from TSS during early embryogenesis. In **c** and **d**, 8-cell, E3.5 and E6.5 RNA-seq data is from Smith 2017.

Extended Data Figure 4. Overlap of DMRs with L1Md elements. A Monte-Carlo simulation was undertaken in which the positions of DMRs were randomly placed in the genome and repeated 100,000 times. The overlap with indicated transposon class is indicated as "Random overlap". This number is contrasted with the "observed overlap" number for actual DMRs.

Extended Data Figure 5. Normal X-inactivation, imprint methylation in *Zmym2*$^{-/-}$. a, Metaplot of DNA methylation over X-chromosomal genes in *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$. **b,** DNA methylation over known murine imprinted regions in *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$. **c,** Hypomethylation, and aberrant transcription from the *Liz/Zdbf2* transient imprint in *Zmym2*$^{-/-}$. Note enrichment of ZMYM2 at the target locus.

Extended Data Figure 6. Upregulation and hypomethylation of transposons in hESCs a, Relative expression of transposon families in naïve and primed hESCs. Note that naïve hESCs show only a very modest increase in L1Hs and L1PA2 levels. **b,** Methylation level of individual CpGs in an L1HS consensus element for the genotype indicated. Only CpGs with >10-fold coverage in the Bar 2021 RRBS data are shown. **c,d** DNA methylation of L1Hs (**c**) and L1PA2 (**d**) control and *ATF7IP*$^{-/-}$ hESCs in B105 hESCs. Note similarity to Figure 4b,c in this second hESC line.

Extended Data Figure 7. Recapitulation of ZMYM2-mediated silencing in an Embryoid body (EB) assay a, Western blots of control and *Zmym2*$^{-/-}$ clonal mESC lines generated on a V6.5 background (left) as well as WT J1 and *Zmym2*$^{-/-}$ mESCs generated by Yang and colleagues. **b,** Immunofluorescent staining of control and *Zmym2*$^{-/-}$ clonal line for ZMYM2. **c,** Light microscopy images of *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ ESCs and EBs. **d,** Relative expression of the pluripotency factor *Pou5f1*, differentiation genes, and ZMYM2-target genes in three *Zmym2*$^{-/-}$ as compared with three control lines, during two separate EB differentiation experiments (n=6 replicates total). **e,** Relative expression of pluripotency, differentiation, and ZMYM2-target genes in control J1 mESCs and *Zmym2*$^{-/-}$ mESCs generated by Yang and colleagues, during EB

differentiation. In **d** and **e,** * indicates $p < 0.05$, ** $p < 0.01$, *** $p< 0.001$, **** $p< 0.0001$, by a two-tailed t-test.

Extended Data Figure 8. Reduced enrichment of H3K4me2 and H3K4me3 over ZMYM2 target loci. Metaplot of H3K4me3 (**a, c**) and H3K4me2 (**b, d)** over the TSS of 94 upregulated genes in *Zmym2*$^{-/-}$ expressed from annotated TSS (**a,b**) or L1MdT elements **(c,d)**. This figure shows individual replicates that were merged in Figure 3h-k. NT=non-targeting sgRNA, indicated as Ctrl in Figure 3. To normalize for ChIP quality, enrichment is normalized based on peak height for all annotated genes.

Extended Data Figure 9. Hypomethylation of PRC1.6 targets in *Zmym2*$^{-/-}$. a-d, Four examples of Cluster 1, PRC1.6-bound genes upregulated in *Zmym2*$^{-/-}$. Note colocalization of ZMYM2 and PRC1.6 targets, as well as hypomethylation in *Zmym2*$^{-/-}$. Hypomethylated DMRs are indicated with light blue boxes. **e, f,** Relative ChIP enrichment of H2AK119Ub in *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ mESCs (**e**) and day 3 EBs (**f**) over genes upregulated in *Zmym2-/-* (transposon-fusion genes excluded). Note that there is no difference between *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$. The modestly increased input enrichment in *Zmym2*$^{-/-}$ may reflect increased locus accessibility as a result of impaired silencing.

Extended Data Figure 10. Hypomethylation of transposons in *Zmym2*$^{-/-}$. a, b, Metaplot of DNA methylation over Cluster 2 and Cluster 4 peaks for *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$. **c,d,** ZMYM2, TRIM28, SUMO2 ChIP-seq and WGBS data plotted over representative L1MdT elements. Note correspondence of the three proteins at L1MdT promoters and corresponding reduction in DNA methylation. **e,** Metaplot of TRIM28, SUMO2 and ZMYM2 mESC ChIP-seq signal over DMRs hypomethylated in *Zmym2*$^{-/-}$ which overlap with L1MdT elements. Dashed lines indicate ChIP input. **f.** Heatmap of input-normalized ZMYM2, TRIM28, SUMO2 ChIP-seq data over hypomethylated L1MdT DMRs.

Extended Data Figure 11. Overlap of ZMYM2 and REST binding sites. a, ZMYM2 plotted over REST binding sites. **b, c,** Examples of overlap of REST and ZMYM2 binding sites. **d,** DNA methylation over Cluster 3 peaks in *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$.

Extended Data Figure 12. Overlap of ZMYM2 and ChAHP complex binding sites. a,b, Examples of overlap of ADNP and ChAHP complex binding sites. Note co-enrichment of ADNP, ZMYM2 and SUMO over indicated SINE elements, with no corresponding TRIM28 enrichment.

Supplementary Table 1. Mapping statistics for next generation sequencing data generated or used in the course of this project.

Supplementary Table 2. RPKMs of all genes in control and *Zmym2*$^{-/-}$ mice.

Supplementary Table 3. Sequences of primers used in this project.

Supplementary Table 4. Annotations of upregulated genes in *Zmym2*$^{-/-}$ mice.

Supplementary Table 5. Locations of DMRs in *Zmym2*$^{-/-}$ mice.

MATERIALS & 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 embryonic day 0.5 (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 of 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 10-µm thick sections as described[58]. Paraffin-embedded samples were processed by the GCI Histology Core Facility and 6µm serial sections were obtained.

Immunofluorescent staining of tissue sections

Immunofluorescence analyses were performed as described[59]. Anti-LINE-1 ORF1p (Abcam cat#ab216324) was used at 1:100 dilution in PBS.

Embryonic Stem Cell Culture Conditions

Both wild-type and *Zmym2*$^{-/-}$ mESCswere generated from the V6.5 line (C57Bl/6 x 129/sv). Cells were cultured on 0.1% gelatin (EmbryoMax, ES-006-B) coated platea and media was refreshed every day. Cells were passaged every 2 days using 0.05% Trypsin-EDTA (Gibco, 25300062). These cells were cultured in Knockout™ D-MEM (Gibco, 10829018) media supplemented with 1000U/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.

Immunofluorescent staining of mESCs

mESCs were cultured on glass coverslips coated with 0.1% gelatin. Cells were then fixed with 4% paraformaldehyde for 15 minutes at room temperature and washed twice with 1xPBS. The mESCs were then permeabilized and blocked using 0.1% Triton-X diluted in 5% solution of donkey serum for 15 minutes at room temperature and washed twice with 1xPBS-T (0.1% Tween20 in 1xPBS). 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 1xPBS-T and secondary antibody and DAPI fluorophores were added (Donkey anti-Rabbit Alexa Fluor™ 488, A-21206 as a 1:500 dilution) for 45 minutes at room temperature. Cells were then washed twice with PBS-T and coverslips were mounted on microscopy slides using ProLong Gold. Cells were imaged on the LSM710 confocal microscope.

Genetic Ablation of Zmym2 by CRISPR-Cas9 Nucleofections
Zmym2$^{-/-}$ mESCs 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 transfect ribonuclear particles within cells. 300K cells were diluted in P3 primary cell solution prior to nucleofection and were hastily passaged to pre-warmed serum-LIF media afterwards. 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.

RNA Isolation, Quantitative PCR and RNA sequencing of embryos
qPCR: Total mRNA was extracted from dissected embryos using a RNeasy micro kit (Qiagen-CAT# 74084) or AllPrep DNA/RNA Micro Kit (Qiagen-CAT# 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. Filtered reads were then aligned to the mm10 reference genome using STAR(v2.7.8a) using 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 the 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 less than 0.05 were selected as significant differentially expressed genes. Gene ontology over-representation was determined using clusterProfiler.

The TEtranscript pipeline was used to determine differential expression of transposons classes in Zmym2$^{-/-}$ E8.5 embryos. RNA-seq reads were first trimmed using Trim Galore! (v0.6.6) using 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.

Reverse-transcriptase quantitative PCR
Total mRNA was isolated from both wild-type, non-target and *Zmym2*$^{-/-}$ mESCs using RNAzol RT (Molecular Research Center Inc, RN 190) and first-strand cDNA was generated using the SensiFast cDNA Synthesis kit (FroggaBio, DD-BIO-65053). Quantitative PCR was then performed using PowerUP SYBR™ Green PCR (Invitrogen, A25742) on the Quantstudio 5 (Applied Biosystems) apparatus with the following cycling conditions: 50 °C 2 minutes, 95 °C

20 seconds, 45x (95 °C 3 seconds, 60 °C 30 seconds), 95 °C 1 second). The qPCR reaction was performed using 1x concentration of PowerUP SYBR Green Master Mix, cDNA from 5ng starting mRNA, and 0.5 µM of primer mix in a total of 7µL reaction. The expression of ZMYM2 target genes were normalized to the housekeeping gene GAPDH and the sequence of primer used for qPCR are provided in Supplementary Table 3.

Western Blot

Protein was extracted using ice cold 1xRIPA lysis buffer supplemented with fresh protease inhibitors (1 mM phenylmethylsulfonyl fluoride, 10mM sodium fluoride and 1 mM sodium orthovanadate). Cell pellets were the subjected to 5 cycles of freeze-thawing using liquid nitrogen to ensure complete breakage of both the cell and nuclear membrane. Lysate protein concentration was measured using a Bradford Assay and 30-40µg 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 in primary antibody in 1xOdyssey Blocking Buffer 0.15% Tween-20 at 4°C. Anti-ZMYM2 antibody (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 three x 5 minutes in PBS supplemented with 0.1% Tween-20 and then incubated in a solution of secondary antibodies (LI-COR IRDye 680RD, 1:20,000 dilution) diluted in Odyssey Blocking Buffer and 0.15% Tween-20 for 1 hour at room temperature. Membranes were then washed for 5 minutes twice in PBS 0.1% Tween and left 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 genomic DNA (500 ng) from E8.5 mouse embryos (QIAamp DNA mini kit, 51304 or AllPrep DNA/RNA Micro Kit, Qiagen, 82084) and spike-in lambda DNA (1.25 ng) (New England Biolabs) were 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 (Zmyo Research, D5005) using manufacturer's instructions. Up to 100 ng of genomic DNA was then used to generate WGBS libraries with the Accel-NGS Methyl-Seq DNA Library prep kit (Swift Biosciences, 30024) according with 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 using the NovaSeq 6000 at the Center of 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)using default parameters. Aligned reads were then deduplicated and reads were filtered for incomplete bisulfite conversion. Methylation calling over cytosines was then done using Bismark and overall DNA methylation was then 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 3Kb

flanking regions. Differentially methylated regions 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.

ChIP of Histone Modifications

Both wild-type and *Zmym2*$^{-/-}$ mESCs or EBs were crosslinked in 1% paraformaldehyde for 10 minutes and quenched with the addition of 1 mM of glycine for 10 minutes to stop crosslinking. Cells were then lysed 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 minutes and Temperature = 7 °C). The sheared DNA was then pre-cleared using magnetic beads (Sera-Mag Protein A/G SpeedBeads, VWR 17152104010150) for 1 hour to remove non-specific binding. Using a magnetic rack, the sonicated lysate was then separated to a new tube and primary antibodies was added (anti-H2AK119Ub Cell Signaling D27C, anti-H3K4me3 Cell Signaling C42D8, anti-H3K4me2 Cell Signaling C64G9) and incubated overnight. Magnetic beads were then washed and added to the sonicated lysate and incubated for 2 hours 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% TritonX-100, 0.1% Deoxycholate, 1mM EDTA, 140 mM NaCl and Wash Buffer B: 50 mM HEPES, 0.1% SDS, 1% TritonX-100, 0.1% Deoxycholate, 1 mM EDTA, 500 mM NaCl) and TE buffer. Protein/DNA complexes of interest were finally eluted in elution buffer (50 mM Tris-HCl, 1 mM EDTA, 1% SDS) incubated at 65 °C for 10 minutes 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 PCR Purification Kit using manufacturer's instructions.

To determine the localization of L3MBTL2 in Zmym2$^{+/+}$ and Zmym2$^{-/-}$ mESCs, 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 kit 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). MACS2 was then used for peak identification using default parameters and inputs as controls. Aligned BAM files were used to generate the bigwigs using DeepTools. The resulting bigwigs were used for metaplot and heatmaps using the computeMatrix and plotHeatmap functions in DeepTools.

Metaplots of histone enrichment were normalized over promoters of upregulated genes and L1MdT elements. This was done by measuring the median ChIP signal over the genomic background ("background signal") and the average peak chip signal over the promoters all genes ("peak signal") for each sample. A scaling factor was then determined by subtracting the background signal from the average peak signal. The individual ChIP sample signals in regions of interest were then normalized by subtracting the background signal and then divided by the scaling factor.

ACKNOWLEDGEMENTS
We thank the Genome Quebec Innovation Centre, 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 for donating *Zmym2*$^{-/-}$ cells. The research was funded by CIHR Project Grant PJT-159768 (M.B), CIHR Project Grant PJT-166169 (W.P.), the New Frontiers in Research Fund Grant NFRFE-2018–00883 (W.P.) and the NSERC Grant RGPIN-2018-04856 (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 Chercheur-boursier de mérite. W.P. was supported by an FRQS Chercheur-boursier.

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.

1. Baubec, T., Ivanek, R., Lienert, F. & Schubeler, D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153**, 480-92 (2013).
2. Smith, Z.D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339-44 (2012).
3. Bestor, T.H., Edwards, J.R. & Boulard, M. Notes on the role of dynamic DNA methylation in mammalian development. *Proc Natl Acad Sci U S A* (2014).
4. Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* **42**, 1093-100 (2010).
5. Hargan-Calvopina, J. *et al.* Stage-Specific Demethylation in Primordial Germ Cells Safeguards against Precocious Differentiation. *Dev Cell* **39**, 75-86 (2016).
6. Pohlers, M. *et al.* A role for E2F6 in the restriction of male-germ-cell-specific gene expression. *Curr Biol* **15**, 1051-7 (2005).
7. Stielow, B., Finkernagel, F., Stiewe, T., Nist, A. & Suske, G. MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *Plos Genetics* **14**(2018).
8. Iyengar, S. & Farnham, P.J. KAP1 protein: an enigmatic master regulator of the genome. *J Biol Chem* **286**, 26267-76 (2011).
9. Rowe, H.M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237-40 (2010).
10. Rowe, H.M. *et al.* De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* **140**, 519-29 (2013).
11. Tao, Y. *et al.* TRIM28-Regulated Transposon Repression Is Required for Human Germline Competency and Not Primed or Naive Human Pluripotency. *Stem Cell Reports* **10**, 243-256 (2018).
12. Rousseaux, M.W. *et al.* Depleting Trim28 in adult mice is well tolerated and reduces levels of alpha-synuclein and tau. *Elife* **7**(2018).

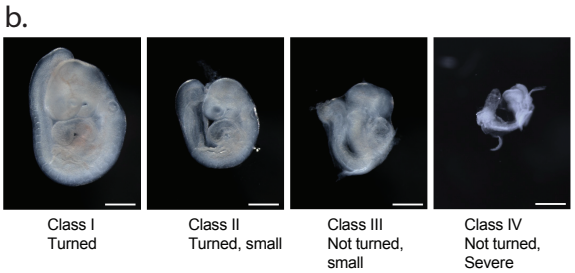
13. Dahlet, T. *et al.* Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity. *Nat Commun* **11**, 3153 (2020).
14. Aguilar-Martinez, E. *et al.* Screen for multi-SUMO-binding proteins reveals a multi-SIM-binding mechanism for recruitment of the transcriptional regulator ZMYM2 to chromatin. *Proc Natl Acad Sci U S A* **112**, E4854-63 (2015).
15. Connaughton, D.M. *et al.* Mutations of the Transcriptional Corepressor ZMYM2 Cause Syndromic Urinary Tract Malformations. *Am J Hum Genet* **107**, 727-742 (2020).
16. Yang, F. *et al.* DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces a Totipotent 2C-like State. *Cell Stem Cell* **26**, 234-250 e7 (2020).
17. Gocke, C.B. & Yu, H. ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. *PLoS One* **3**, e3255 (2008).
18. Guzzo, C.M. *et al.* Characterization of the SUMO-binding activity of the myeloproliferative and mental retardation (MYM)-type zinc fingers in ZNF261 and ZNF198. *PLoS One* **9**, e105271 (2014).
19. Gill, G. Something about SUMO inhibits transcription. *Curr Opin Genet Dev* **15**, 536-41 (2005).
20. Hendriks, I.A. & Vertegaal, A.C. A comprehensive compilation of SUMO proteomics. *Nat Rev Mol Cell Biol* **17**, 581-95 (2016).
21. Shi, Y.J. *et al.* Regulation of LSD1 histone demethylase activity by its associated factors. *Mol Cell* **19**, 857-64 (2005).
22. Hakimi, M.A., Dong, Y., Lane, W.S., Speicher, D.W. & Shiekhattar, R. A candidate X-linked mental retardation gene is a component of a new family of histone deacetylase-containing complexes. *J Biol Chem* **278**, 7234-9 (2003).
23. Hackett, J.A. *et al.* Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nat Commun* **9**, 4292 (2018).
24. Tsusaka, T., Fukuda, K., Shimura, C., Kato, M. & Shinkai, Y. The fibronectin type-III (FNIII) domain of ATF7IP contributes to efficient transcriptional silencing mediated by the SETDB1 complex. *Epigenetics Chromatin* **13**, 52 (2020).
25. Lezmi, E. *et al.* The Chromatin Regulator ZMYM2 Restricts Human Pluripotent Stem Cell Growth and Is Essential for Teratoma Formation. *Stem Cell Reports* **15**, 1275-1286 (2020).
26. Bar, S. *et al.* Identifying regulators of parental imprinting by CRISPR/Cas9 screening in haploid human embryonic stem cells. *Nat Commun* **12**, 6718 (2021).
27. Smith, Z.D. *et al.* Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**, 543-547 (2017).
28. Jin, Y., Tam, O.H., Paniagua, E. & Hammell, M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593-9 (2015).
29. Sookdeo, A., Hepp, C.M., McClure, M.A. & Boissinot, S. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob DNA* **4**, 3 (2013).
30. Greenberg, M.V. *et al.* Transient transcription in the early embryo sets an epigenetic state that programs postnatal growth. *Nat Genet* **49**, 110-118 (2017).
31. Pastor, W.A. *et al.* Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. *Cell Stem Cell* **18**, 323-9 (2016).

32. Krogan, N.J. *et al.* The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* **11**, 721-9 (2003).
33. Lee, J.H. & Skalnik, D.G. 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**, 609-18 (2008).
34. Noh, K.M. *et al.* Engineering of a Histone-Recognition Domain in Dnmt3a Alters the Epigenetic Landscape and Phenotypic Features of Mouse ESCs. *Mol Cell* **59**, 89-103 (2015).
35. Otani, J. *et al.* Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep* **10**, 1235-41 (2009).
36. Cossec, J.C. *et al.* SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. *Cell Stem Cell* **23**, 742-757 e8 (2018).
37. Yang, B.X. *et al.* Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* **163**, 230-45 (2015).
38. McGann, J.C. *et al.* Polycomb- and REST-associated histone deacetylases are independent pathways toward a mature neuronal phenotype. *Elife* **3**, e04235 (2014).
39. Wagner, R.T. & Cooney, A.J. Minireview: the diverse roles of nuclear receptors in the regulation of embryonic stem cell pluripotency. *Mol Endocrinol* **27**, 864-78 (2013).
40. Moutier, E. *et al.* Retinoic acid receptors recognize the mouse genome through binding elements with diverse spacing and topology. *J Biol Chem* **287**, 26328-41 (2012).
41. Chatagnon, A. *et al.* RAR/RXR binding dynamics distinguish pluripotency from differentiation associated cis-regulatory elements. *Nucleic Acids Res* **43**, 4833-54 (2015).
42. Zhu, L., Santos, N.C. & Kim, K.H. Small ubiquitin-like modifier-2 modification of retinoic acid receptor-alpha regulates its subcellular localization and transcriptional activity. *Endocrinology* **150**, 5586-95 (2009).
43. Choi, S.J. *et al.* Negative modulation of RXRalpha transcriptional activity by small ubiquitin-related modifier (SUMO) modification and its reversal by SUMO-specific protease SUSP1. *J Biol Chem* **281**, 30669-77 (2006).
44. Sharova, L.V. *et al.* Emergence of undifferentiated colonies from mouse embryonic stem cells undergoing differentiation by retinoic acid treatment. *In Vitro Cell Dev Biol Anim* **52**, 616-24 (2016).
45. Tagliaferri, D. *et al.* Retinoic Acid Induces Embryonic Stem Cells (ESCs) Transition to 2 Cell-Like State Through a Coordinated Expression of Dux and Duxbl1. *Front Cell Dev Biol* **7**, 385 (2019).
46. Iturbide, A. *et al.* Retinoic acid signaling is critical during the totipotency window in early mammalian development. *Nat Struct Mol Biol* **28**, 521-532 (2021).
47. Ostapcuk, V. *et al.* Activity-dependent neuroprotective protein recruits HP1 and CHD4 to control lineage-specifying genes. *Nature* **557**, 739-743 (2018).
48. Tsusaka, T., Shimura, C. & Shinkai, Y. ATF7IP regulates SETDB1 nuclear localization and increases its ubiquitination. *EMBO Rep* **20**, e48297 (2019).
49. Tharp, M.E., Malki, S. & Bortvin, A. Maximizing the ovarian reserve in mice by evading LINE-1 genotoxicity. *Nat Commun* **11**, 330 (2020).
50. Brind'Amour, J. *et al.* LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. *Nat Commun* **9**, 3331 (2018).

51. Pontis, J. *et al.* Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell* **24**, 724-735 e5 (2019).
52. Jachowicz, J.W. *et al.* LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat Genet* **49**, 1502-1510 (2017).
53. Percharde, M. *et al.* A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **174**, 391-405 e19 (2018).
54. Charlton, J. *et al.* TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nat Genet* **52**, 819-827 (2020).
55. van der Maarel, S.M. *et al.* Cloning and characterization of DXS6673E, a candidate gene for X-linked mental retardation in Xq13.1. *Hum Mol Genet* **5**, 887-97 (1996).
56. Philips, A.K. *et al.* X-exome sequencing in Finnish families with intellectual disability--four novel mutations and two novel syndromic phenotypes. *Orphanet J Rare Dis* **9**, 49 (2014).
57. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
58. Nguyen, A.H. *et al.* Gata3 antagonizes cancer progression in Pten-deficient prostates. *Hum Mol Genet* **22**, 2400-10 (2013).
59. Narlis, M., Grote, D., Gaitan, Y., Boualia, S.K. & Bouchard, M. Pax2 and pax8 regulate branching morphogenesis and nephron differentiation in the developing kidney. *J Am Soc Nephrol* **18**, 1121-9 (2007).

a.

	Expected	E8.5 n=67	E9.5 n=34	E10.5 n=32	E11.5 n=26
+/+	25%	26.8	29.4	31.25	38.4
+/-	50%	49.4	44.2	56.25	38.4
-/-	25%	23.8	26.4	0	0
Resorbed		0	0	12.5	23.2



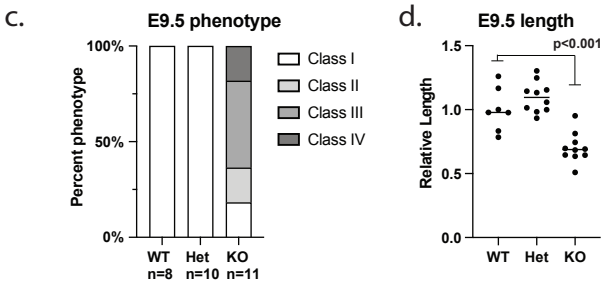


Figure 1. *Zmym2*$^{-/-}$ embryos are embryonic lethal by E10.5 and show variable developmental delays at E9.5. a, Embryonic viability by genotype from embryonic day E8.5 to E11.5 compared to the expected Mendelian ratios from *Zmym2*$^{+/-}$; *Zmym2*$^{+/-}$ crosses. Each embryonic age includes pooled data from 3+ litters. **b,** Images taken of E9.5 *Zmym2*$^{-/-}$ embryos representing four different classifications of phenotype severity. Scale bars= 500µm. **c,** Distribution of E9.5 phenotype class 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.

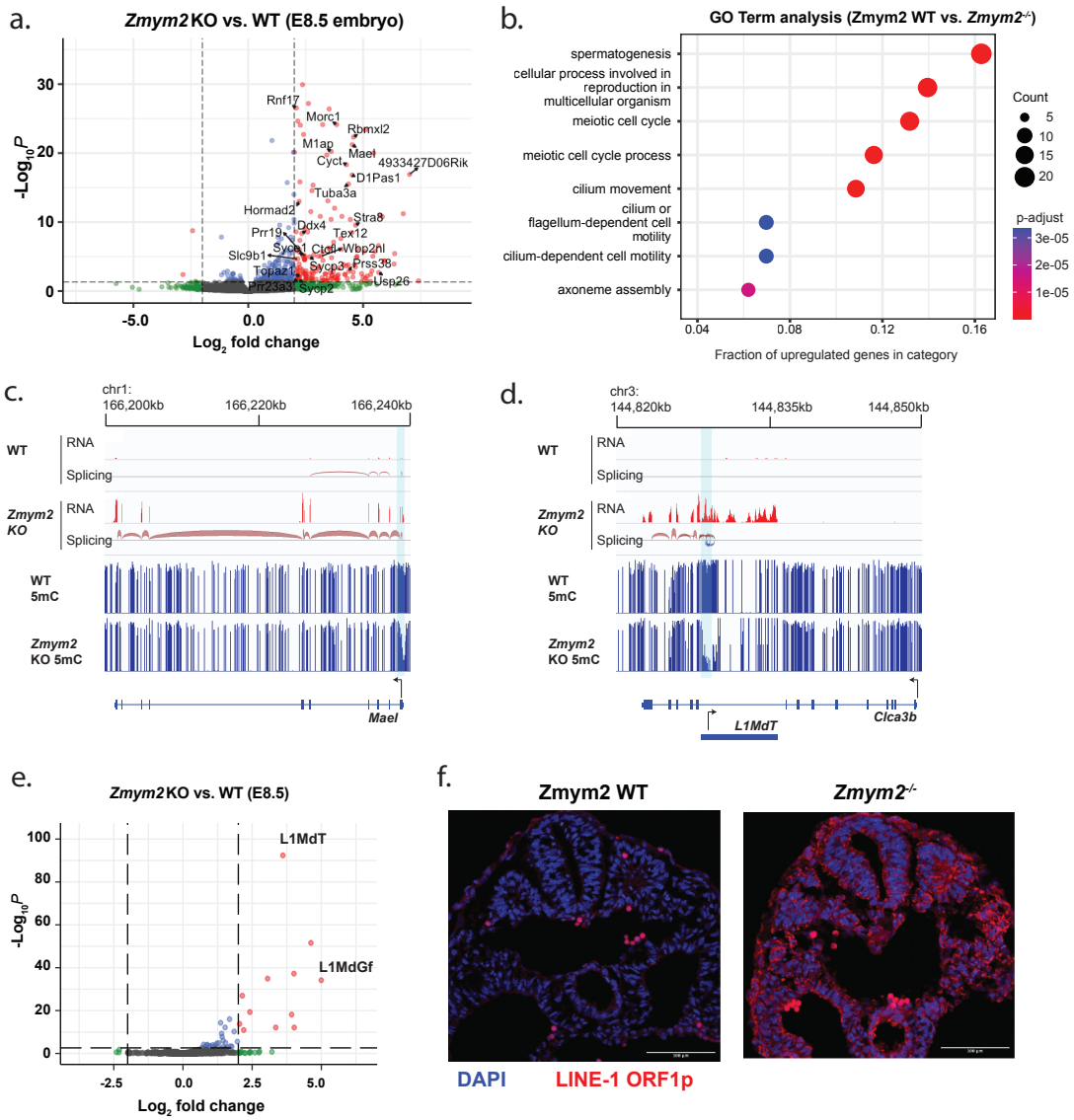


Figure 2. Germline genes and young LINE elements are upregulated in *Zmym2*$^{-/-}$ embryos. **a,** Volcano plot of differentially expressed genes comparing *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ E8.5 embryos. Significant differentially expressed genes (fold ≥ 4, q-value < 0.05) are expressed in red. Germline 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. 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. 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 ≥ 4, q-value < 0.05) are coloured red. **f,** Immunofluorescence staining of LINE-1 ORFp (red) and DAPI (blue) in E9.5 *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ embryonic tissue. Note that the few, bright red cells in *Zmym2*$^{+/+}$ are red blood cells, which autofluoresce. Scale bars= 500μm.

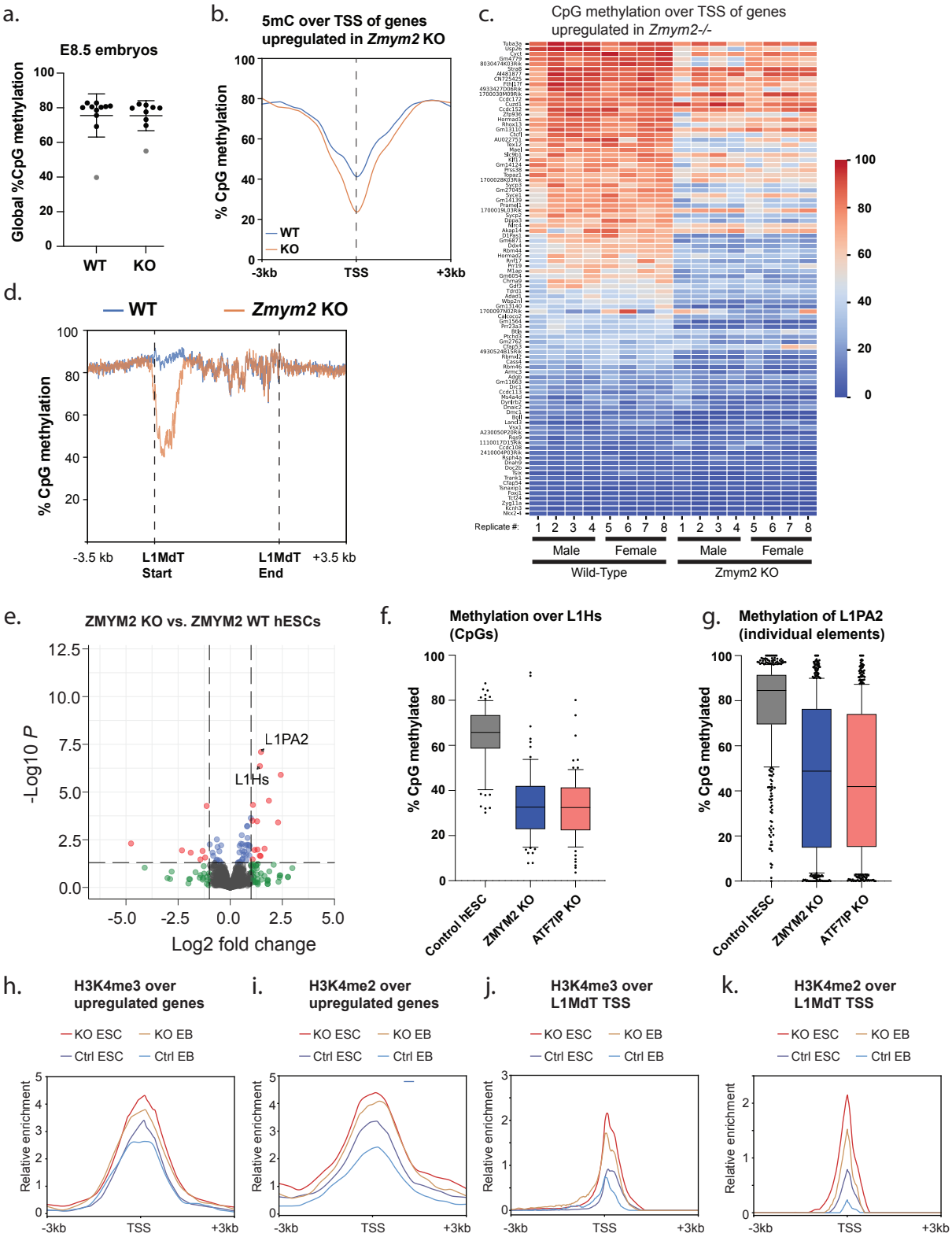

a.
E8.5 embryos
Global %CpG methylation
WT
KO
b.
5mC over TSS of genes upregulated in Zmym2 KO
% CpG methylation
WT
KO
-3kb
TSS
+3kb
c.
CpG methylation over TSS of genes upregulated in Zmym2-/-
Replicate #:
Male
Female
Male
Female
Wild-Type
Zmym2 KO
d.
WT
Zmym2 KO
% CpG methylation
-3.5 kb
L1MdT Start
L1MdT End
+3.5 kb
e.
ZMYM2 KO vs. ZMYM2 WT hESCs
-Log10 P
L1PA2
L1Hs
Log2 fold change
f.
Methylation over L1Hs (CpGs)
% CpG methylated
Control hESC
ZMYM2 KO
ATF7IP KO
g.
Methylation of L1PA2 (individual elements)
% CpG methylated
Control hESC
ZMYM2 KO
ATF7IP KO
h.
H3K4me3 over upregulated genes
i.
H3K4me2 over upregulated genes
j.
H3K4me3 over L1MdT TSS
k.
H3K4me2 over L1MdT TSS
KO ESC
KO EB
Ctrl ESC
Ctrl EB
Relative enrichment
-3kb
TSS
+3kb

Figure 3. Impaired DNA methylation in *Zmym2*$^{-/-}$ **a,** Dot 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 genes upregulated in *Zmym2*$^{-/-}$ embryos. Transposon fusion genes are excluded. **c,** Heatmap of DNA methylation of the TSS of 94 upregulated genes in *Zmym2*$^{-/-}$ expressed from annotated TSS. Average methylation within 500bp of transcription start site of the 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 (>6kb) L1MdT elements in the genome. **e.** Volcano plot of differentially expressed transposons comparing *ZMYM2*$^{+/+}$ and *ZMYM2*$^{-/-}$ hESCs. Significant differentially expressed transposons (fold ≥ 2, q-value < 0.05) are coloured red. **f,** A boxplot of average CpG methylation across 1Kb start regions of full length L1Hs transposable elements. **g,** A boxplot of average CpG methylation across L1PA2 transposable elements. **h – k,** Metaplot of H3K4me3 (**h, j**) and H3K4me2 (**i, k**) over the TSS of 94 upregulated genes in *Zmym2*$^{-/-}$ expressed from annotated TSS (**h,i**) or L1MdT elements (**j,k**) in *Zmym2*$^{-/-}$ (KO) and control (Ctrl) mESCs.. To normalize for ChIP quality, enrichment is normalized based on peak height for all annotated genes.

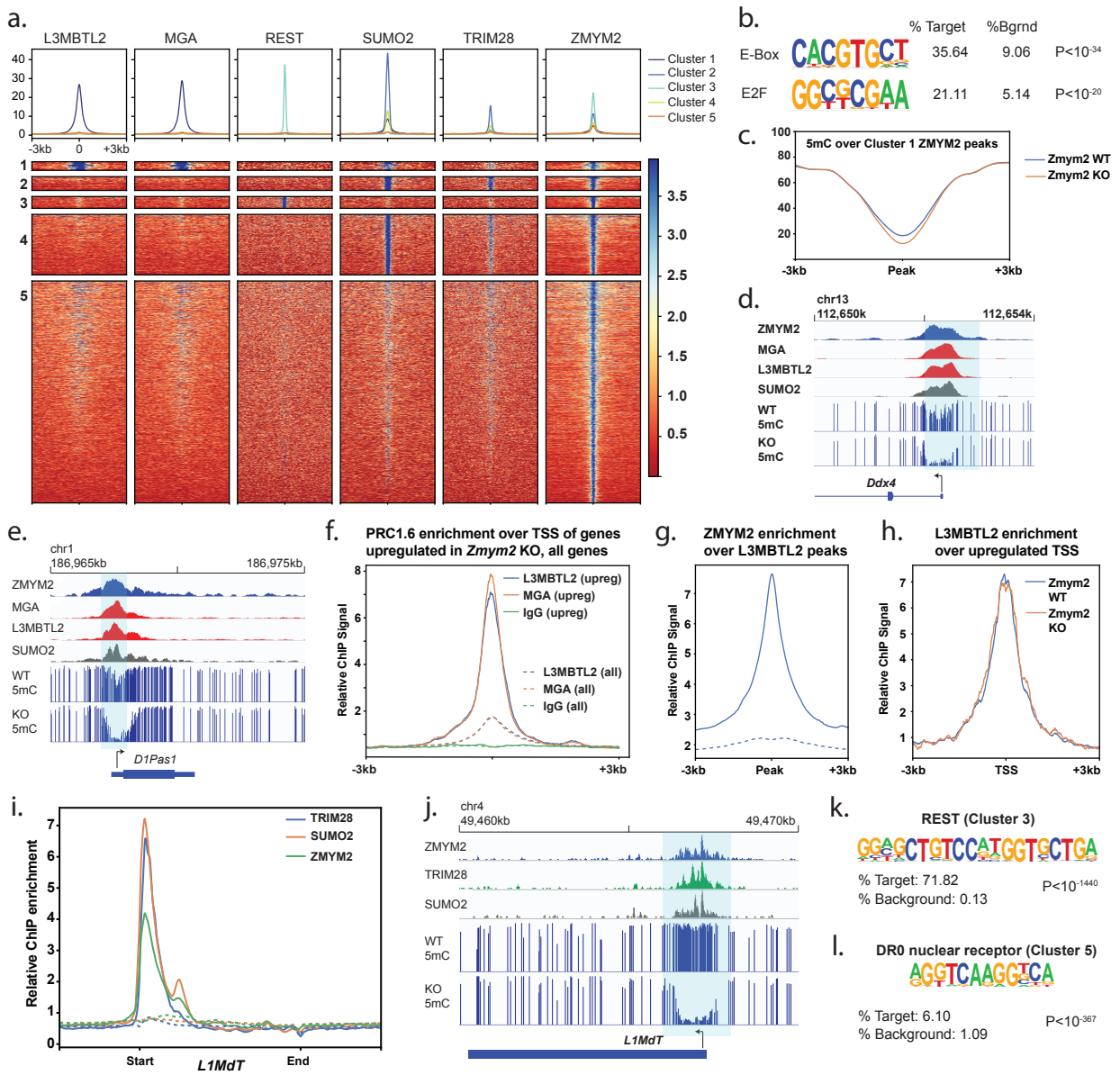


Figure 4. ZMYM2 is recruited to chromatin by a variety of distinct complexes. a, Heatmaps of ZMYM2, L3MBTL2, MGA, REST, SUMO2, and TRIM28 enrichment over ZMYM2 ChIP-seq peak summits and the flanking 3Kb regions (n = 21,549 peaks). Peaks were partitioned into five clusters using a k-means algorithm. Peak numbers: Cluster 1, 290 peaks; Cluste 2, 931 peaks; Cluster 3, 796 peaks; Cluster 4, 4080 peaks; Cluster 5, 14969 peaks. Cluster 1 peaks are shown at twice the width of other peaks to enhance visibility. **b,** HOMER motif analysis of Cluster 1 ZMYM2 peaks show enrichment of E-Box and E2F motifs. **c,** Metaplot of CpG methylation spanning Cluster 1 ZMYM2 peaks and the flanking 3Kb region. **d,e,** ChIP-seq and DNA methylation data over the Ddx4 (**d**) and D1Pas (**e**) loci. Note colocalization of ZMYM2 with PRC1.6 components and hypomethylation in *Zmym2*$^{-/-}$. **f,** A 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). **g,** Metaplot of ZMYM2 ChIP-seq signal over L3MBTL2 peaks. **h,** Metaplot of L3MBTL2 CUT&RUN ChIP-seq signal over the annotated TSS of upregulated genes in both *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ mESCs. **i,** Metaplot of ZMYM2, TRIM28, SUMO2 ChIP-seq over all full length L1MdT elements. Dashed line indicates ChIP input. **j,** ChIP-seq and DNA methylation data over a representative LINE element. Note overlap of TRIM28 and ZMYM2 peaks, and hypomethylation over the peak site. **k,** HOMER motif analysis of Cluster 3 ZMYM2 peaks show enrichment of REST-binding motif. **l,** HOMER motif analysis of Cluster 5 ZMYM peaks show enrichment of DR0 nucleic acid motif.

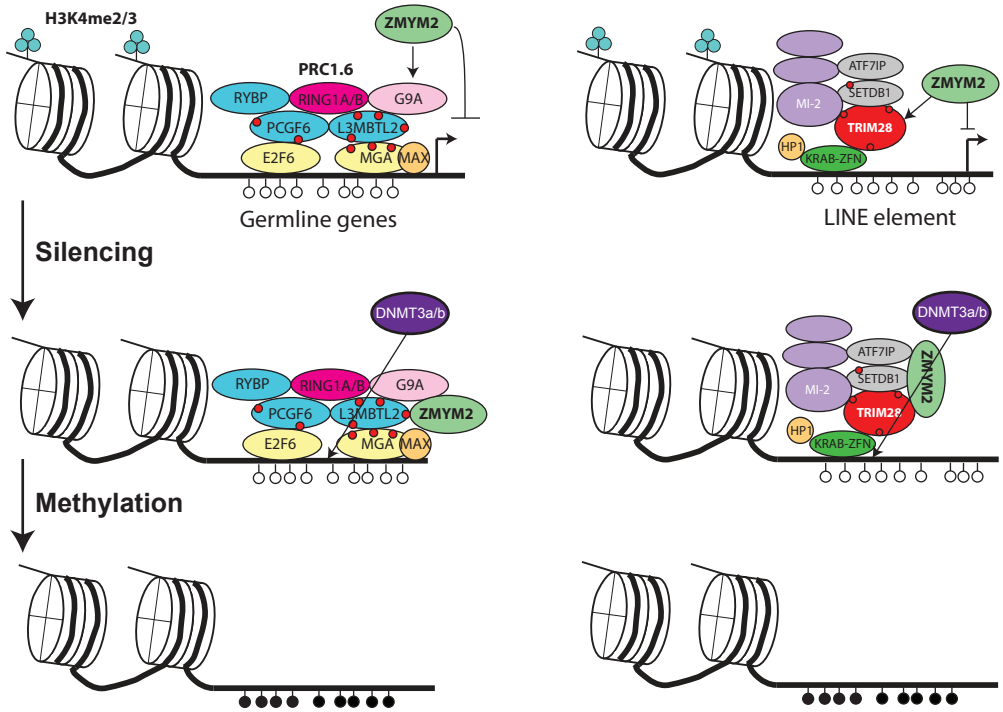
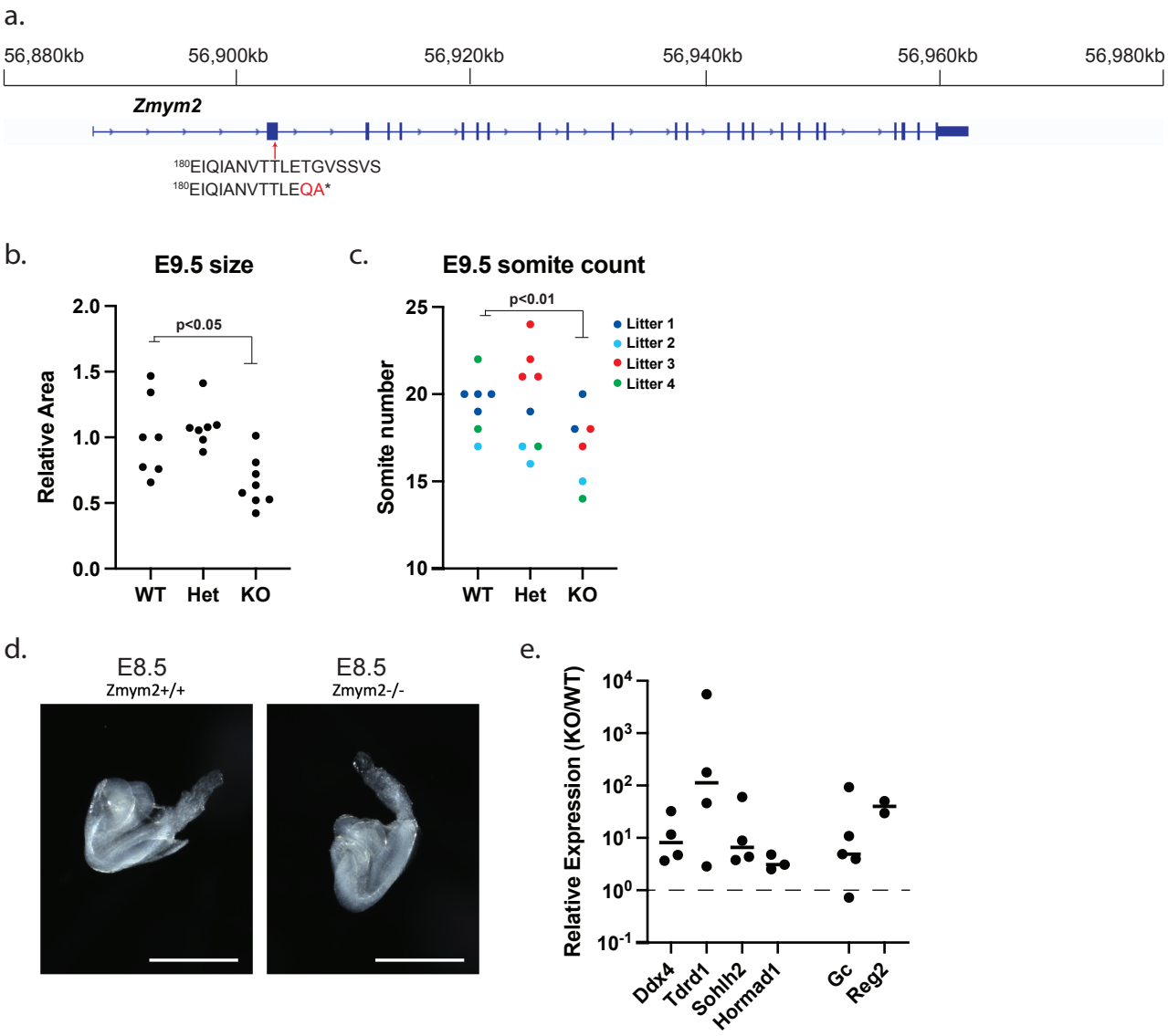
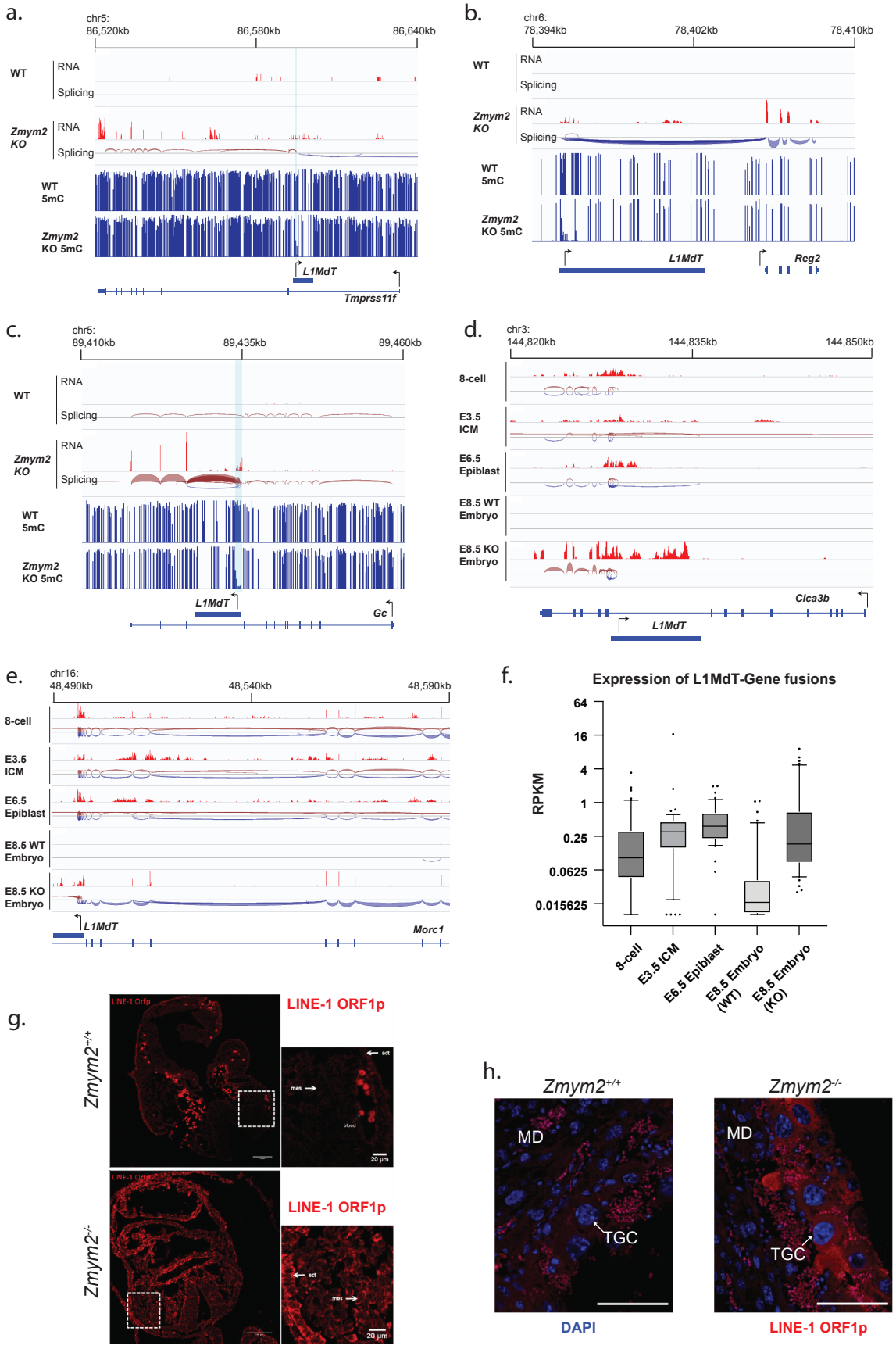


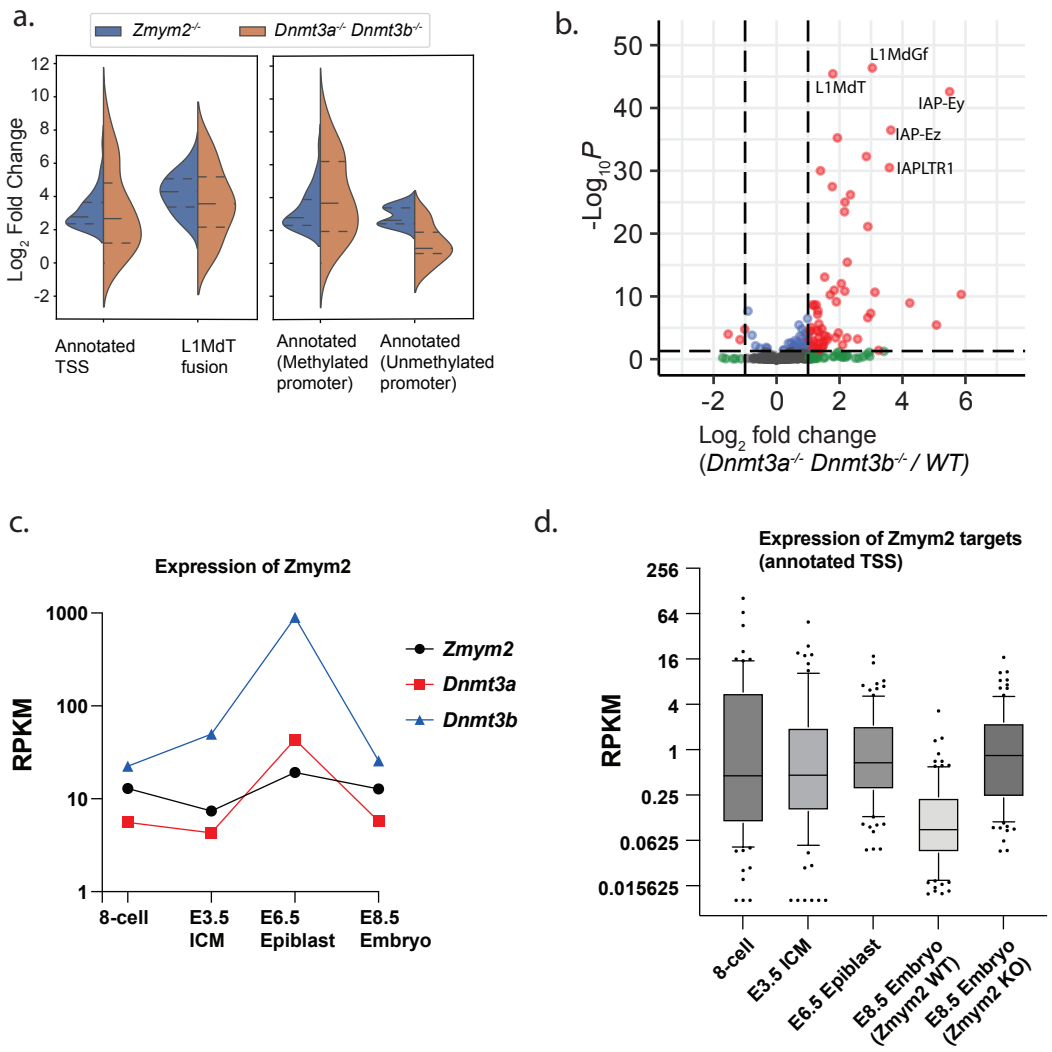
Figure 5. Proposed mechanism for ZMYM2-mediated DNA methylation. ZMYM2 is recruited to PRC1.6 and TRIM28 binding sites respectively, potentially via binding to SUMOylation sites (indicated as red circles) or interaction with complex components such as ATF7IP. ZMYM2 mediates silencing of target loci, resulting in loss of H3K4 methylation and acquisition of DNA methylation.



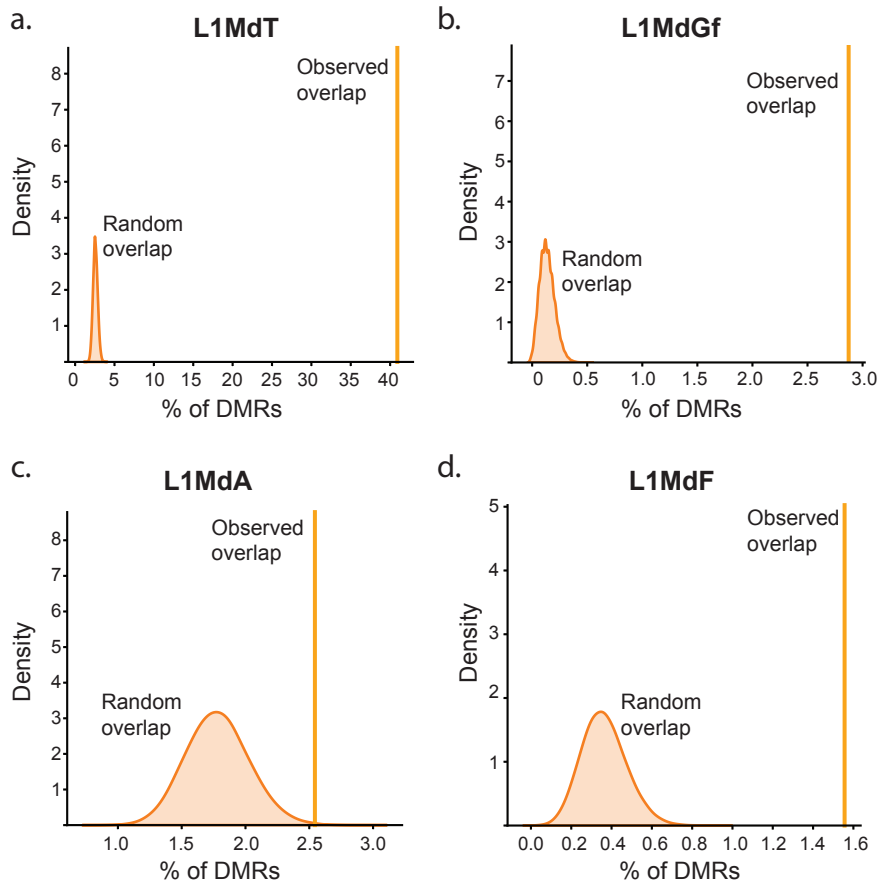
Extended Data Figure 1. Reduced size and developmental delays in *Zmym2*$^{-/-}$ mice and confirmation of gene dysregulation. a, Schematic representation of point mutation and truncation point in *Zmym2*- allele. **b,** Relative size of embryos measured as 2D area in embryos in sagittal position normalized by litter as a ratio to average *Zmym2*$^{+/+}$ area. p-value calculated on basis of one-tailed t-test. **c,** Somite counts of E9.5 embryos per genotype. Each litter is represented in a different color to account for litter variability. Note that *Zmym2*$^{-/-}$ mice show reduced somite count within each litter. p-value calculated on basis of one-tailed t-test. **d,** Images of E8.5 *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ mice. **e,** Quantitative reverse-transcription PCR (qRT-PCR) of four upregulated genes expressed from promoter (*Ddx4*, *Tdrd1*, *Sohlh2*, *Hormad1*) and LINE-fusions (*Gc*, *Reg2*). Each dot reflects the ratio of expression in KO/WT littermate embryos, with expression normalized to *RPS16*.


a.
chr5:
86,520kb
86,580kb
86,640kb
WT
RNA
Splicing
Zmym2 KO
RNA
Splicing
WT 5mC
Zmym2 KO 5mC
L1MdT
Tmprss11f
b.
chr6:
78,394kb
78,402kb
78,410kb
WT
RNA
Splicing
Zmym2 KO
RNA
Splicing
WT 5mC
Zmym2 KO 5mC
L1MdT
Reg2
c.
chr5:
89,410kb
89,435kb
89,460kb
WT
RNA
Splicing
Zmym2 KO
RNA
Splicing
WT 5mC
Zmym2 KO 5mC
L1MdT
Gc
d.
chr3:
144,820kb
144,835kb
144,850kb
8-cell
E3.5 ICM
E6.5 Epiblast
E8.5 WT Embryo
E8.5 KO Embryo
Clca3b
L1MdT
e.
chr16:
48,490kb
48,540kb
48,590kb
8-cell
E3.5 ICM
E6.5 Epiblast
E8.5 WT Embryo
E8.5 KO Embryo
L1MdT
Morc1
f.
Expression of L1MdT-Gene fusions
RPKM
64
16
4
1
0.25
0.0625
0.015625
8-cell
E3.5 ICM
E6.5 Epiblast
E8.5 Embryo (WT)
E8.5 Embryo (KO)
g.
Zmym2+/+
LINE-1 ORF1p
Zmym2-/-
LINE-1 ORF1p
20 µm
h.
Zmym2+/+
Zmym2-/-
MD
TGC
DAPI
LINE-1 ORF1p

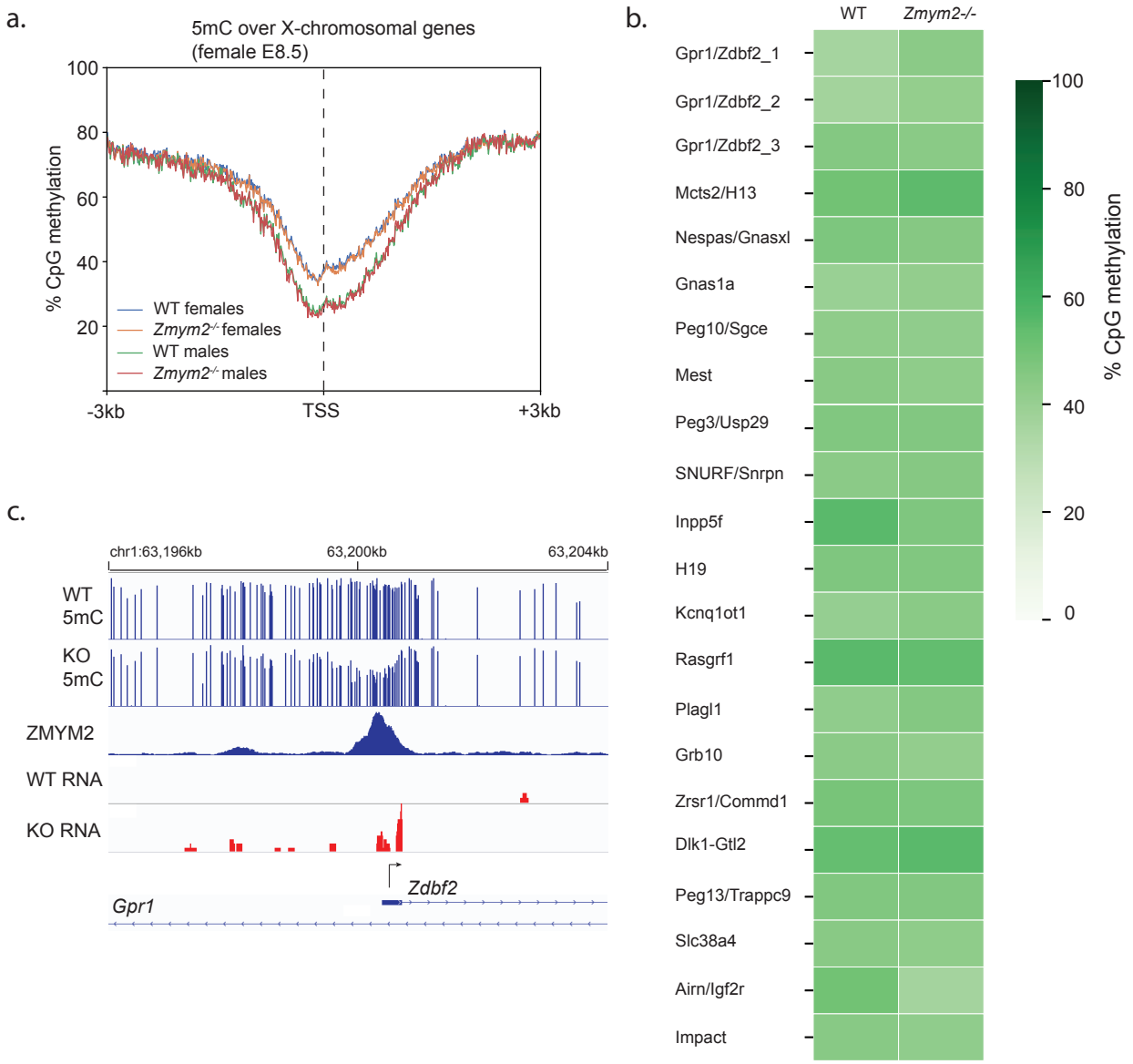
Extended Data Figure 2. Expression of L1MdT-gene fusion transcripts in *Zmym2*$^{-/-}$ embryos and early embryogenesis. a-c, Representative examples of L1MdT-gene fusion transcripts. Note examples of bidirectional transcription of an internal L1MdT element (**a**), splicing from an upstream L1MdT in a forward direction into a gene (**b**) and splicing from an internal L1MdT promoter in a forward direction into a gene (**c**). Hypomethylated DMRs are indicated in light blue. **d, e,** Representative examples of L1MdT-gene fusion transcript expression in early embryogenesis. **f,** Box plot of L1MdT-gene fusion transcripts during early mouse embryonic development. In **d-f,** 8-cell, E3.5 and E6.5 RNA-seq data is from Smith 2017. **g,** Immunofluorescent staining of LINE-1 ORFp (red) in E8.5 *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ embryos. Regions of mesoderm and ectoderm in inset are indicated. Scale bar= 20µm
h, Immunofluorescent staining of LINE-1 ORFp (red) and DAPI (blue) of E9.5 *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ implantation sites. TGC= trophoblast giant cell, MD= maternal decidua. Small red cells visible in both samples are maternal blood cells. Scale bars= 100µm.



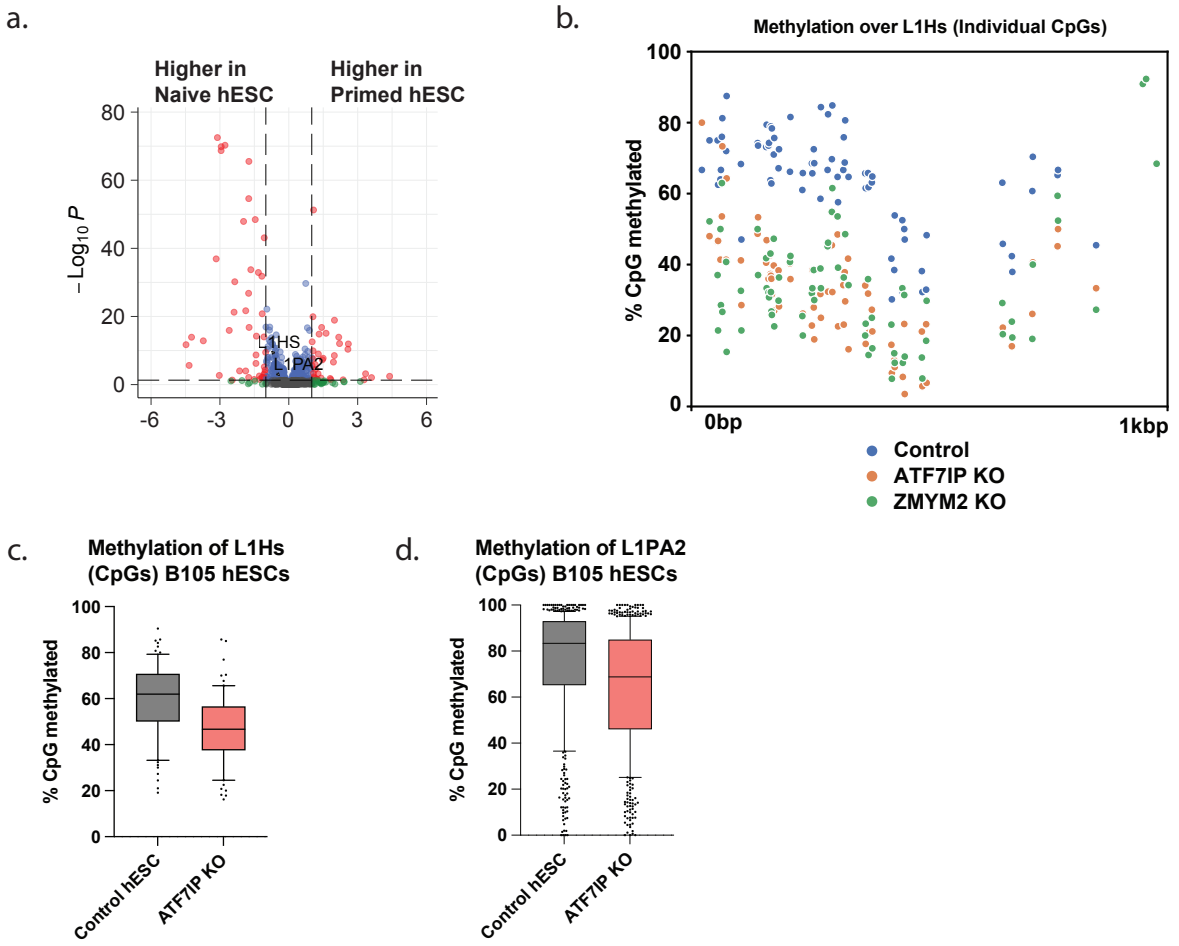
Extended Data Figure 3. Hypomethylation of germline genes in *Zmym2*$^{-/-}$. a, Expression of each gene in the given category is shown (KO/WT) in *Zmym2*$^{-/-}$ and *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$. Note that most genes upregulated in Zmym2-/- are also upregulated in the Dnmt double knockout. **b,** Volcano plot of differentially expressed transposable elements comparing *Dnmt3a*$^{-/-}$ *Dnmt3b*$^{-/-}$ and control E8.5 embryos. Significant differentially expressed transposable elements (fold ≥ 2, q-value < 0.05) are coloured red. **c,** Expression of ZMYM2 and de novo methyltransferases DNMT3A and DNMT3B during early embryogenesis. **d,** Expression of the 95 upregulated genes expressed from TSS during early embryogenesis. In **c** and **d**, 8-cell, E3.5 and E6.5 RNA-seq data is from Smith 2017.



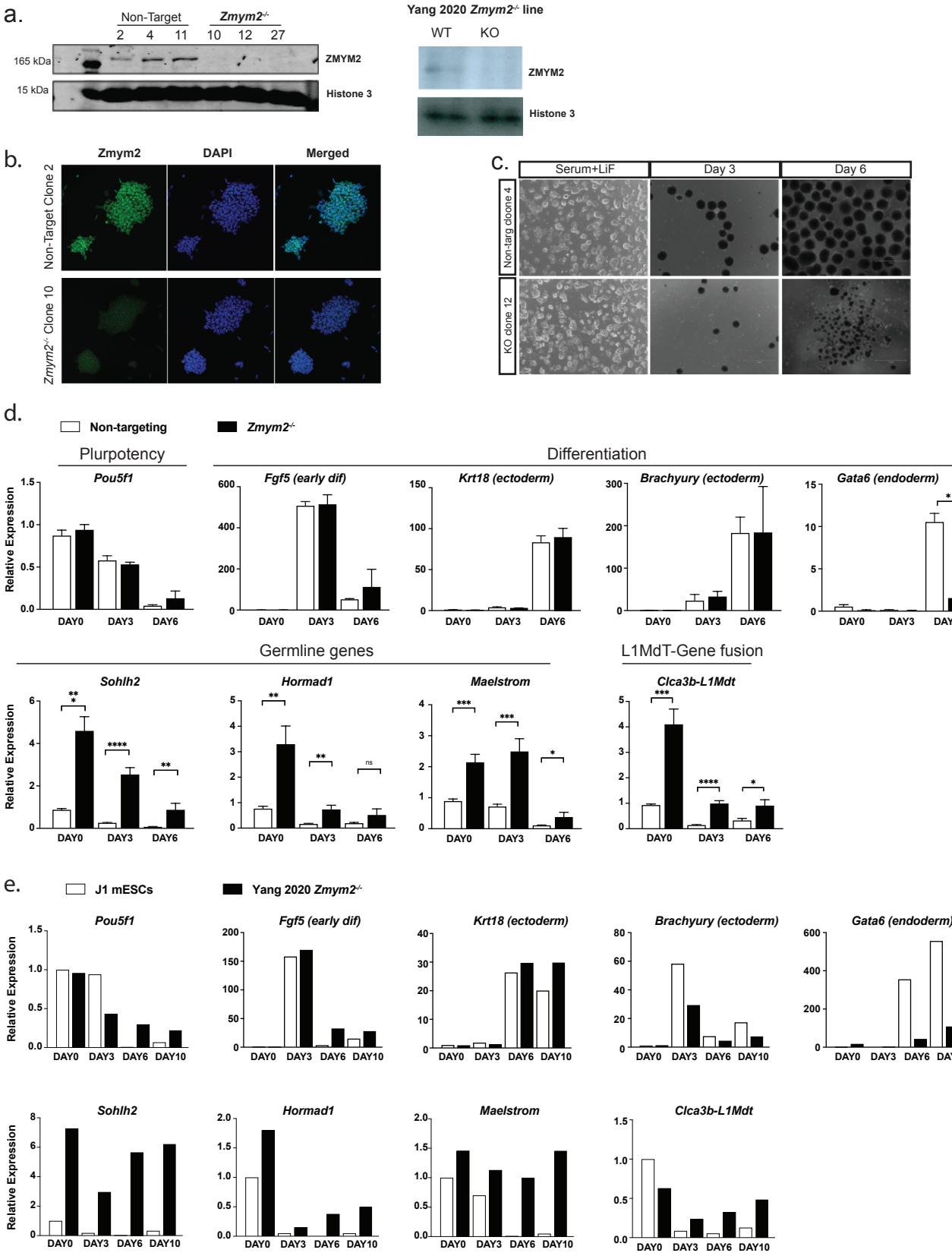
Extended Data Figure 4. Overlap of DMRs with L1Md elements. A Monte-Carlo simulation was undertaken in which the positions of DMRs were randomly placed in the genome and repeated 100,000 times. The overlap with indicated transposon class is indicated as “Random overlap”. This number is contrasted with the “observed overlap” number for actual DMRs.



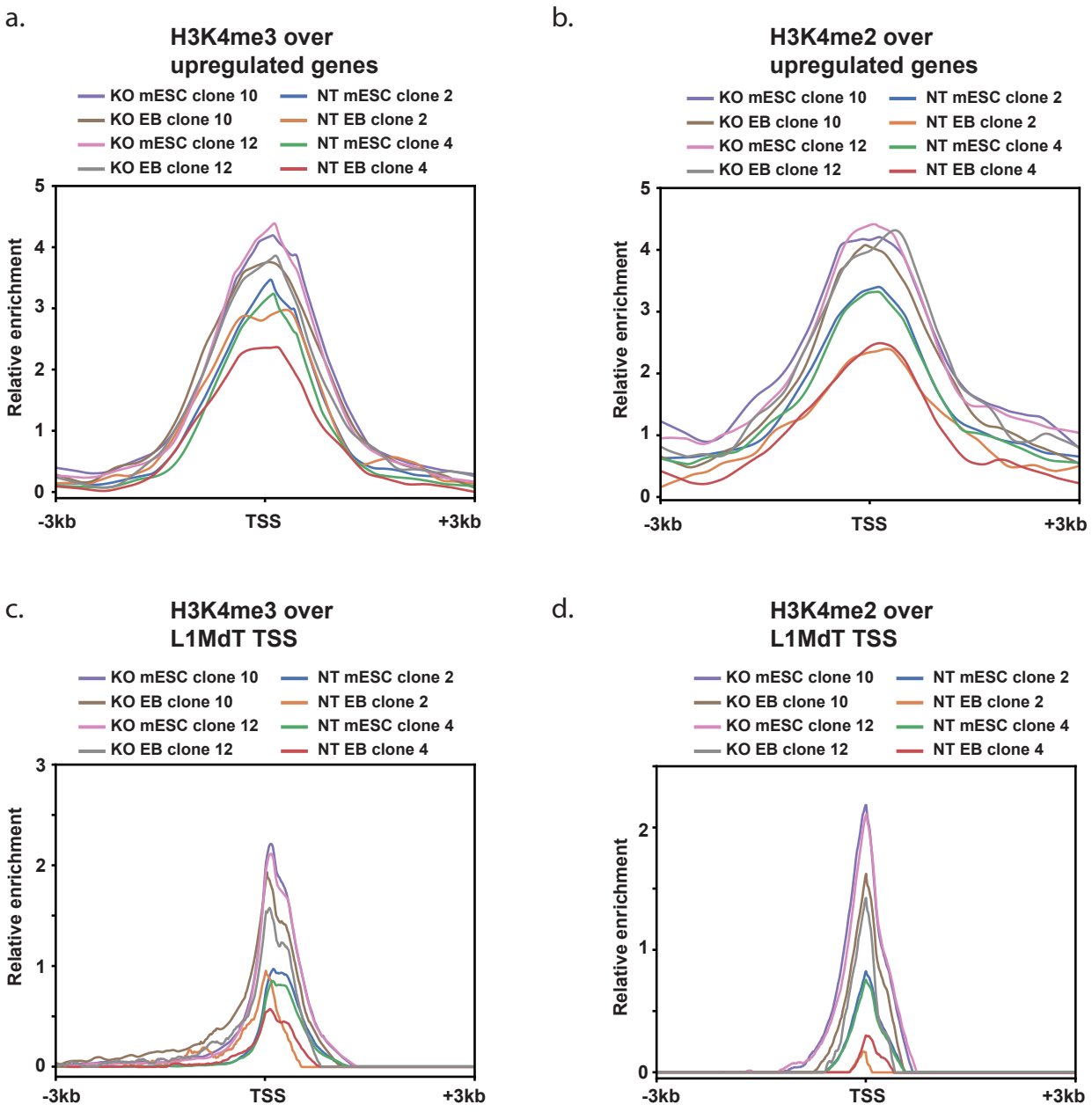
Extended Data Figure 5. Normal X-inactivation, imprint methylation in *Zmym2*$^{-/-}$. a, Metaplot of DNA methylation over X-chromosomal genes in *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$. **b,** DNA methylation over known imprinted regions in mice in *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$. **c,** Hypomethylation, and aberrant transcription from the *Liz/Zdbf2* transient imprint in *Zmym2*$^{-/-}$. Note enrichment of ZMYM2 at the target locus.



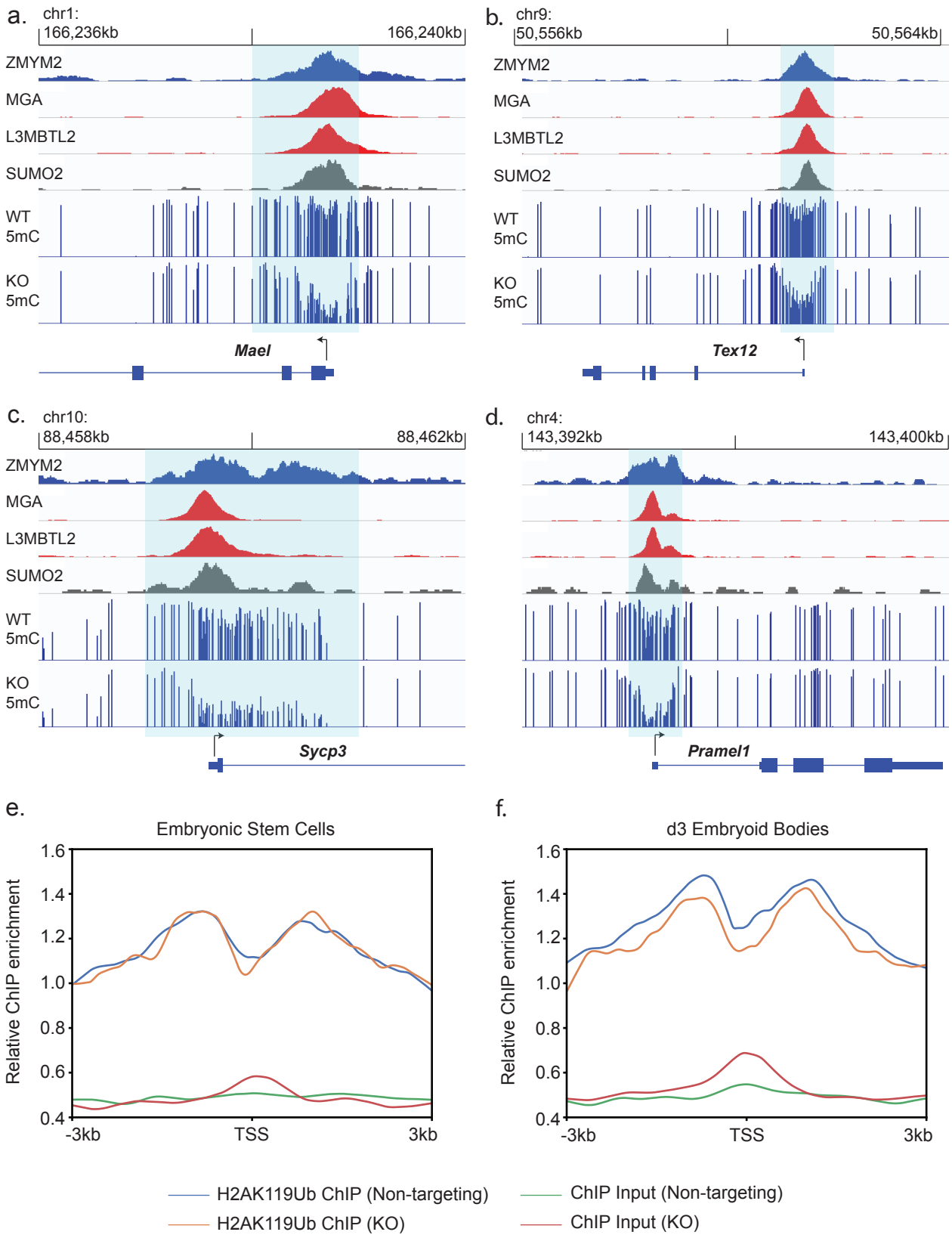
Extended Data Figure 6. Upregulation and hypomethylation of transposons in hESCs a, Relative expression of transposon families in naïve and primed hESCs. Note that naïve hESCs show only a very modest increase in L1Hs and L1PA2 levels. **b,** Methylation level of individual CpGs in an L1HS consensus element for the genotype indicated. Only CpGs with >10-fold coverage in the Bar 2021 RRBS data are shown. **c,d,** DNA methylation of L1Hs (**c**) and L1PA2 (**d**) control and *ATF7IP*$^{-/-}$ hESCs in B105 hESCs. Note similarity to Figure 4b,c in this second hESC line.


a.
Non-Target
Zmym2-/-
2 4 11 10 12 27
165 kDa
15 kDa
ZMYM2
Histone 3
Yang 2020 Zmym2-/- line
WT KO
ZMYM2
Histone 3
b.
Zmym2
DAPI
Merged
Non-Target Clone 2
Zmym2-/- Clone 10
c.
Serum+LiF
Day 3
Day 6
Non-targ clone 4
KO clone 12
d.
Non-targeting
Zmym2-/-
Plurpotency
Differentiation
Pou5f1
Fgf5 (early dif)
Krt18 (ectoderm)
Brachyury (ectoderm)
Gata6 (endoderm)
Relative Expression
DAY0 DAY3 DAY6
Germline genes
L1MdT-Gene fusion
Sohlh2
Hormad1
Maelstrom
Clca3b-L1Mdt
e.
J1 mESCs
Yang 2020 Zmym2-/-
DAY0 DAY3 DAY6 DAY10

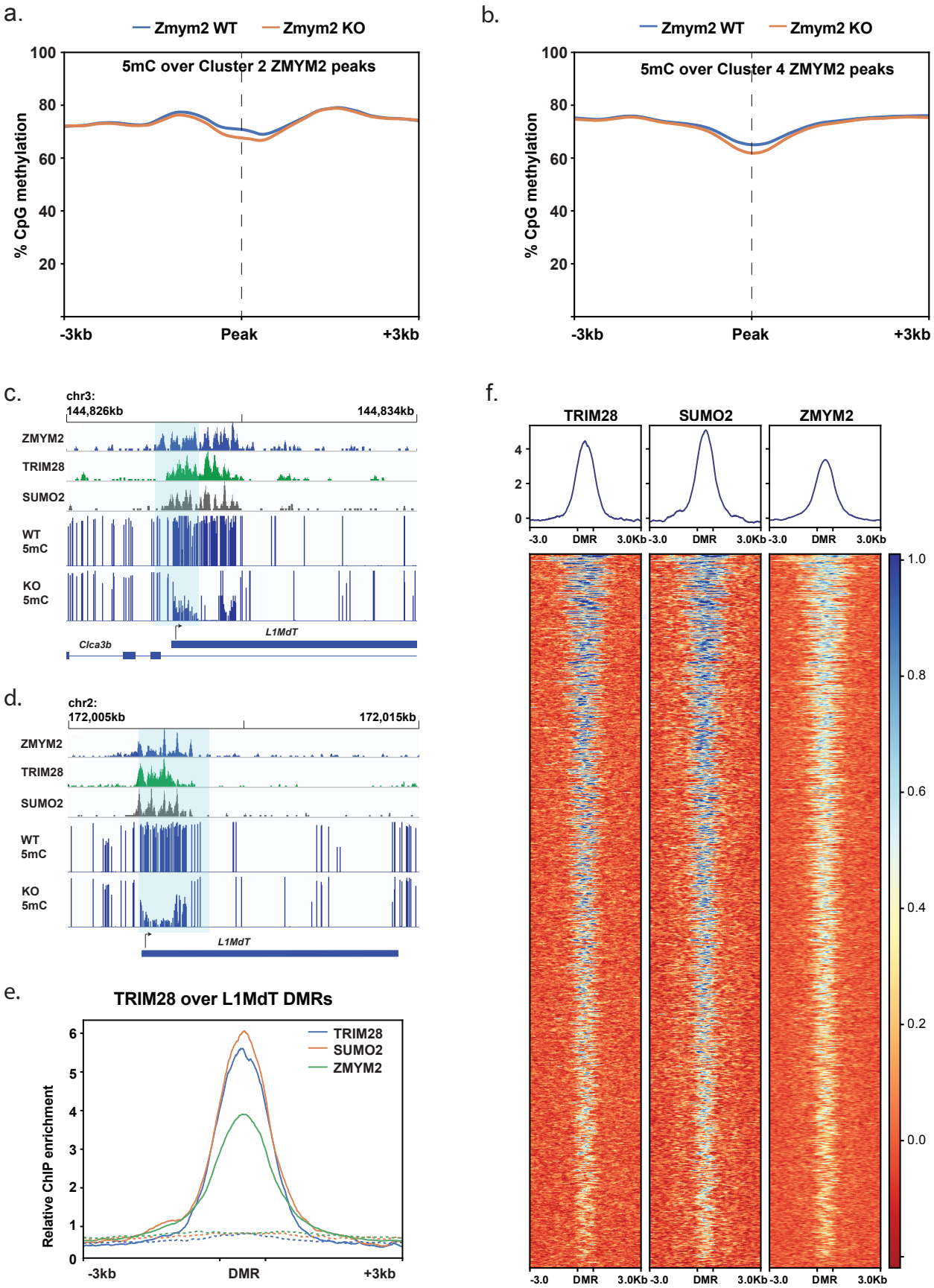
Extended Data Figure 7. Recapitulation of ZMYM2-mediated silencing in an embryoid body (EB) assay a, Western blots of control and *Zmym2*$^{-/-}$ clonal mESC lines generated on a V6.5 background (left) as well as WT J1 and *Zmym2*$^{-/-}$ mESCs generated by Yang and colleagues. **b,** Immunofluorescent staining of control and *Zmym2*$^{-/-}$ clonal line for ZMYM2. **c,** Light microscopy of *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ ESCs and EBs. **d,** Relative expression of the pluripotency factor *Pou5f1*, differentiation genes, and ZMYM2-target genes in three *Zmym2*$^{-/-}$ as compared with three control lines, during two separate EB differentiation experiments (n=6 replicates total). **e,** Relative expression of pluripotency, differentiation, and ZMYM2-target genes in control J1 mESCs and *Zmym2*$^{-/-}$ mESCs generated by Yang and colleagues, during embryoid body differentiation. In **d** and **e**, * indicates $p < 0.05$, ** $p < 0.01$, *** $p< 0.001$, **** $p< 0.0001$, by a two-tailed t-test.



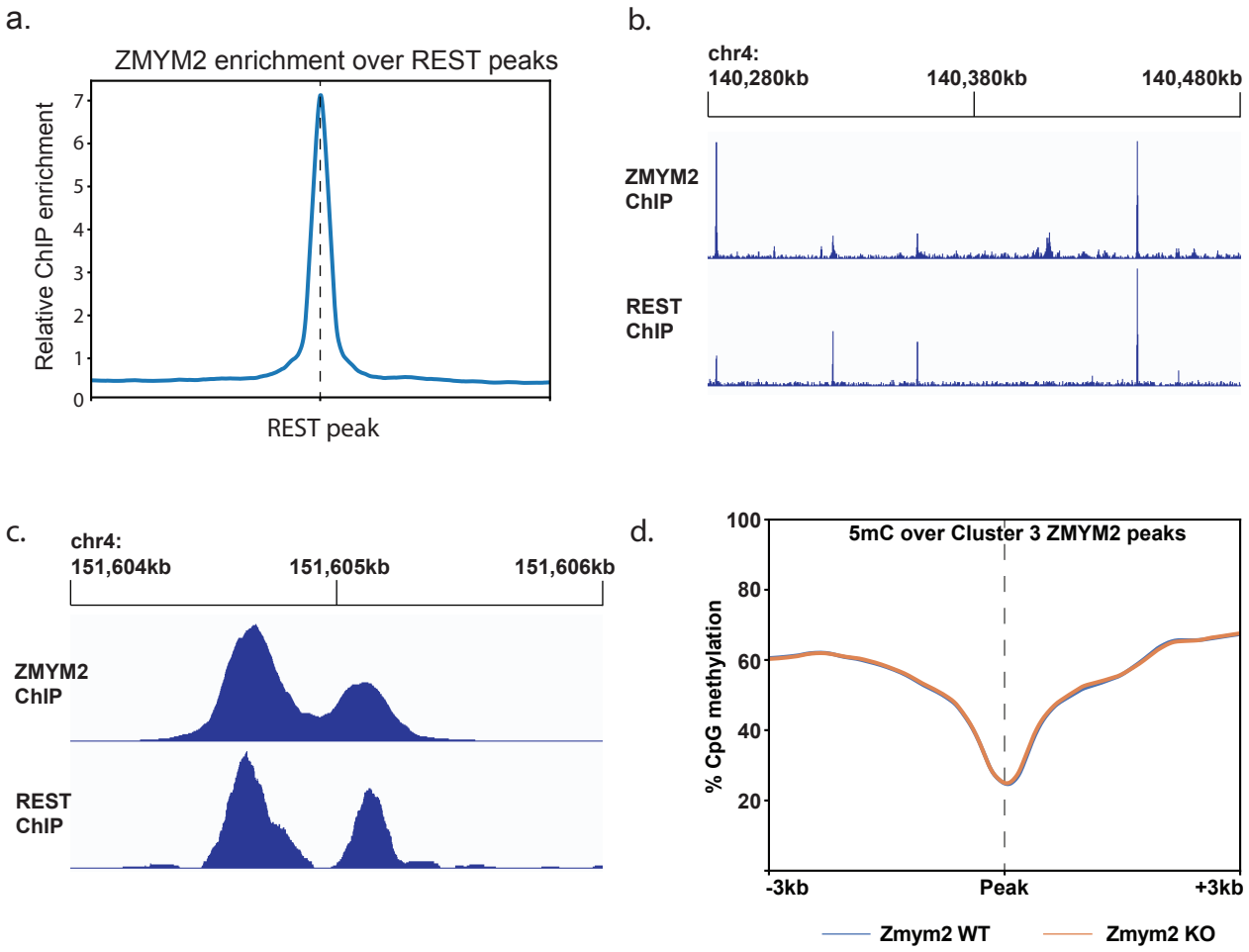
Extended Data Figure 8. Reduced enrichment of H3K4me2 and H3K4me3 over ZMYM2 target loci. Metaplot of H3K4me3 **(a, c)** and H3K4me2 **(b, d)** over the TSS of 94 upregulated genes in *Zmym2*$^{-/-}$ expressed from annotated TSS **(a,b)** or L1MdT elements **(c,d).** This figure shows individual replicates that were merged in Figure 3h-k. NT=non-targeting sgRNA, indicated as Ctrl in Figure 3. To normalize for ChIP quality, enrichment is normalized based on peak height for all annotated genes.


a.
chr1:
166,236kb
166,240kb
ZMYM2
MGA
L3MBTL2
SUMO2
WT
5mC
KO
5mC
Mael
b.
chr9:
50,556kb
50,564kb
ZMYM2
MGA
L3MBTL2
SUMO2
WT
5mC
KO
5mC
Tex12
c.
chr10:
88,458kb
88,462kb
ZMYM2
MGA
L3MBTL2
SUMO2
WT
5mC
KO
5mC
Sycp3
d.
chr4:
143,392kb
143,400kb
Pramel1
e.
Embryonic Stem Cells
Relative ChIP enrichment
1.6
1.4
1.2
1.0
0.8
0.6
0.4
-3kb
TSS
3kb
f.
d3 Embryoid Bodies
Relative ChIP enrichment
1.6
1.4
1.2
1.0
0.8
0.6
0.4
-3kb
TSS
3kb
H2AK119Ub ChIP (Non-targeting)
H2AK119Ub ChIP (KO)
ChIP Input (Non-targeting)
ChIP Input (KO)

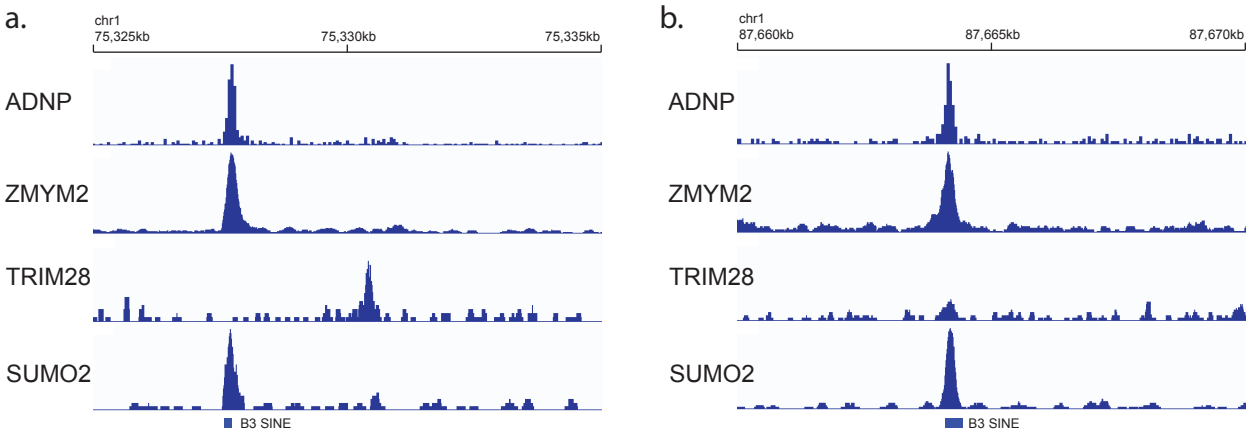
Extended Data Figure 9. Hypomethylation of PRC1.6 targets in *Zmym2*$^{-/-}$. a-d, Four examples of Cluster 1, PRC1.6-bound genes upregulated in *Zmym2*$^{-/-}$. Note colocalization of ZMYM2 and PRC1.6 targets, as well as hypomethylation in *Zmym2*$^{-/-}$. Hypomethylated DMRs are indicated with light blue boxes. **e, f,** Relative ChIP enrichment of H2AK119Ub in *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$ mESCs (**e**) and day 3 EBs (**f**) over genes upregulated in *Zmym2*$^{-/-}$ (transposon-fusion genes excluded). Note that there is no difference between *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$. The modestly increased input enrichment in *Zmym2*$^{-/-}$ may reflect increased locus accessibility as a result of impaired silencing.


a.
Zmym2 WT
Zmym2 KO
5mC over Cluster 2 ZMYM2 peaks
% CpG methylation
100
80
60
40
20
-3kb
Peak
+3kb
b.
Zmym2 WT
Zmym2 KO
5mC over Cluster 4 ZMYM2 peaks
% CpG methylation
100
80
60
40
20
-3kb
Peak
+3kb
c.
chr3:
144,826kb
144,834kb
ZMYM2
TRIM28
SUMO2
WT 5mC
KO 5mC
L1MdT
Clca3b
d.
chr2:
172,005kb
172,015kb
ZMYM2
TRIM28
SUMO2
WT 5mC
KO 5mC
L1MdT
e.
TRIM28 over L1MdT DMRs
TRIM28
SUMO2
ZMYM2
Relative ChIP enrichment
6
5
4
3
2
1
0
-3kb
DMR
+3kb
f.
TRIM28
SUMO2
ZMYM2
4
2
0
-3.0
DMR
3.0Kb
1.0
0.8
0.6
0.4
0.2
0.0

Extended Data Figure 10. Hypomethylation of transposons in *Zmym2*$^{-/-}$. a, b, Metaplot of DNA methylation over Cluster 2 and Cluster 4 peaks for *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$. **c,d,** ZMYM2, TRIM28, SUMO2 ChIP-seq and WGBS data plotted over representative L1MdT elements. Note correspondence of the three proteins at L1MdT promoters and corresponding reduction in DNA methylation. **e,** Metaplot of TRIM28, SUMO2 and ZMYM2 mESC ChIP-seq signal over DMRs hypomethylated in *Zmym2*$^{-/-}$ which overlap with L1MdT elements. Dashed lines indicate ChIP input. **f.** Heatmap of input-normalized ZMYM2, TRIM28, SUMO2 ChIP-seq data over hypomethylated L1MdT DMRs.



Extended Data Figure 11. Overlap of ZMYM2 and REST binding sites. a, ZMYM2 plotted over REST binding sites. **b, c,** Examples of overlap of REST and ZMYM2 binding sites. **d,** DNA methylation over Cluster 3 peaks in *Zmym2*$^{+/+}$ and *Zmym2*$^{-/-}$.



Extended Data Figure 12. Overlap of ZMYM2 and ChAHP complex binding sites. a,b, Examples of overlap of ADNP and ChAHP complex binding sites. Note co-enrichment of ADNP, ZMYM2 and SUMO over indicated SINE elements, with no corresponding TRIM28 enrichment.