



Published in final edited form as:

Nature. 2020 July ; 583(7817): 625–630. doi:10.1038/s41586-020-2500-9.

N(6)-methyladenine in DNA antagonizes SATB1 in early development

Zheng Li^{1, #}, Shuai Zhao^{2, #}, Raman V. Nelakanti^{1, #}, Kaixuan Lin^{1, #}, Tao P. Wu^{1, 7, #}, Myles H. Alderman III¹, Cheng Guo^{3, 4}, Pengcheng Wang³, Min Zhang², Wang Min⁵, Zongliang Jiang⁶, Yinsheng Wang³, Haitao Li^{2, *}, Andrew Z. Xiao^{1, *}

¹Department of Genetics and Yale Stem Cell Center, Yale School of Medicine, New Haven, Connecticut 06520, USA.

²MOE Key Laboratory of Protein Sciences, Beijing Advanced Innovation Center for Structural Biology, Beijing Frontier Research Center for Biological Structure, Tsinghua-Peking Joint Center for Life Sciences, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, China

³Department of Chemistry, University of California, Riverside, CA 92521, USA

⁴Cancer Institute, Key Laboratory of Cancer Prevention and Intervention, China National Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, 310009, China.

⁵Interdepartmental Program in Vascular Biology and Therapeutics, Department of Pathology, Yale University School of Medicine

⁶School of Animal Sciences, AgCenter, Louisiana State University, Baton Rouge, LA 70803, USA

⁷Current address: Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA

Abstract

The recent discovery of N^6 -mA in mammalian genomes suggests that it may serve as an epigenetic regulatory mechanism¹. However, the biological role of N^6 -mA and molecular pathways exerting

Reprints and permissions information is available at www.nature.com/reprints. Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

* **Corresponding authors** Correspondence and requests for materials should be addressed to Andrew Xiao (andrew.xiao@yale.edu) or Haitao Li (lht@tsinghua.edu.cn).

Author Contributions

Z.L., R.V.N and T.P.W. designed the cell biology, genetics and genomic experiments and analyzed the data. T.P.W. established the *iCdx2* approach and initiated the study. S.Z., M.Z. and H.L. designed and performed SATB1 binding experiments. R.V.N. developed the ssDNA-seq approach and analyzed the data. K.L. and T.P.W. analyzed the sequencing data. M.H.A. and R.V.N. generated the *Alkbh1* KO mice and characterized the phenotype. C.G., P.W.W. and Y.W. performed the MS experiments and analyzed the data. W.M. and Z.J. helped to characterize the early embryonic defects. A.Z.X. conceived the hypothesis and designed the study. R.V.N. and A.Z.X. wrote the manuscripts.

#: These authors equally contribute to this work.

Competing interests

The authors declare no competing interests.

Additional Information

Supplementary Information is available for this paper.

its function remain elusive. Herein, we demonstrate that N^6 -mA plays a critical role in changing the epigenetic landscape during cell fate transitions in early development. We found that N^6 -mA is upregulated during trophoblast stem cell development, specifically at Stress Induced DNA Double Helix Destabilization (SIDD) regions²⁻⁴. It is well-known that SIDD regions are conducive to topological stress-induced double helix unpairing and play critical roles in organizing large-scale chromatin structures^{3,5,6}. We demonstrated that the presence of N^6 -mA abolishes (>500-fold) the *in vitro* interactions between SIDD and SATB1, a critical chromatin organizer interacting with SIDD regions; N^6 -mA deposition also effectively antagonizes SATB1 function *in vivo* by preventing its binding to chromatin. Concordantly, N^6 -mA functions at the boundaries between eu-/hetero- chromatin to restrict the spreading of euchromatin. N^6 -mA mediated repression is critical for gene regulation during trophoblast development in cell culture models and *in vivo*. Overall, our study discovers an unexpected molecular mechanism for N^6 -mA function via SATB1, and reveals surprising connections between DNA modification, DNA secondary structures and large chromatin domains in early embryonic development.

Several recent studies have implicated an intriguing role of N^6 -mA in epigenetic silencing, especially at long interspersed element 1 (LINE-1) endogenous transposons in mouse embryonic stem cells (mESCs), brain tissues under environmental stress, and in human lymphoblastoid cells (LCLs) and tumorigenesis^{1,7-9}. Although these findings underscore the importance of N^6 -mA in mammalian biology and human disease, the underlying mechanisms of N^6 -mA mediated silencing have thus far remained elusive.

N^6 -mA is enriched at AT-rich regions in mammalian genomes¹, especially for those interacting with the nuclear architecture, such as Matrix/Scaffold attachment regions (M/SAR). M/SARs are liable to undergo DNA unpairing under torsional stress (known as the Base Unpairing Regions, BURs or SIDD)²⁻⁴ in replication, transcription, and recombination^{3,5,6}. A recent study by genome-wide permanganate and S1 endonuclease mapping (ssDNA-seq¹⁰) demonstrated a strong correlation between the SIDD regions identified experimentally and those predicated by computational approaches¹¹. SIDD/BUR regions in S/MAR play critical roles in organizing large chromatin domains, such as establishing and maintaining heterochromatin-euchromatin boundaries¹²⁻¹⁴ and facilitating long-range interactions¹⁵.

SATB1, a well-known SIDD regulating protein, is mainly expressed in developing T cells¹⁶, epidermis¹⁷, and trophoblast stem cells (TSCs)^{18,19}. SATB1 binds to SIDD/BUR and then stabilizes the DNA double helix^{2,20} whereby it establishes and maintains large-scale euchromatin and heterochromatin domains¹²⁻¹⁴. For example, SATB1 directly binds to critical enhancers in pro- and pre-T cells, thereby activating gene expression in these cells while repressing the mature T cell fate²¹. In early embryogenesis, SATB1 and SATB2 form an intricate regulatory network in ESC differentiation²². Additionally, emerging evidence shows that *Satb1* is markedly upregulated during extraembryonic tissue development, thereby promoting extra-embryonic fates including trophoblast and primitive endoderm lineages^{18,19,23}.

In this work, we show that N^6 -mA plays a critical role during TSC development and differentiation by antagonizing SATB1 function at SIDD. Moreover, our latest work²⁴

demonstrates that ALKBH1, the DNA demethylase of N^6 -mA, highly prefers SIDD/BUR sequences as substrates which further strengthens the connection between N^6 -mA and DNA secondary structures during cell fate transitions.

N^6 -mA enrichment at SIDD during TSC development

Since N^6 -mA is present at low levels (6-7 ppm) in mESCs¹, we strove to search for conditions that upregulate N^6 -mA levels. Intriguingly, N^6 -mA levels are positively correlated with the *in vivo* developmental potential (pluripotency) of these conditions as determined by tetraploid complementation (4N); N^6 -mA is greatly diminished under traditional 2i conditions (ERK and GSK3b inhibitors) (4N negative) but retained under some of the alternative 2i conditions (4N positive) (Extended data Fig. 1a), which may explain the discrepancy in the literature²⁵.

Since previous studies showed that N^6 -mA demethylase *Alkbh1* deficient mice developed TE defects²⁶, we next investigated the role of N^6 -mA in TSC development and differentiation. To this end, we leveraged a *Cdx2* inducible expression system²⁷ (*iCdx2*) to model TSC development in cell culture, which manifests a well-synchronized and efficient cell fate transition process (Fig. 1a). As corroborated by a previous study²⁸, our RNA-seq analysis revealed that *iCdx2* cells first go through a transition state, then a TSC-LC state, before undergoing TE lineage differentiation (Extended data Fig. 1b). We found that the N^6 -mA level transiently increases at a time window that coincides with the emergence of TSC-LC, and then tapers off during TE lineage differentiation (Fig. 1a). Accordingly, the expression level of *Alkbh1* inversely correlates with N^6 -mA level during this cell-fate transition (Extended data Fig. 1b).

We then used DIP-seq (DNA immunoprecipitation followed by high-throughput sequencing) to interrogate N^6 -mA distribution at different stages during cell fate transition. To ensure the specificity towards N^6 -mA in DNA, samples were first treated with extensive RNase digestion (Methods). This protocol does not pull down any significant signal from RNA/DNA hybrids above background (unmodified DNA controls) (Extended data Fig. 1c). 95.55% of the sequencing reads were mapped (with unique mapping) to the murine genome with non-significant contributions from mitochondrial DNA and essentially no contributions from microbes. Peaks were called with high confidence using uniquely mapping reads against various controls, such as IgG or whole genome amplification (WGA) (Extended data Fig. 1d, and Methods). Consistently, the DIP-seq approach also detected more N^6 -mA peaks at the transition state than at the other stages (Extended data Fig. 1e) and the highest aggregate signal over peaks at the transition (Fig. 1b). Bioinformatic analysis revealed that N^6 -mA is mainly found in intergenic regions with a very high AT content (61.6% AT) (Extended data Fig. 1f), such as LINE-1s, but not other classes of transposons (Extended data Fig. 1g&h and Supplementary Table 1). Interestingly, 60-66% of N^6 -mA peaks are predicted to be SIDD with an established algorithm¹¹ (Extended data Fig. 1i). To confirm these results, we extracted M/SAR DNA with a well-established protocol²⁹ and quantified N^6 -mA levels with mass spectrometry (MS), which can distinguish methylated and unmethylated dA as well as RNA m⁶A (not detected in the samples) (Fig. 1c and Extended data Fig. 1j). This approach confirmed a 10-fold the enrichment of

N^6 -mA in M/SAR DNA and its upregulation during the cell fate transition. We next used ssDNA-seq^{10,11} to interrogate the SIDD regions in our experimental system (Fig. 1d). In self-renewing mESCs (TT2 WT), ssDNA-seq reads are enriched at SINE elements, and more signal is observed at the gene bodies and upstream of promoters of expressed versus non-expressed genes, similar to a previous study in B-cells¹¹ (Extended data Fig. 1k&l). Interestingly, ssDNA-seq peaks in *iCdx2* cells become enriched at LINE-1 elements, where N^6 -mA deposition is increased (Extended data Fig. 1l-n). Uniquely mapped N^6 -mA peaks significantly overlap with ssDNA regions (Fig. 1e). Similar overlap was also observed when all sequencing reads, including unique and non-unique, were included in the analysis²⁴. A representative trophoblast gene locus shows N^6 -mA directly overlapping with ssDNA peaks (Fig. 1f). Collectively, our genomic and biochemical data demonstrated that N^6 -mA is enriched at BUR/SIDD regions during ESC-to-TSC fate transition.

N^6 -mA abolishes SATB1 binding with DNA in vitro

We further investigated whether the presence of N^6 -mA affects the binding of regulators of SIDD, such as SATB1. First, we quantified the *in vitro* binding of SATB1 towards N^6 -mA modified oligodeoxynucleotides or unmodified controls. The structure of the CUT 1 domain, a DNA binding domain of SATB1^{20,30}, showed that the α 3 of SATB1 inserted into the major groove of dsDNA³⁰. Q390 and Q402 formed hydrogen bonds with the 7th deoxyadenosine of the substrates, while the 10th deoxyadenosine (referred to as 7th N^6 -mA or 10th N^6 -mA, respectively) is located outside the binding pocket³¹ (Fig. 2a). We quantitatively determined the binding affinities between SATB1 and SIDD sequences with N^6 -mA modified oligos at the 7th and 10th positions using isothermal titration calorimetry (ITC) (Fig. 2b). Of note, the binding affinities were determined with a SATB1 monomer, and the *in vivo* binding of SATB1 can be augmented by multimerization^{2,20}. The results showed that unmodified dsDNA binds to SATB1 with similar affinities to previous reports³¹ ($K_D=5.9 \mu\text{M}$). Importantly, this interaction is abolished by the 7th N^6 -mA (K_D undetectable), while the 10th N^6 -mA has no obvious effects ($K_D=7.0 \mu\text{M}$).

We next sought to engineer SATB1 mutants that can tolerate 7th N^6 -mA, which would be helpful in elucidating N^6 -mA function *in vivo*. The surface plasmon resonance imaging (SPRi) results revealed that the Q390RQ402A mutant binds similarly to both unmodified and 7th N^6 -mA SIDD sequences, while Q402N and Q402T mutants cannot bind either presumably due to the loss of coordinated interactions (Fig. 2c&d and Extended data Fig. 2a). Biolayer interferometry (BLI) further confirmed these results (Extended data Fig. 2b&c).

In addition, SATB1 DNA binding properties are also influenced by sequence specificity (Extended data Fig. 2d&e). A few other known AT-rich DNA binding proteins tested do not display N^6 -mA specificity. ARID3a recognized neither the unmodified nor the 7th N^6 -mA substrates, while FOXM1 and FOXD3 bind to both with similar binding affinities (Extended data Fig. 2f&g).

***N*⁶-mA antagonizes SATB1 binding *in vivo*.**

The biochemical results prompted us to investigate the impact of *N*⁶-mA on SATB1 interaction with chromatin. Consistent with previous studies^{18,19}, we found marked upregulation of SATB1 during the cell fate transition from ESC to a TSC-like fate, especially when *N*⁶-mA levels culminate. Paradoxically, despite this global increase SATB1 ChIP-seq (chromatin immunoprecipitation followed by high-throughput sequencing) demonstrated a 50% reduction in SATB1 peaks at days 5-6 when *N*⁶-mA levels culminate (Fig. 3a).

Further analysis revealed that SATB1 binding to chromatin is inversely correlated with *N*⁶-mA deposition and SATB1 sites are excluded from *N*⁶-mA sites (Extended data Fig. 3a&b). The differentially decreased SATB1 peaks (day 5 vs. day 0) significantly overlap with both the gained *N*⁶-mA sites at day 5 and the ssDNA-seq peaks (Extended data Fig. 3c). Concordantly, aggregation profiling also demonstrated that SATB1 ChIP-seq signals decrease where *N*⁶-mA signals increase during the cell fate transition (day 5 vs. day 0) (Fig. 3b).

We further interrogated *N*⁶-mA function by overexpressing *Alkbh1*, the DNA *N*⁶-mA demethylase whose activities have been identified by previous studies^{1,9,32,33} and recent structural studies^{24,34}. Consistent with several previous studies^{1,9,35}, overexpressed ALKBH1 is primarily localized to the nucleus, but not the mitochondria, in ESCs or TSC-LCs (Extended data Fig. 3d). As expected, *Alkbh1* overexpression (OE) greatly reduced *N*⁶-mA upregulation during cell fate transition (Extended data Fig. 3e), and a significant proportion of ALKBH1-regulated *N*⁶-mA peaks occur at SIDD/BURs identified by ssDNA-Seq (Extended data Fig. 3f). Upon *Alkbh1* overexpression, SATB1 ChIP-seq peak numbers increase greater than 2-fold (Fig. 3c), while overall SATB1 levels remain unchanged (Extended data Fig. 3g). The increased SATB1 peaks in *Alkbh1* OE significantly intersect with differentially decreased *N*⁶-mA peaks (*Alkbh1* OE vs. control); greater than 50% of those intersected peaks (SATB1 increased and *N*⁶-mA decreased) overlap with ssDNA peaks (Fig 3d). There is also a conspicuous increase in SATB1 ChIP-seq signal at differentially decreased *N*⁶-mA sites in *Alkbh1* OE in comparison to controls (Fig. 3e). Additionally, the increased SATB1 peaks are enriched on young LINE-1s, but not other classes of endogenous transposons, and overlapped with ssDNA regions (Extended data Fig. 3h&i and Supplementary table 1). Collectively, our data demonstrate that *N*⁶-mA antagonizes SATB1 binding to specific SIDD sequences *in vitro* and with chromatin *in vivo*.

***N*⁶-mA controls euchromatin boundaries.**

Given the well-known role of SIDD and SATB1 in facilitating long-range chromatin interactions, we used ATAC-seq to interrogate the chromatin accessibility in *iCdx2* cells (Extended data Fig. 4 and Fig. 4). Although ATAC-seq and *N*⁶-mA peaks are mutually exclusive at each other's peak centers (Fig. 4a), further analysis revealed a striking "peak and valley" distribution pattern. *N*⁶-mA signals culminate at the boundaries of the ATAC-seq inserts, mostly 0.5-1.0 kb away from the center of the ATAC-seq peaks; similarly, ATAC-seq signals also culminate at the boundaries of the *N*⁶-mA peaks (Fig.

4a). Consistently, correlation analysis demonstrated that although N^6 -mA and ATAC-seq peaks do not directly overlap at their peak summits, these peaks are significantly enriched at each other's boundaries (~0.5-1 kb) (Fig. 4b). Moreover, with *Alkbh1* OE, chromatin accessibility increases at the original (control) N^6 -mA peaks and spreads to much larger chromatin domains beyond the boundaries of N^6 -mA peaks, such as at the *Dkl1-Dio3* locus (Fig. 4c&d). In summary, N^6 -mA is enriched at boundaries of euchromatin as defined by ATAC-seq, and thereby prevents euchromatin from spreading into heterochromatin.

The vital role of N^6 -mA in early embryogenesis

Gene Set Enrichment Analysis (GSEA) demonstrated that LINE1 and genes involved in trophoblast differentiation are significantly upregulated in *Alkbh1* OE cells, with specificity for genes upregulated in spiral artery trophoblast giant cells (SpA-TGCs)³⁶, but not in the other trophoctoderm lineages (Fig. 5a and Extended data Fig. 5a&b), which are confirmed by RT-qPCR (Extended data Fig. 5c). Furthermore, a majority of imprinted genes, such as *H19*, which play critical roles in TSC and placental development^{37,38}, are upregulated upon N^6 -mA removal by *Alkbh1* OE; maternally expressed genes (MEGs) appear to be more severely affected than paternally expressed genes (PEGs) (Extended data Fig. 5d). We then confirmed the accelerated differentiation TGC-like cells with two methods to quantify polyploid (> 4N) cells (Extended data Fig. 5e&f). Thus, the acceleration of TGC formation only by the overexpression of catalytically active *Alkbh1* suggests that N^6 -mA maintains the TSC fate.

To investigate an early developmental phenotype in an *Alkbh1* deficient background, which was reported differently by previous studies^{26,39}, we generated a new strain of *Alkbh1* KO mice with the “KO-first” strategy⁴⁰ targeting exon 2 (tm1a allele, henceforth called 1a), and its floxed rescue (tm1c allele, henceforth called 1c) (Extended data Fig. 5g). Although heterozygous mice were born alive at expected Mendelian ratios, no homozygous *Alkbh1*^{1a/1a} mice were observed at birth, indicating a fully-penetrant embryonic lethality phenotype ($n = 68$ pups, 12 litters; Extended data Fig. 5h). This phenotype was completely rescued by removing the targeting cassette via Flp to create a floxed exon 2 (1c), which rules out the concerns of off-target or secondary mutations; *Alkbh1*^{1c/1c} mice are viable, fertile and indistinguishable from WT mice ($n = 52$ pups, 9 litters; Extended Data Fig. 5g&i). Furthermore, we leveraged an immunofluorescence approach⁹ to interrogate N^6 -mA levels in *iCdx2* cells and E8.5 cohort of embryos (Extended data Fig. 5j&k). Digesting with S1 nuclease, micrococcal nuclease, or DNase I (in addition to RNase A/T1) greatly reduces or completely abolishes the signal, whereas RNase H treatment does not affect the signal at all, indicating that the signal comes from DNA exclusively. Moreover, examination of the cohort of embryos from a timed mating experiment demonstrated that *Alkbh1*^{1a/1a} embryos develop at expected mendelian frequencies, but are all deceased and undergoing resorption by E17.5 ($n = 25$ embryos, 4 matings; Fig. 5b). Histological examination of E8.5 and E12.5 maternal-fetal interfaces, including the chorioallantoic plate and surrounding decidua, demonstrated significant reduction of TGCs by *Alkbh1* deletion (Fig. 5c&d). RNA-seq of the maternal-fetal interface at E10.5 (Extended data Fig. 5l and Supplementary Table 2) showed that a few key factors involved in trophoblast development were significantly reduced, including *Gpc3* which is expressed by differentiating human syncytiotrophoblasts⁴¹. Genes involved

in the regulation of maternal/fetal blood pressure and the hypoxia response, such as *Agt*⁴² and *Nppc*⁴³, were also aberrantly regulated. Of note, the TGC depletion phenotype and gene expression changes cannot be explained by expression of the gene *Nrp*, which partially overlaps with exon 1 of *Alkbh1*, because it is minimally expressed in placental tissue at similar levels across genotypes (Extended data Fig. 5m). In summary, *Alkbh1* deficiency *in vivo* increases placental *N*⁶-mA and inhibits TSC differentiation into TGCs, corroborating the results that *Alkbh1* overexpression facilitates TSC differentiation in the cell culture model.

Since our results showed that *N*⁶-mA antagonizes SATB1 interaction with chromatin, we next used a CRISPR-Cas9 approach to knockout *Satb1* in the *iCdx2* system (Extended data Fig. 6a). RNA-seq demonstrated that TE genes are inefficiently induced in *Satb1* KO cells (Extended data Fig. 6b). Consequently, *Satb1* deficiency significantly impairs TSC-LC differentiation into TGC-LC, as assayed by flow cytometry and morphology (Extended data Fig. 6c&d). 34% of the genes activated by *Satb1* are also repressed by *N*⁶-mA, which were identified in *Alkbh1* OE cells (Extended data Fig. 6e).

We next leveraged the aforementioned *N*⁶-mA tolerant SATB1 mutant (Q390RQ402A) (Fig. 2 and above) to further elucidate the epistasis between *Satb1* and *N*⁶-mA (Fig. 5e). We reconstituted *Satb1* KO cells with *Satb1* WT or the *N*⁶-mA tolerant *Satb1* mutant and demonstrated that reconstitution with the tolerant mutant generated more polyploid TGC-LCs at a faster rate than with the WT (Extended data Fig. 6c&d). Concordantly, the tolerant mutant rescued the expression of the downregulated trophoblast lineage markers in the *Satb1* KO cells more efficiently than the WT (Fig. 5e and Extended data Fig. 6f). Thus, these results lend further support to the notion that *N*⁶-mA modulates the ESC-to-TSC cell fate transition by antagonizing SATB1 function.

Discussion

In this work, we elucidate the molecular function of *N*⁶-mA in the regulation of chromatin structure by antagonizing SATB1 at SIDD during early development (Fig. 5f). *N*⁶-mA accumulates at the boundaries between euchromatin and heterochromatin and restricts euchromatin regions from spreading, thereby preventing ESCs from adopting a TSC fate. These findings are corroborated by genetic studies of *Alkbh1* KO mice. Of note, our most recent study has also demonstrated that ALKBH1, the DNA *N*⁶-mA demethylase, also prefers the unpairing SIDD/BUR sequences²⁴. These studies in conjunction reveal an unexpected mode of epigenetic regulation by the newly identified DNA modification *N*⁶-mA via DNA secondary structures during early development.

SATB1, a well-known SIDD binding factor, is a critical chromatin organizer for orchestrating large-scale chromatin structures^{12-14,20}. Our current work reveals the unexpected role of *N*⁶-mA as an antagonist of SATB1 function *in vitro* and *in vivo*. Strikingly, SATB1 binding to SIDD sequences is completely abolished in the presence of *N*⁶-mA (K_D undetectable), which is quite unusual for epigenetic effector binding proteins which usually manifest several-fold differences. Consistent with the long-appreciated role of SATB1, *N*⁶-mA is specifically enriched at the eu- and hetero- chromatin boundaries,

thereby controlling the spread of euchromatin. The major targets of the N^6 -mA/SIDD/SATB1 pathway appear to be imprinting genes, such as *H19*, which plays important roles in TSC differentiation³⁷. In parallel, this pathway may regulate gene expression by controlling the chromatin landscape at long-ranges, as indicated by chromatin accessibility changes in the current study, which also warrants further research into chromatin folding in the future.

In summary, our work demonstrates that N^6 -mA is a critical component in epigenetic regulation via antagonizing the function of SATB1. This study sheds new light on the mechanisms whereby epigenetic modification and DNA secondary structures regulate chromatin structure and gene expression in early development.

Methods

i*Cdx2* ES cell differentiation

The Tet-inducible-*Cdx2*-Flag ES cell line (i*Cdx2* cells) was obtained from NIA mouse ES cell bank²⁷. i*Cdx2* cells were maintained on feeder cells in DMEM supplemented with 15% fetal bovine serum (Gibco, Thermo Fisher Scientific), 1% non-essential amino acids, 2 mM l-glutamine, 1,000 units of mLIF (EMD Millipore), 0.1 mM β -mercaptoethanol (Sigma), antibiotics and 2 μ g/ml doxycycline (Sigma). When starting differentiation⁴⁴, i*Cdx2* cells were selectively seeded on gelatin-coated plates after stepwise elimination of the feeder MEF, then *Cdx2* overexpression was induced by removing doxycycline. The cells were split at day 1 after doxycycline withdrawal and cell samples were collected in the following days.

Mouse ES cells in 2i/L were cultured in commercially available 2i medium kit (Milipore, SF016-200). a2i/L medium contains a 1:1 mixture of DMEM/F12 supplemented with N2 (Invitrogen) and Neurobasal media with glutamine (Invitrogen) supplemented with B27 (Invitrogen), 1X Pen/Strep (Invitrogen), 1000 units of LIF (Milipore), 1.5 μ M CGP77675 (Tocris) and 3 μ M CHIR99021 (Tocris). PKCi/L medium contains a 1:1 mixture of DMEM/F12 supplemented with N2 (Invitrogen) and Neurobasal media with glutamine (Invitrogen) supplemented with B27 (Invitrogen), PenStrep (Invitrogen), 1000 units of mLIF (Milipore) and 5 μ M Gö6976 (Tocris).

Generation of *Satb1* knockout cell line

Two guide RNAs target flanking exon 6 were chosen. The guide RNA oligos were annealed and cloned into the lentiCRISPRv1 vector with hygromycin resistance. Then two constructs were transfected into ES cells by jetPRIME transfect reagent (Polyplus 11407). 24 hours after transfection, 400 μ g/ml hygromycin was added to the medium for 2 days. ES cells were then seeded at low density to obtain single-derived colonies. Then, 72 ES cell colonies were randomly picked up and screened by PCR genotyping that is illustrated in Extended Data Fig. 6a. PCR screening primer sequences can be found in Supplementary Table 3.

Establishment of *Satb1* overexpression cell line

The WT and mutant (Q390RQ402A) human *Satb1*-myc DNA sequence were inserted into pLV-EF1a-IRES-Hygro (pLV), a lentivirus-based vector with hygromycin resistance. For the *Satb1* rescue experiment, wild-type and Q390RQ402A mutant *Satb1* pLV constructs were

introduced to *Satb1* knockout ES cells by lentivirus; the original pLV empty vector was chosen as a control. After viral infections, the cells were selected with hygromycin at 400 µg/ml for 4 days, and then the cells were expanded for following experiments.

DNA extraction

Cellular genomic DNA was purified using the DNeasy kit (QIAGEN 69504) with minor modifications. In brief, the cell lysate was treated with RNase A1/T mix (Thermo Fisher EN0551) at 37 °C for 30 min. Then Proteinase K was added to the lysate and treated at 56 °C for 3 hours or overnight. The remaining steps were performed as described in the kit, and the DNA was eluted with ddH₂O. Furthermore, to remove trace RNA contamination, eluted DNA was treated with once more with RNase A/T1 for 2 hours and then RNase H (NEB M0297S) overnight, followed by phenol-chloroform extraction and ethanol precipitation.

N⁶-mA dot blotting

DNA samples were denatured at 95 °C for 5 min, cooled down on ice, and neutralized with 10% vol of 6.6 M ammonium acetate. Samples were spotted on the membrane (Amersham Hybond-N+, GE) and air dried for 5 min, then UV-crosslinked (2× auto-crosslink, 1800 UV Stratalinker, STRATAGENE). Membranes were blocked in blocking buffer (5% milk in PBST) for 1 hr at room temperature, then incubated with N⁶-mA antibody (1:1000, Synaptic Systems 202-003) at 4 °C overnight. Membranes were washed with PBST for 3 times and then incubated with HRP-linked secondary anti-rabbit IgG antibody (1:5000, Cell Signaling 7074S) for 1 hr at room temperature. After 3 washes with PBST, the membrane signal was detected with SuperSignal™ West Dura Extended Duration Substrate kit (Thermo Fisher 34076).

Matrix associated region (MAR) DNA extraction

The MAR DNA extraction was performed according to Mirkovitch *et al.*²⁹ and Pathak *et al.*⁴⁵ with modifications. Briefly, the cell pellets were resuspended with Buffer A (10 mM HEPES, PH 7.0, 10 mM KCl, 1.5 mM MgCl₂, 0.34 M sucrose, 10% glycerol, and proteinase inhibitor cocktail (Roche)) and treated with 0.1% Triton X-100 for 7 minutes on ice. After wash with Buffer A, the cells were resuspended with Buffer A and MNase buffer (NEB). 5 µl MNase was added into the cell suspension and incubated at 37 °C for 5 minutes, then 0.5 mM EGTA was added to terminate the digestion reaction. After washing with low salt buffer (Buffer A with 400 mM NaCl) for 10 minutes, the cell pellets were resuspended with high salt buffer (Buffer A with 2M NaCl) and rotated for 30 min at 4 °C. After high speed centrifugation (14,000 rpm) for 15 minutes, the cell pellets were resuspended with Buffer A and treated with RNase A/T1 for 2 hours and then RNase H at 37°C, then Proteinase K was added and incubated at 56 °C for 2 hours. The DNA was finally purified by phenol/chloroform extraction and ethanol precipitation.

Flow cytometry

Cells were collected using trypsin treatment. Cells were washed with ice cold PBS 3 times and fixed with cold 70 % ethanol overnight at 4 °C. After washing with cold PBS 3 times,

cells were resuspended and stained with 0.5 µg/ml DAPI for 10 min, then washed with cold PBS 2 more times. Samples were analyzed by a BD LSR II.

LC-MS/MS analysis of *N*⁶-methyl-2'-deoxyadenosine

To 1 µg of DNA, 0.1 U of nuclease P1, 0.25 nmol of *erythro*-9-(2-hydroxy-3-nonyl)-adenine hydrochloride (EHNA), and a 3-µL solution containing 300 mM sodium acetate (pH 5.6) and 10 mM ZnCl₂ were added. EHNA served as an inhibitor for adenine deaminase to minimize the deamination of adenosine. The reaction mixture was incubated at 37 °C for 24 h. Next, 0.1 U of alkaline phosphatase, 2.5 × 10⁻⁴ U of phosphodiesterase 1 and 4-µL of 0.5 M Tris-HCl buffer (pH 8.9) were added. After digestion at 37 °C for 2 h, the resulting digestion mixture was neutralized with 3 µL of 1.0 M formic acid. Then 200 pmol of ¹⁵N-labeled 2'-deoxyadenosine and 25 fmol of D₃-labeled *N*⁶-methyl-2'-deoxyadenosine were spiked into the digestion mixture, followed by chloroform extraction to remove the enzymes used for the DNA digestion. The resulting aqueous layer was dried and reconstituted in 10 µL of ddH₂O, to which 90 µL acetonitrile was added. The solution was subsequently centrifuged, and the supernatant was dried and re-dissolved in 50 µL of acetonitrile/ddH₂O (95:5, v/v) for LC-MS/MS analysis.

LC-MS/MS measurements were conducted on an Agilent 1200 series (Waldbronn, Germany) coupled with LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with a heated electrospray ionization source. Chromatographic separation was conducted on an Agilent Zorbax HILIC Plus column (2.1 mm × 100 mm, 3.5 µm). The mobile phase was (A) H₂O containing 0.2% formic acid and 10 mM ammonium formate (B) acetonitrile containing 0.2% formic acid, 2 mM ammonium formate and 0.06 mM malic acid. An isocratic mode of 5% A and 95% B was utilized, and the flow rate was 100 µL/min. The injection volume was 25 µL. High energy collisional dissociation was utilized and the transitions of *m/z* 252.1 → 136.0 (2'-deoxyadenosine), *m/z* 257.1 → 141.0 ([¹⁵N₅]2'-deoxyadenosine), *m/z* 266.1 → 150.0 (*N*⁶-methyl-2'-deoxyadenosine) and *m/z* 269.1 → 153.0 ([D₃]*N*⁶-methyl-2'-deoxyadenosine) were monitored for quantifications.

Protein expression and purification

The cDNA encoding SATB1 (amino acids 368-456), FOXM1 (amino acids 225-326) and FOXD3 (amino acids 140-236) were cloned into the pRSFDuet vector. Point mutations were generated using a site-directed mutagenesis kit (Stratagene). All unmodified and modified DNA oligos were synthesized by GenScript.

Proteins SATB1, FOXM1 and FOXD3 were expressed in *Escherichia coli* strain BL21 (DE3) in the presence of 0.4 mM IPTG. Overnight-induced cells were collected by centrifugation and re-suspended in lysis buffer: 100 mM NaCl, 20 mM Tris pH 7.5. Then cells were lysed with an Emulsiflex C3 (Avestin) high-pressure homogenizer. After centrifugation at 16,770×g, the supernatant was applied to the HisTrap column (GE Healthcare). The resultant protein was further purified by anion-exchange chromatography. The peaks eluted were applied to a Superdex 75 10/300GL (GE Healthcare) gel filtration column. All proteins were stored in 100 mM NaCl, 20 mM Tris pH 7.5 at ~20 mg/ml in a

–80 °C freezer. Mutants were expressed and purified using essentially the same procedure as for the wild-type proteins.

Surface plasmon resonance imaging

To measure the interaction between immobilized proteins and flowing nucleic acids, an SPR imaging instrument (Kx5, Plexera, USA) was used to monitor the whole procedure in real-time. Briefly, a chip with well-prepared biomolecular microarray was assembled with a plastic flow cell for sample loading. The DNA samples were prepared at a series of concentrations (2.5, 1.2, 0.6 and 0.3 μM) in TBS running buffer (20 mM Tris, 100 mM NaCl, pH 7.5) while a 10 mM glycine-HCl buffer (pH 2.0) was used as regeneration buffer. A typical binding curve was obtained by flowing DNA samples at 2 $\mu\text{l/s}$ for 300 s association and then flowing running buffer for 300 s dissociation, followed by 200 s regeneration buffer at 3 $\mu\text{l/s}$. Binding data were collected and analyzed by a commercial SPRi analysis software (Plexera SPR Data Analysis Module, Plexera, USA).

Biolayer interferometry assays

His-tagged wild type and mutant SATB1 proteins were immobilized onto anti-HIS antibody-coated biosensors (Fortebio, Pall Life Sciences). After being washed with binding buffer (20 mM Tris pH 7.5, 100 mM NaCl), the sensors were dipped into binding buffers containing unmodified and modified double-strand DNA oligos at various concentrations for 5 min for complete association followed by 5 min disassociation in binding buffer without DNA. Kinetics were recorded using Octet RED96 (Fortebio, Pall Life Sciences) at 1,000 r.p.m. vibration at 25 °C.

Isothermal titration calorimetry

For ITC measurement, synthetic DNA oligos and the purified proteins were extensively dialyzed against ITC buffer: 20 mM Tris pH 7.5, 100 mM NaCl. The titration was performed using a MicroCal iTC200 system (GE Healthcare) at 25 °C. Each ITC titration consisted of 17 successive injections with 0.4 μl for the first and 2.4 μl for the rest. Usually, proteins at 0.5 mM were titrated into DNA oligos at 0.05 mM. The resultant ITC curves were processed using Origin 7.0 software (OriginLab) according to the ‘One Set of Sites’ fitting model.

*N*⁶-mA-DNA-immunoprecipitation (DIP) sequencing and analysis.

*N*⁶-mA-DIP was performed as described in Wu. *et al*¹. Briefly, 5 μg extracted genomic DNA was sonicated to 200–500 bp with a Bioruptor. Then, NEBNext adaptors were ligated to genomic DNA fragments after end repair following the NEBNext-UltraII library prep manual. The ligated DNA fragments were denatured at 95 °C for 10 min and chilled on ice to form single-stranded DNA fragments. *N*⁶-mA enriched DNA fragments were enriched and purified using *N*⁶-mA antibody (5 μg for each reaction, Synaptic Systems 202-003) according to the Active Motif hMeDIP protocol. Immunoprecipitated and input DNA were PCR amplified with NEBNext Multiplex Oligos for Illumina indexing primers. Sequencing was performed with the Illumina HiSeq 4000 platform. Basecalling and adapter sequence trimming were processed using the standard Illumina workflow. After sequencing and filtering, high-quality raw reads were aligned to the mouse genome (UCSC, mm9)

with bowtie (2.3.1, default settings)⁴⁶. To identify uniquely mapped N^6 -mA peaks prior to peak calling, multiple-alignment reads were filtered using Samtools to retain reads with map quality (MapQ) scores above 20 and with properly oriented read mates. N^6 -mA enriched regions were called against single-stranded input DNA as control with SICER (version 1.1, window size 200, gap 600, false discovery rate (FDR) < 0.01)⁴⁷. Including IgG pull-down and whole genome wide amplification as controls in peak calling generate very similar results (see Extended data Fig. 1). High overlaps were observed among biological and technical replicates of N^6 -mA DIP-seq peaks during cell fate transition. On average, 73.4% of total peaks during transition are overlapped in pairwise comparison between replicates; and 59.6% of unique peaks that excluded some of the multiple mapping reads are overlapped. Differential N^6 -mA deposition regions among two samples of interest were identified with the SICER-df setting. N^6 -mA genome browser tracks were generated by deepTools v3.1.3 with reads extended to match the fragment size defined by the two read mates and normalized to reads per genomic content (RPGC). Categorical mapping and quantification on endogenous transposons was performed by counting reads on each repeat element based on the UCSC Genome Browser RepeatMaster track of the reference genome mm9. Deposition on each endogenous transposon category was enumerated by BEDTools coverage and normalized by per million mapped reads. LINE1 transposons were grouped by evolutionary age into young: < 1.5 Myr (L1Md_A, L1MdT, L1Md_Gf); middle: > 1.5 Myr & < 6 Myr (L1Md_F, L1VL); old > 6 Myr (Lx, L1_Mus) as previously described⁴⁸. Intersections were performed by USeq⁴⁹. Part of the data analysis was done by in-house customized scripts in R, Python or Perl.

To test the N^6 -mA antibody binding to DNA and RNA-DNA hybrid in DIP, 1 pmol of different oligos (single N^6 -mA containing oligo, double N^6 -mA containing oligos, single unmodified DNA oligo and DNA-RNA hybrid with RNA m6A modification) were added into 5 μ g sonicated genomic DNA. The spiked-in oligos were heated at 95°C for 10 minutes then slowly cooled down in a PCR machine and tested by agarose electrophoresis. Then DIP was performed according to the denaturing DNA IP protocol (55010, Active Motif). After denaturing under 95°C for 10 minutes, DNA was quickly transferred on ice for 5 minutes, then 5 μ g N^6 -mA antibodies or Rabbit IgG were added and incubated overnight at 4 °C. Then Sheep Anti-Rabbit IgG dynabeads (Invitrogen 11203D) were added and incubated for 2 hours at 4 °C. After wash and elution using the buffer from the Active Motif hMeDIP kit, the enriched DNA fragments were purified by Phenol-chloroform extraction followed by ethanol precipitation. Real-time PCR was performed with the iQ™ SYBR® Green Supermix (BioRAD 170-8882) and quantified by a CFX96 system or CFX384 system (BioRAD).

Single stranded DNA-sequencing (ssDNA-seq).

ssDNA-seq was performed as previously described with minor modifications¹¹. 8×10^7 ES cells cultured in feeder-free conditions. Low salt buffer (15 mM Tris-HCl, pH 7.5, 60 mM KCl, 15 mM NaCl, 5 mM MgCl₂, 0.5 mM EGTA, and 300 mM sucrose) at 37 °C was added for 5 min. Cells were then treated with 100 mM KMnO₄ for 80 s at 37 °C (“Treated” sample), and the reaction was quenched by the addition of 50 mM EDTA, 700 mM β -mercaptoethanol, and 1% (w/v) SDS. In parallel, the same number of

cells were treated with water (“Blank” sample) and processed similarly. After overnight proteinase K digestion at 37 °C, DNA was extracted twice with phenol and once more with phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v; PCI), and precipitated with 2 M ammonium acetate in ethanol (PCI extraction with ethanol precipitation).

Genomic DNA was resuspended in 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA buffer (TE buffer) and treated with RNase A/T1 (1:50 dilution, 40 ug/mL and 100 U/mL, respectively) for 1 h at 37 °C, then PCI extracted with ethanol precipitation and resuspended in TE buffer. Free 3' ends formed due to random DNA breakage during sample preparation were blocked by treatment with 100 µM cordycepin-5'-triphosphate sodium salt (Sigma Aldrich C9137) and 400 U of Terminal transferase (NEB M0315; TdT) in 1 × TdT buffer in a reaction volume of 3 mL for 2 hrs at 37 °C. DNA was PCI extracted with ethanol precipitation and resuspended in TE buffer.

Digestion of ssDNA was carried out by dividing each of the Treated and Blank samples into separate reactions with 0, 50, 100, or 200 U of S1 nuclease (Thermo Fisher EN0321) in 500 µL of the supplied reaction buffer and incubating for 30 min at 37 °C. DNA was PCI extracted with ethanol precipitation and resuspended in TE buffer. Based on a fragment size distribution between 2-10 kb, the 50 U S1-treated samples (Treated and Blank) were chosen for further processing. Next, 35 ug of DNA was biotinylated with 300 U TdT, 250 µM dATP, 250 µM dCTP, and 50 µM Biotin-16-dUTP (Roche 11093070910) in a final volume of 300 µL of 1X TdT buffer at 37°C for 30 min. Reactions were stopped with 15 µL of 0.5 M EDTA. PCI extraction with ethanol precipitation was performed, followed by a second ethanol precipitation to remove free biotin, and samples were resuspended in TE buffer.

Samples were sonicated with a Covaris S220 to generate DNA fragments between 200 and 700 bp and PCI extracted with ethanol precipitation to remove free biotin. DNA was resuspended in TE buffer, and 10% of the Treated sample was saved for the input control. With the remaining sample, biotinylated fragments were pulled down using streptavidin-coated beads using the manufacturer's protocol (Dynabeads kilobase BINDER Kit, Thermo Fisher 60101). Fragments were released by incubating beads with 50 U S1 nuclease in 100 µL reaction buffer for 15 min at 37°C. Finally, DNA was purified with a MinElute PCR Purification Kit (Qiagen) and eluted in 30 µL of the manufacturer's elution buffer. Sample concentrations were measured using a Qubit 3.0 fluorometer, and at least a 10-fold greater pulldown efficiency for Treated versus Blank was confirmed.

ssDNA-seq library preparation and high-throughput sequencing data processing and analysis.

“Treated” samples and their input controls were further processed for high-throughput sequencing. Sequencing libraries were made using the NEBNext Ultra II DNA Library Prep kit with 10-100 ng of each sample and input. Libraries were prepared according to the manufacturer's instructions without size selection. Library concentrations were measured with Qubit and quality control performed with Agilent 2100 Bioanalyzer. Libraries were pooled and sequenced paired-end 2x100 bp in one lane of an Illumina HiSeq4000.

Sequencing reads were filtered and pre-analyzed with the Illumina standard workflow. After filtering, raw reads (in fastq format) were aligned to the mouse genome (UCSC, mm9) with Bowtie2 (2.3.1) using the default settings. Non-uniquely mapped reads were filtered out using a MapQ > 10. PCR and optical duplicates were then removed with Picard (2.9.0; <http://broadinstitute.github.io/picard>). BigWigs for genome browser tracks were generated using deepTools and normalized to reads per genomic content (RPGC). Aggregation profiles and heatmaps were generated from bigWigs over regions of interest, and in indicated plots, presented as a log₂ fold change over input. After alignment and processing, ssDNA enriched regions were called with SICER (version 1.1, FDR < 0.01, input DNA as control)⁴⁷, with window size 200 bp and gap size 600 bp.

Analysis of ssDNA enrichment in transposable elements.

Raw fastq reads were analyzed by SalmonTE (version 0.8.2) with default parameters and fold enrichment of samples over input was calculated for each family of transposable elements in the mouse index⁵⁰. Standalone aggregation profiles were constructed from bigWigs without filtering for uniquely mapped reads over RepeatMasker annotations of full-length (longer than 6 kb) young LINE1 elements (L1Md_A, Tf, and Gf). Heatmaps used only uniquely mapped data over the same annotated LINE1 regions.

Chromatin-immunoprecipitation

10⁷ Cells were washed once with DPBS and then fixed in 5 mL DMEM medium with 1% formaldehyde for 10 min at RT with rotation. Fixation was terminated with addition of 100 mM glycine followed by 5 min rotation at RT. Cells were then scraped and spun down at 4 °C at 1000 rpm for 5 min, followed by two washes of ice-cold PBS. Cell pellets were resuspended in 1 mL Lysis Buffer 1 (50 mM Hepes-KOH, pH7.5, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40, and 0.25% Triton X-100) followed by rotation at 4 °C for 10 min. Cell pellets were spun down at 1400×g for 5 min and resuspended in 1 mL Lysis Buffer 2 buffer (10 mM Tris-HCl, pH8.0, 200 mM NaCl, 1 mM EDTA, and 0.5 mM EGTA). Following another 10 min incubation at 4 °C, cell pellets were again spun down at 1400×g for 5 min and resuspended in 200 to 300 μL Lysis Buffer 3 (10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.5% Na-Deoxycholate and 0.5% N-lauroylsarcosine). Cells were then sonicated to about 500 bp fragments using a Covaris S200 with the following parameters: peak power: 120, duty factor: 2.0, cycle/burst: 200, time: 10 min, and power output: 2.1. The sonicated cell solution was then diluted with Lysis buffer 3 plus 1% Triton X-100, incubated for 15 min with rotation at 4 °C and spun down at 20,000×g for 15 min. 10 μL anti-SATB1 antibody (Abcam ab109122) was added to the supernatant; an equivalent amount of rabbit IgG antibody was used as a control. After overnight incubation at 4 °C, antibodies were pulled down using anti-Rabbit Dynabeads (10 μL per 1 μg antibody). After incubation with beads for 2 hr at 4 °C, samples were extensively washed 3 times with RIPA buffer (50 mM Hepes-KOH, pH 7.6, 500 mM LiCl, 1 mM EDTA, 1% NP40 and 0.7% Na-deoxycholate), rotating at 4°C for 10 min each time. After the last wash, beads were washed once with 1×TE, 50 mM NaCl buffer. DNA was then eluted in elution buffer (50 mM Tris-HCl, pH8.0, 10 mM EDTA and 1% SDS) twice at 65 °C for 20 min. Eluted DNA was then incubated at 65 °C overnight to reverse cross-link. The solution was treated with RNase A1/T mix and incubated at 37 °C for 1 hr, followed by the

addition of 3 μ L of proteinase K and incubation for 2 hr at 56 °C. The solution was then phenol-chloroform extracted once followed by DNA extraction using the MinElute PCR Purification Kit (Qiagen 28004). DNA was eluted with EB buffer for downstream library preparation using the NEBNext UltraII library preparation kit. Sequencing was performed with HiSeq 2000, and the output sequencing reads were filtered and pre-analyzed with Illumina standard workflow. After filtration, the qualified tags (in fastq format) were aligned to the mouse genome (mm9) with bowtie2 (2.3.1)⁴⁶. After mapping, multiple-alignment reads were filtered using Samtools to maintain reads with quality score above 20 and properly paired. Then SATB1 enriched regions were called with SICER (window size 200, gap 600, FDR < 0.01, input DNA as control). Quantification of transposable elements was performed as described for N⁶-mA-DIP-seq.

Statistics on peak intersections

USeq was used to intersect peaks and determine an empirical p-value from two given sets of peaks. Each peak in the second set was shuffled within the same chromosome, and this “genome random” set was intersected with the first peak set. This process was repeated 1×10^6 times, after which the mean intersections of the randomized set and the number of times there were more intersections in the genome random than the original set was determined. A count of zero trials where genome random was greater than the original was reported as $P < 1 \times 10^6$. It was also verified that the original had > 2-fold enrichment of intersections compared to the mean of genome random.

ATAC-seq

ATAC-seq was performed as described⁵¹. In brief, 50,000 cells were collected and resuspended in ice-cold ATAC-RSB buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin). After incubation on ice for 3 min, the cells were washed with ice-cold ATAC-RSB buffer without NP40. Then cells were resuspended in transposition mix composed of 0.33 \times DPBS, 0.01% Tween-20, 1 \times TD buffer (Nextera DNA Library Prep Kit, Illumina) and Transposase (100 nM final concentration, Illumina). Transposition reaction was incubated at 37 °C for 30 minutes in a thermomixer with mixing. Then transposed DNA was purified by the MinElute PCR Purification Kit. Eluted DNA was amplified using NEB Next UltraII Q5 Master Mix for 5 cycles. A real-time qPCR amplification was performed to determine additional cycles. The amplified DNA libraries were size-selected and purified by AMPure beads according to the manufacturer’s instructions. After sequencing and filtering, high quality raw reads were aligned to the mouse genome (UCSC, mm9) with bowtie2 (2.3.1, default settings). PCR duplicates were removed by Samtools. ATAC peaks were identified using MACS2 (2.1.1)⁵² with BAMPE mode. Peaks overlapping with Encode blacklist regions were removed. To generate bigwig files, reads with insertion sizes greater than 120 bp were retained and normalized by reads per genomic content (RPGC). Quantification of intersections was performed by USeq⁴⁹.

Maternal/fetal interface RNA extraction

At 10.5 days post conception, whole maternal fetal interfaces were collected by excising the decidua and placenta from the embryo/myometrium. Samples were stored at –80 °C until homogenization and RNA extraction. Care was taken to reduce excess myometrial,

perimetrium, amnion, and umbilicus tissue. Tissues were lysed using the QIAzol lysis reagent (Qiagen 79306) in conjunction with the Tissuelyser II homogenization system (Qiagen 85300) as per the manufacturer's instructions.

RNA-seq and RT-qPCR approaches

RNA was extracted with miRNeasy kit (Qiagen 217004) and standard RNA protocol. The quality of RNA samples was measured using the Agilent Bioanalyzer. RNA-seq libraries were constructed with Illumina Truseq stranded mRNA library prep kit and sequenced on an Illumina HiSeq 2000. Transcriptome mapping was performed with STAR version 2.5.2b using mm9 Gencode vM1 release exon/splice-junction annotation. Differential expression analysis was performed using cufflinks version 2.2.1 using UCSC gene annotation with first-strand library type and the correction of sequencing bias and multiple reads mapping. Gene set enrichment analysis (GSEA) on RNAseq data was conducted through the GSEA Preranked tool⁵³ based on cuffdiff output and the target gene sets (SpA-TGCs or Prdm1-TGCs from Nelson *et al.*, *Nat Commun* 2016). Statistical significance for GSEA is determined empirically by permuting the phenotype labels 1,000 times to generate a null distribution of the enrichment score, then comparing the actual enrichment score to the null. Gene ontology enrichment analysis was performed using PANTHER version 14⁵⁴. Part of the data analysis was done by in-house customized scripts in R. Supplemental RNA-seq analysis was performed using HTSEQ -count and DESeq2 using the local alignment paradigm.

For RT-qPCR, cDNA libraries were generated with the iScript™ cDNA Synthesis Kit (BioRAD 170-8891). Real-time PCR was performed with the iQ™ SYBR® Green Supermix (BioRAD 170-8882) and quantified by a CFX96 system or CFX384 system (BioRAD).

Mouse breeding

Animal experiments were performed in accordance with all relevant ethical regulations and guidelines of Yale University's Institutional Animal Care and Use Committee and the NIH. C57BL/6N-*Alkbh1^{tm1a}(EUCOMM)Hmgul^leg* (*Alkbh1^{tm1a/wt}*) males and females were purchased from the EMMA repository (derived from a C57BL/6N background ES-line)⁵⁵. Heterozygous females were backcrossed once onto C57BL/6J males. Next, females were crossed with FLPe-expressing males (Jackson Laboratory) to generate mice with the *Alkbh1^{tm1c}* allele. Brother-sister matings of 6-10 week-old *Alkbh1^{tm1a/wt}* or *Alkbh1^{tm1a/tm1c}* mice were performed to remove the FLPe allele and then maintained for > 5 generations before performing subsequent experiments.

Timed matings were performed by synchronizing the estrus cycle of female *Alkbh1^{tm1a}* heterozygotes via exposure to male soiled bedding, and subsequently placing 2 females per *Alkbh1^{tm1a}* male heterozygote per cage. Females were checked for vaginal plugs daily in the early morning, and the presence of a plug was standardized as day/E 0.5. Pregnant animals were sacrificed via CO₂ per university guidelines, and embryos/placental tissue were dissected out. Sample sizes were not predetermined for mating experiments; 2-3

biological samples with multiple sections were chosen for staining and histology. Blinding was not performed, and randomization was not applicable to this study.

Placental tissue histological and N⁶-mA immunofluorescence (IF) staining

Placental tissue was immediately fixed in 4% formaldehyde overnight and transferred to 70% ethanol. Placental tissue was paraffin embedded and sectioned by Yale Pathology Tissue Services (YPTS); subsequent H&E staining of placental tissues were performed by YPTS.

N⁶-mA IF of paraffin embedded tissues was performed via the following protocol: slides were heated to 55 °C for 45 min, and then deparaffinized in xylenes for 15 min. Rehydration of the tissue was performed via submersion in 100%, 90%, 80%, 70% EtOH for 5 min each, followed by 10 min in PBS. Cells were permeabilized by incubation in 0.2% Triton-X 100 in PBS for 20 min followed by washes with 0.1% Tween 20 in PBS (PBS-T). DNA was hydrolyzed via incubation in 2N HCl for 20 min, followed by neutralization with 0.1 M sodium borate for 30 min. After washes with PBS-T, slides were blocked with 6 µL/mL RNase A/T1 (Thermo Fisher EN0551) with additional 50 µg/mL RNase A (Qiagen 19101), and 50 U/mL RNase H (NEB M0297) as indicated in 10% goat serum PBS-T overnight at 37 °C. Slides were incubated in 1:400 α-N⁶-Methyladenine monoclonal antibody (Cell Signaling Technologies 56593, Lot 1) overnight at 4 °C. After rinsing in PBS-T, slides were incubated with 1:1000 Goat-anti-rabbit IgG Alexa Fluor 488 (Invitrogen A11008) at RT for 60 min. Slides were rinsed in PBS-T for 5 min, incubated in 1:1000 DAPI for 3 min and rinsed in PBS 3x. After IF staining slides were preserved using Fluoromount-G (Electron Microscopy Sciences, 17984-25) and stored at 4 °C before imaging.

Immunofluorescence staining of cell culture samples

The stable ES cells carrying pLV-Flag-HA-*Alkbh1* and pLV-*Alkbh1*-HA constructs were grown on gelatin treated slides for 24 h. Cells were fixed with 1% paraformaldehyde and permeabilized with 0.2% triton X-100 in PBS. Then the slides were incubated in blocking buffer (2% bovine serum albumin, 0.2% triton X-100 in PBS) for 1 hour, incubated with 1:500 HA-tag antibody (Cell Signaling 3724S) in blocking buffer at 4 °C overnight, and then incubated with the secondary antibody. DNA was stained with DAPI for 5 minutes. Slides were mounted with Fluormount-G.

N⁶-mA staining in cultured cells was performed as described in Xie *et al.*⁹ with minor modifications. Cells were grown on gelatin treated Millicell® EZ SLIDES (Millipore PEZGS0816). Cells were then fixed with 1% paraformaldehyde and permeabilized with 0.2% triton X-100 in PBS. After PBS wash, the slides were treated with 2N HCl for 20 minutes, then neutralized with 0.1 M sodium borate for 30 min. After PBS washes, the slides were blocked in blocking buffer (2% bovine serum albumin, 0.2% triton X-100 in PBS) with RNase A/T1 for 1 hour, incubated with 1:400 N⁶-mA antibody (202-003, Synaptic Systems) in blocking buffer at 4 °C overnight. After PBS washes, the slides were incubated with Goat anti-Rabbit IgG, Alexa Fluor Plus 594 (Invitrogen A32740) for 1 hour. DNA was stained with DAPI for 5 minutes. Slides were mounted with Fluormount-G. Images were acquired with a Leica SP5 confocal laser microscope. As nuclease controls, slides were treated with

20 U/ml DNase I (NEB M0303S) or 20,000 gel units/ml MNase (NEB M0247S) overnight at 37 °C before the 2 N HCl DNA denaturing step. 100 U/ml RNase H (NEB M0297S) or 2 U/ml S1 nuclease (Thermo Fisher EN0321) treatment were carried out overnight at 37 °C after DNA denaturing with 2 N HCl.

Microscopy

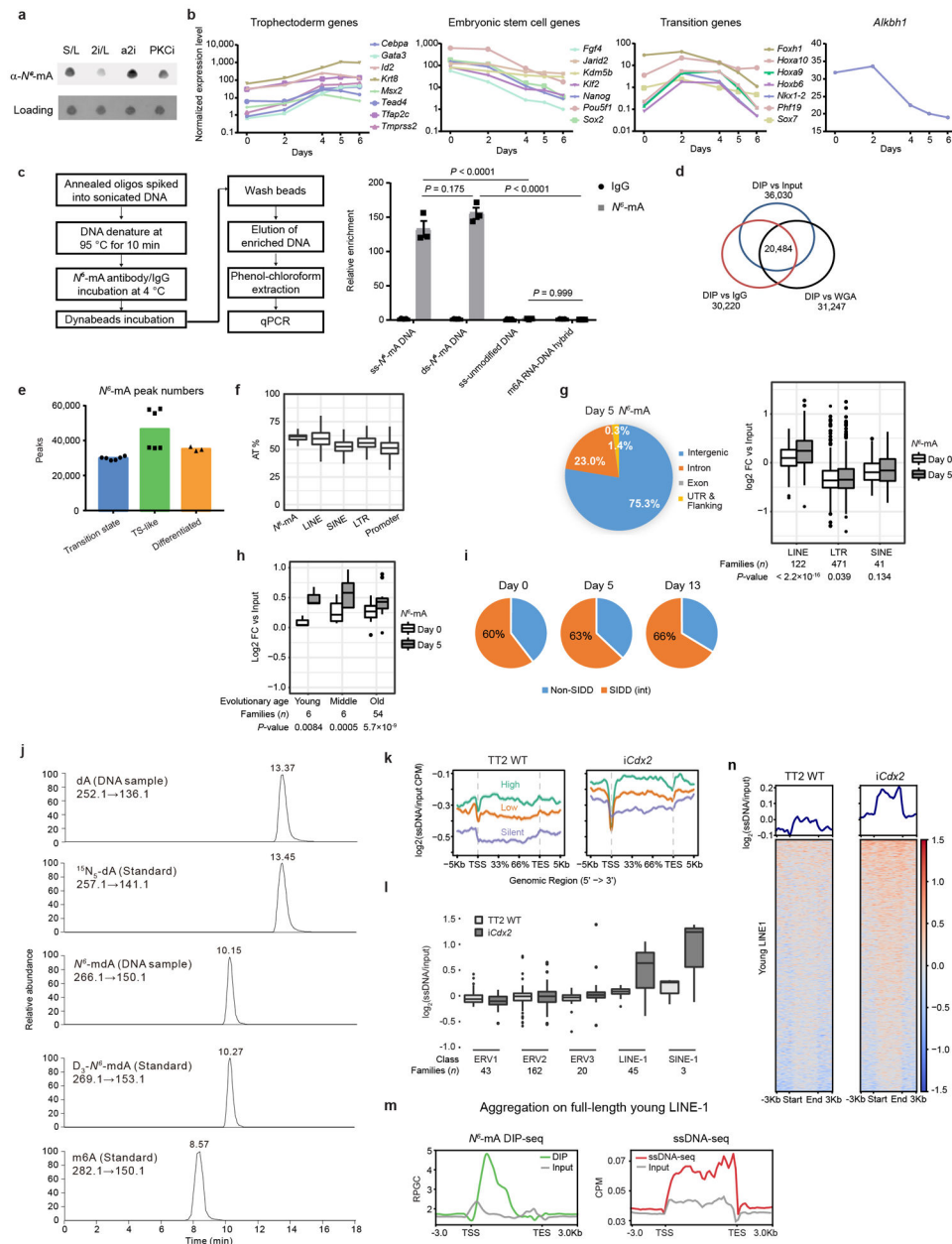
Immunofluorescence microscopy was performed within 7 days of staining using (1) an Olympus IX81 Fluorescencexs microscope at 20x magnification and equal exposure time for each respective channel for tissue sections and (2) a Leica TCS SP5 Spectral Confocal Microscope at 40x magnification and 2.0x zoom for cultured cells. Images were captured and prepared using (1) MetaMorph imaging software (Molecular Devices LLC, v7.7.0) or (2) LAS AF software suite (Leica Microsystems, v2.6.0). Brightfield microscopy was performed using an Olympus BX51 microscope at 10x and processed using CellSens (Olympus, v1.17).

For H&E TGC quantification, maternal-fetal interfaces were sectioned to 8 µm thick, and 16 sections were collected from each implantation site. 5 sections spanning 128 µm were chosen for TGC quantification from representative interfaces, and TGC's between the chorioallantoic plate (E8.5)/labyrinth (E12.5) and decidua were counted manually. Quantitative analysis was performed on 3 litter matched sets of embryos from 3 different litters by individuals blinded to the genotype of the sample.

Data Availability Statement

Sequencing data generated for this study have been deposited in the Gene Expression Omnibus database with accession numbers GSE126077 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126077>) and GSE126227 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126227>). Source data for animal-based data in Fig. 5 and Extended Data Fig. 5 are available with the paper. All other data, materials, and custom code are available from the corresponding author upon reasonable request.

Extended Data

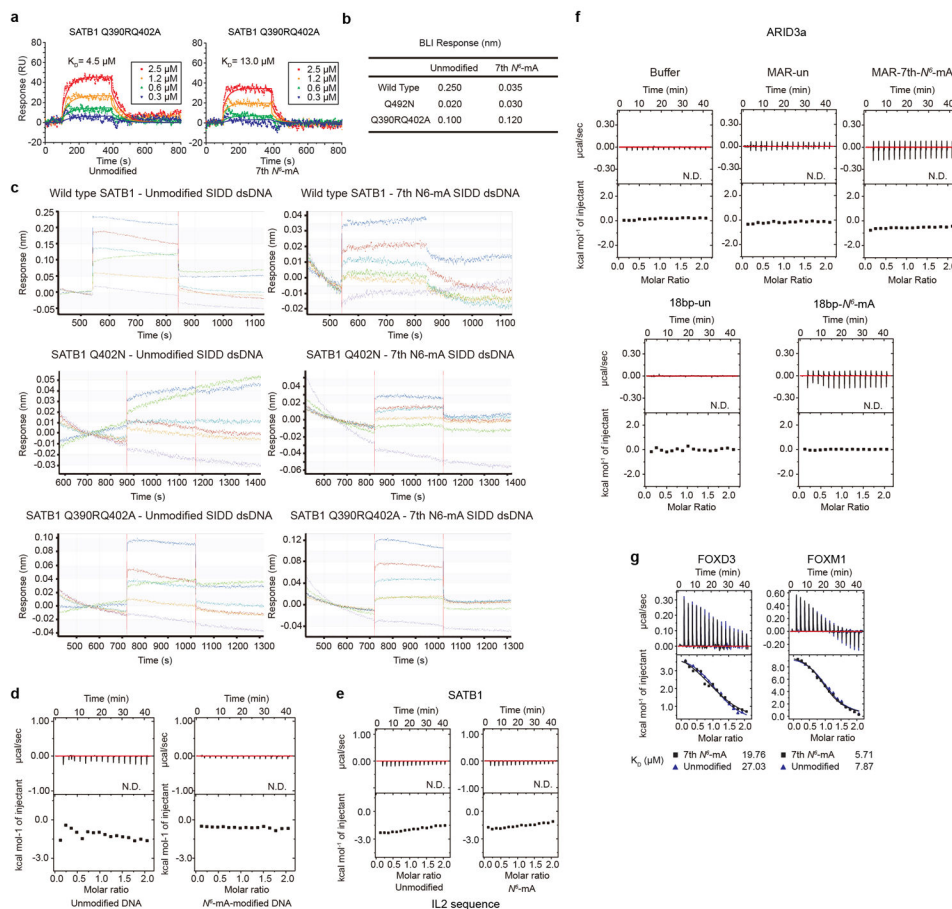
Extended Data Figure 1. N^6 -mA upregulation during TSC development

a. Dot blotting of N^6 -mA in ES cells cultured in various conditions. Methylene blue staining was performed as a DNA loading control. S/L, serum and LIF; 2i/L, Gsk3 β , ERK inhibitor, and LIF; a2i, Gsk3 β and Src inhibitor; PKCi, PKC inhibitor. The experiments were repeated independently 3 times with similar results. For blot source data, see Supplementary Figure 1.

b. RNA-seq analysis revealed a transition state with unique gene expression signatures prior to TSC-LC. Gene expression in fragments per kilobase of transcript per million mapped reads (FPKM) of selected genes, as well as *Alkbh1*, is shown during cell differentiation at indicated days.

- c.** qPCR analysis of N^6 -mA DIP on 1 pmol of the indicated spiked-in oligos. Fold enrichment was calculated relative to IgG. One-way ANOVA ($P < 0.0001$) followed by Tukey's multiple comparisons test. Mean \pm s.e.m. of 3 independent samples. ss- N^6 -mA DNA, single strand DNA with N^6 -mA modification; ss-unmodified DNA, single strand DNA with no modification; ds- N^6 -mA DNA, annealed double strand DNA with N^6 -mA modification; m6A RNA-DNA hybrids, DNA and RNA hybrid with RNA m6A modification. Oligo sequences are in Supplementary Table 3.
- d.** Intersection of peaks called against single-stranded input, IgG, and whole genome amplification (WGA) shows high specificity of N^6 -mA DIP-seq.
- e.** Number of N^6 -mA DIP-seq peaks at indicated points of cell fate transition. Transition state, days 0 and 3; TS-like, days 5 and 7; differentiated, day 13. Three biological replicates per day.
- f.** AT percentages at day 5 versus day 0 differentially increased N^6 -mA peaks (24,072); LINE (962,656), SINE (1,505,855), LTR (824,426) retrotransposon loci from RepeatMasker; and promoters (55,419) defined by transcription start sites from UCSC Genes to 1 kilobase upstream. AT percentage at each genomic location was calculated by $AT\% = (A+T)/(A+T+G+C) \times 100\%$ and the distribution for each group displayed in a box plot.
- g.** *Left:* Annotation of N^6 -mA peaks identified at day 5 by Cis-regulatory Element Annotation System (CEAS). *Right:* Log₂ fold-change (FC) enrichment of N^6 -mA DIP-seq over input reads for three major classes of transposable elements (TE). Day 5 reads are significantly enriched for LINE-1 over day 0. Number of families plotted per TE class and P values from two-sided paired t -tests of day 5 vs day 0 are provided. Raw data are in Supplementary Table 1.
- h.** Log₂ FC enrichment of N^6 -mA DIP-seq over input reads for different LINE1 families grouped by evolutionary age (young: < 1.5 Myr; middle: > 1.5 Myr & < 6 Myr; old > 6 Myr). Number of families plotted per group and P values from two-sided paired t -tests of day 5 vs day 0 are provided. Raw data are in Supplementary Table 1.
- i.** More than 60% of N^6 -mA peaks identified at days 0, 5 and 13 are bioinformatically predicted to be SIDD.
- j.** LC-MS/MS analysis of dA and N^6 -mdA from digested DNA samples. Stable isotope labeled N^6 -mdA and dA standards are indicated. The elution time and molecular weight of N^6 -mdA are different from the m6A RNA standard shown at the bottom. The experiments were repeated independently 3 times with similar results.
- k.** Average log₂ FC signal of ssDNA-seq in TT2 wild type (WT) and *iCdx2* ES at highly expressed (FPKM ≥ 10), low ($0.5 \leq$ FPKM < 10), and silent genes (FPKM < 0.5). Kb, kilobase; TSS, transcription start site; TES, transcription end site.
- l.** Log₂ FC enrichment of ssDNA-seq over input at each locus for selected major retrotransposon classes. Number of loci per class are provided below each box plot.
- m.** N^6 -mA DIP-seq (*left*) and ssDNA-seq (*right*) aggregation of mapped reads over annotated full-length young LINE-1 elements. TSS, transcription start site; TES, transcription end site; CPM, counts per million.
- n.** Average signal and heatmaps of ssDNA-seq in TT2 WT and *iCdx2* ES cells at young LINE1 elements.

For box plots, boxes represent the interquartile range (IQR), horizontal line the median, and whiskers the minimum and maximum values; whiskers extend up to $1.5 \times$ IQR when outliers that fall outside that range are present.



Extended Data Figure 2. Summary data and BLI binding curves of SATB1 and mutants. ITC titration curves and fitting curves of control DNA sequences and proteins.

a. The SPRi fitting curves of unmodified and 7th N⁶-mA modified SIDD dsDNA binding to SATB1 Q390RQ402A. RU, response units.

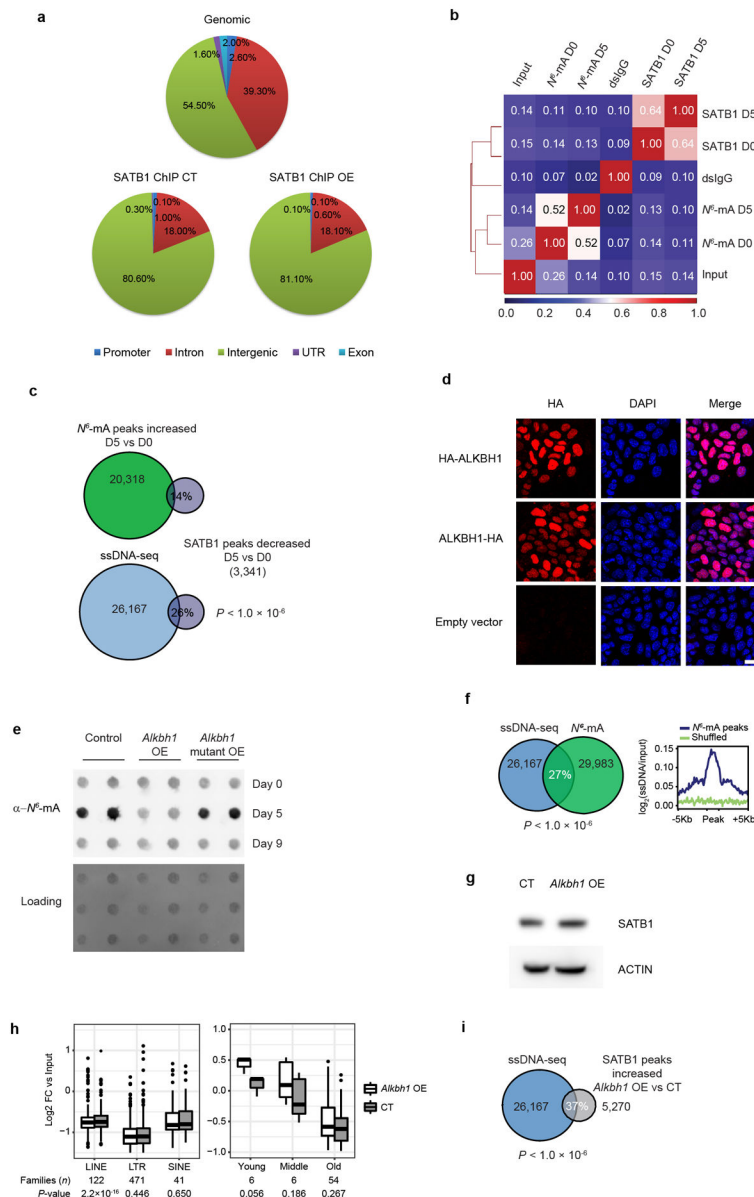
b. Summary of BLI response signals from unmodified and 7th N⁶-mA modified SIDD dsDNA binding to wild type and mutant SATB1 proteins.

c. Original BLI binding curves of SATB1 and mutants with unmodified, and N⁶-mA modified SIDD dsDNA substrates at the 7th position respectively.

d. ITC titration curves and fitting curves of SATB1 titrated into unmodified, and N⁶-mA modified 18 bp dsDNA substrates at the 9th position respectively. N.D., not detected. The sequence of 18 bp dsDNA substrate is 5'-CTTATGGAAAGCATGCTT-3'.

e. ITC titration curves and fitting curves of SATB1 titrated into unmodified or N⁶-mA modified SIDD dsDNA substrates derived from IL-2 sequences. The sequence is 5'-AGATGAAAAGAATAAATGTTTAGATTTGTTGATTAAA-3', with the modification at the 11th position.

f. ITC titration curves and fitting curves of ARID3a titrated into different unmodified (un) and N^6 -mA modified dsDNA substrates as indicated.
g. ITC titration curves and fitting curves of FOXD3 and FOXM1 titrated into unmodified or 7th N^6 -mA modified SIDD dsDNA sequences.
 Data shown are representative of 3 independent experiments with similar results.



Extended Data Figure 3. SATB1 binding to chromatin is abolished by N^6 -mA and regulated by *Alkbh1*.

a. Annotation of SATB1 ChIP-seq peaks in vector control (CT) and *Alkbh1* overexpression (OE) cells by CEAS. The genomic distribution is shown above.
b. Spearman correlation analysis of SATB1 ChIP-seq and N^6 -mA DIP-seq reads, as well as input and IgG controls. $n = 1$ sequencing sample per condition.

c. Venn diagram showing the percentage of SATB1 ChIP-seq differentially decreased (day 5 vs. day 0) peaks that intersect with differentially increased N^6 -mA peaks (*top*) or ssDNA-seq peaks (*bottom*). Number of peaks per dataset given in diagram; empirical P value computed versus genome random (see Methods).

d. Nuclear localization of ALKBH1 in *iCdx2* ES cells with transient overexpression of HA-tagged ALKBH1 and staining with anti-HA antibody. HA-ALKBH1, N-terminus HA tag; ALKBH1-HA, C-terminus HA tag. Scale bar = 25 μ m. The experiment was repeated independently 3 times with similar results.

e. DNA dot blotting of N^6 -mA in empty vector control, *Alkbh1*-overexpressing (OE), or *Alkbh1* mutant OE cell lines at three time points of fate transition. The experiments were repeated independently 3 times with similar results.

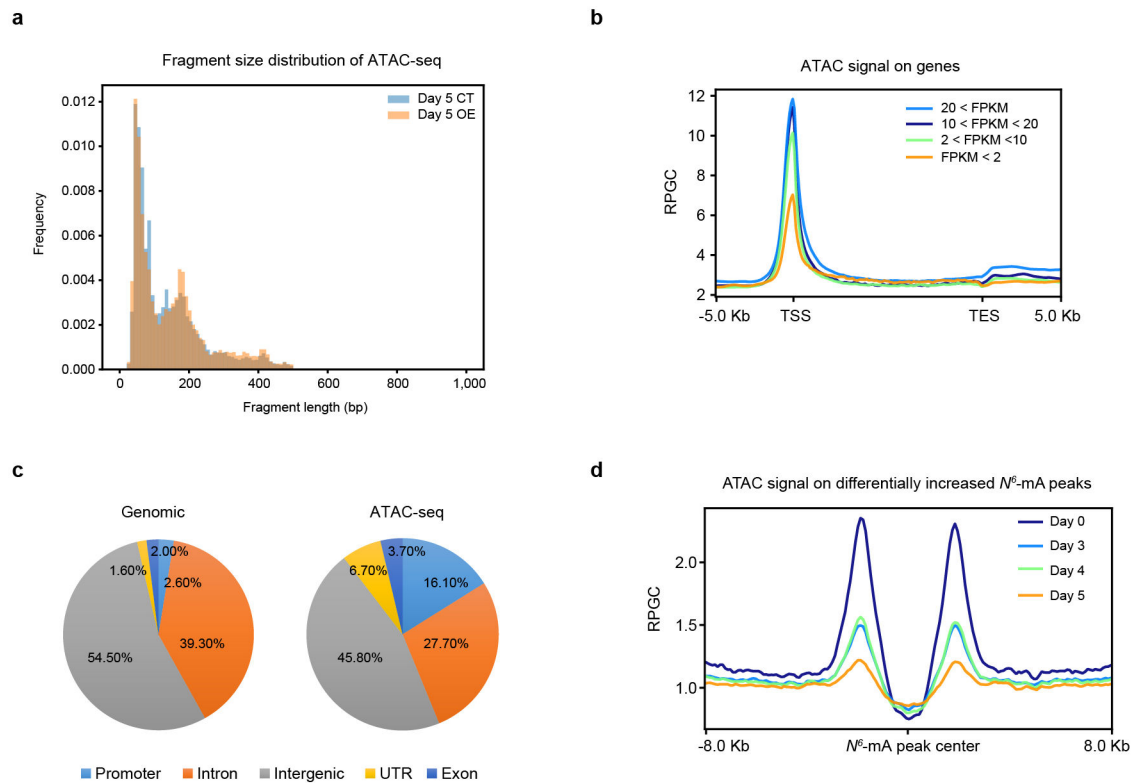
f. *Left:* Venn diagram showing the percentage of N^6 -mA DIP-seq differentially decreased *Alkbh1* OE vs CT peaks that intersect with ssDNA-seq peaks within a 5 kb window. Number of peaks per dataset given in diagram; empirical P value computed versus genome random. *Right:* Average signal of ssDNA-seq over N^6 -mA DIP-seq differentially decreased (*Alkbh1* OE vs CT) peaks.

g. Western blot showing that SATB1 protein levels are similar in control and *Alkbh1* OE cells at day 5. The experiment was repeated independently 3 times with similar results.

h. *Top:* Log₂ fold-change (FC) enrichment of Satb1 ChIP-seq over input reads for three major classes of transposable elements (TE). Day 5 reads are significantly enriched for LINE-1 over day 0. *Bottom:* Log₂ fold-change (FC) enrichment of Satb1 ChIP-seq over input reads for different LINE1 elements grouped by evolutionary age (young: < 1.5 Myr; middle: > 1.5 Myr < 6 Myr; old < Myr). Number of families plotted per TE class or category and P values from two-sided paired t -tests are provided. Raw data are in Supplementary Table 1. Boxes represent the interquartile range (IQR), horizontal line the median, and whiskers the minimum and maximum values; whiskers extend up to $1.5 \times$ IQR when outliers that fall outside that range are present.

i. Venn diagram showing the percentage of SATB1 differentially increased *Alkbh1* OE vs CT peaks that intersect with ssDNA-seq peaks within a 5 kb window. Number of peaks per dataset given in diagram; empirical P value computed versus genome random.

For blot source data, see Supplementary Figure 1.



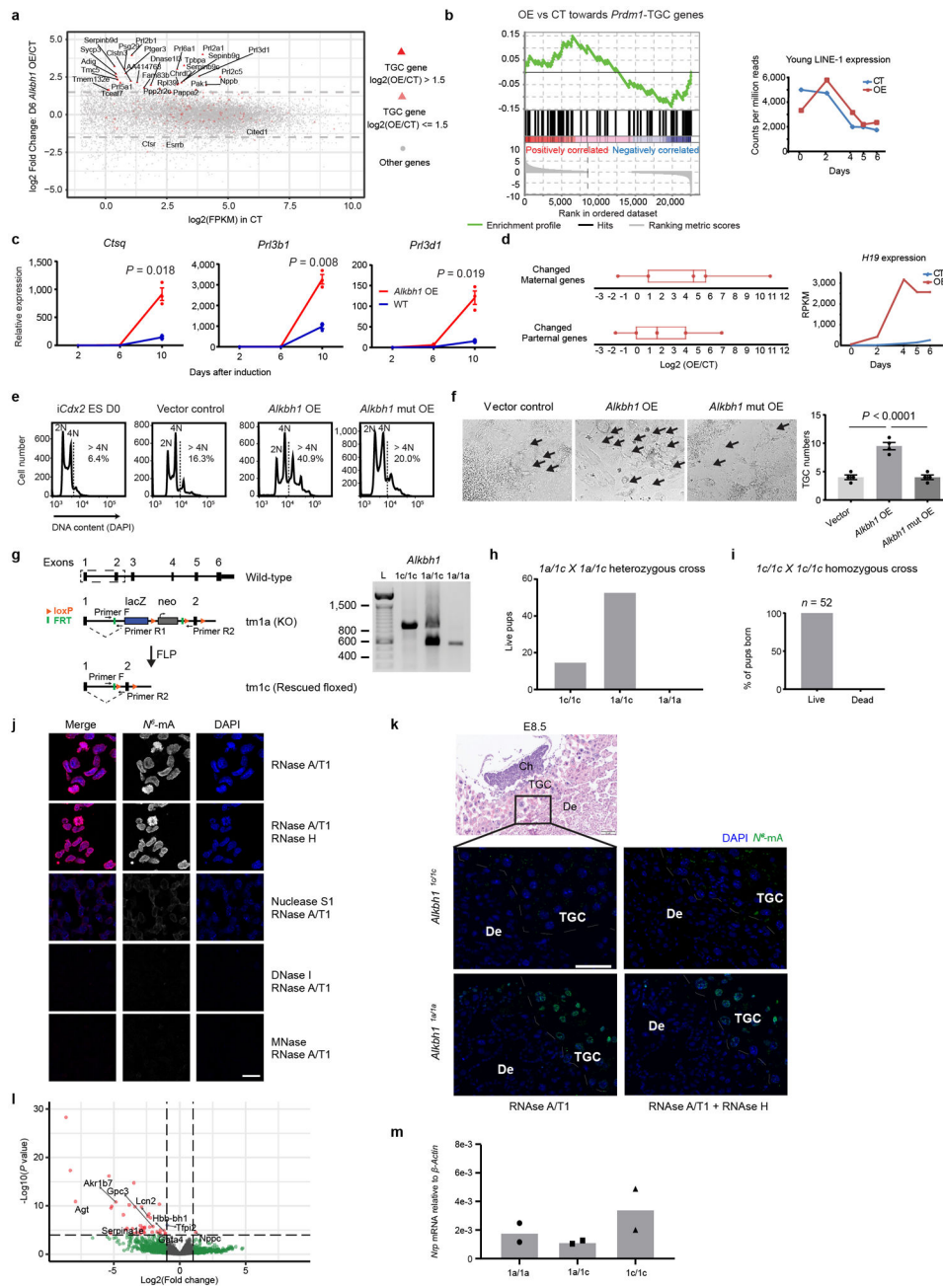
Extended Data Figure 4. ATAC-seq signal is diminished at N^6 -mA DIP-seq peaks.

a. Fragment size distribution of ATAC-seq inserts, which culminates around 100 bp, is unchanged in vector control (CT) and *Alkbh1*-overexpressing (OE) cells at day 5.

b. ATAC-seq signal in control cells at day 5 is enriched at transcription start sites (TSS) and increases corresponding to gene expression level, which is consistent with previous reports.

c. The distribution of given genomic features compared to the distribution of annotated ATAC-seq peaks in day 5 CT cells calculated by CEAS.

d. ATAC-seq signal in CT cells at indicated time points centered at N^6 -mA DIP-seq differentially increased peaks (day 5 vs day 0). The signal diminishes at regions that gain N^6 -mA signals (see Fig. 1b) during cell fate transition.



Extended Data Figure 5: *Alkbh1* overexpression promotes TGC formation and *Alkbh1* knockout leads to loss of TGC and embryonic lethality.

a. RNA-seq analysis of *Alkbh1*-overexpressing (OE) vs vector control (CT) cells in the *iCdx2* system. Genes involved in trophoblast giant cell (TGC) differentiation are enriched in differentially upregulated genes. Dotted grey lines mark a log₂ fold-change of ± 1.5. FPKM, fragments per kilobase of transcript per million mapped reads.

b. Left: Gene set enrichment analysis of 79 *Prdm1*+ trophoblast giant cell (*Prdm1*-TGC) genes shows *Alkbh1* OE does not upregulate genes specific to *Prdm1*-TGCs ($P = 0.1094$). Enrichment score reflects the degree to which TGC genes are overrepresented at the top or bottom of a ranked list of genes from RNA-seq data, and significance is determined

empirically (see Methods). *Right*: Young LINE-1 expression from RNA-seq data is down-regulated during cell fate transition. *Alkbh1* OE modestly affects expression.

c. RT-qPCR analysis of TGC-specific marker gene expression at indicated time points in *Alkbh1*-overexpressing and control cells. Holm-Sidak method for multiple comparisons with all timepoints and genes; adjusted *P* values that reached significance are provided. Mean \pm s.e.m. of 3 biological replicates.

d. *Left*: Maternal expressed genes (Megs) appear to be more severely affected than the paternal expressed genes (Pegs) in *Alkbh1* OE versus control cells. *Right*: *H19* expression is upregulated during cell fate transition and by *Alkbh1* OE. Data from RNA-seq.

e. Histograms of cellular DNA content (DAPI) by flow cytometry analysis of i*Cdx2*ES at day 0 and empty vector control, *Alkbh1* OE, or *Alkbh1* mutant (mut) OE cells at day 10 of cell fate transition. Dotted lines mark threshold used for polyploid (> 4N) cells. The experiments were repeated independently 2 times with similar results.

f. More TGC-LCs formed in *Alkbh1* overexpressed cells than in WT cells. Black arrows indicate the TGC-LCs. Quantification of TGC cells shown as mean \pm s.e.m. of 4 randomly chosen high-power fields. One-way ANOVA ($P < 0.0001$) followed by Tukey's multiple comparisons test. The experiments were repeated independently 2 times with similar results.

g. The strategy of *Alkbh1* knockout mice in this paper. Genotyping PCR shown at right, representative of > 5 independent experiments with similar results. For gel source data, see Supplementary Figure 1.

h. Crosses of *Alkbh1*^{1a/1c} heterozygotes generate 0 live *Alkbh1*^{1a/1a} knockout pups, indicating embryonic lethality. Significance of lethality was determined by the chi-squared test (X^2 (df = 2, $n = 68$) = 27.85; $P < 0.0001$). Data from 12 independent litters.

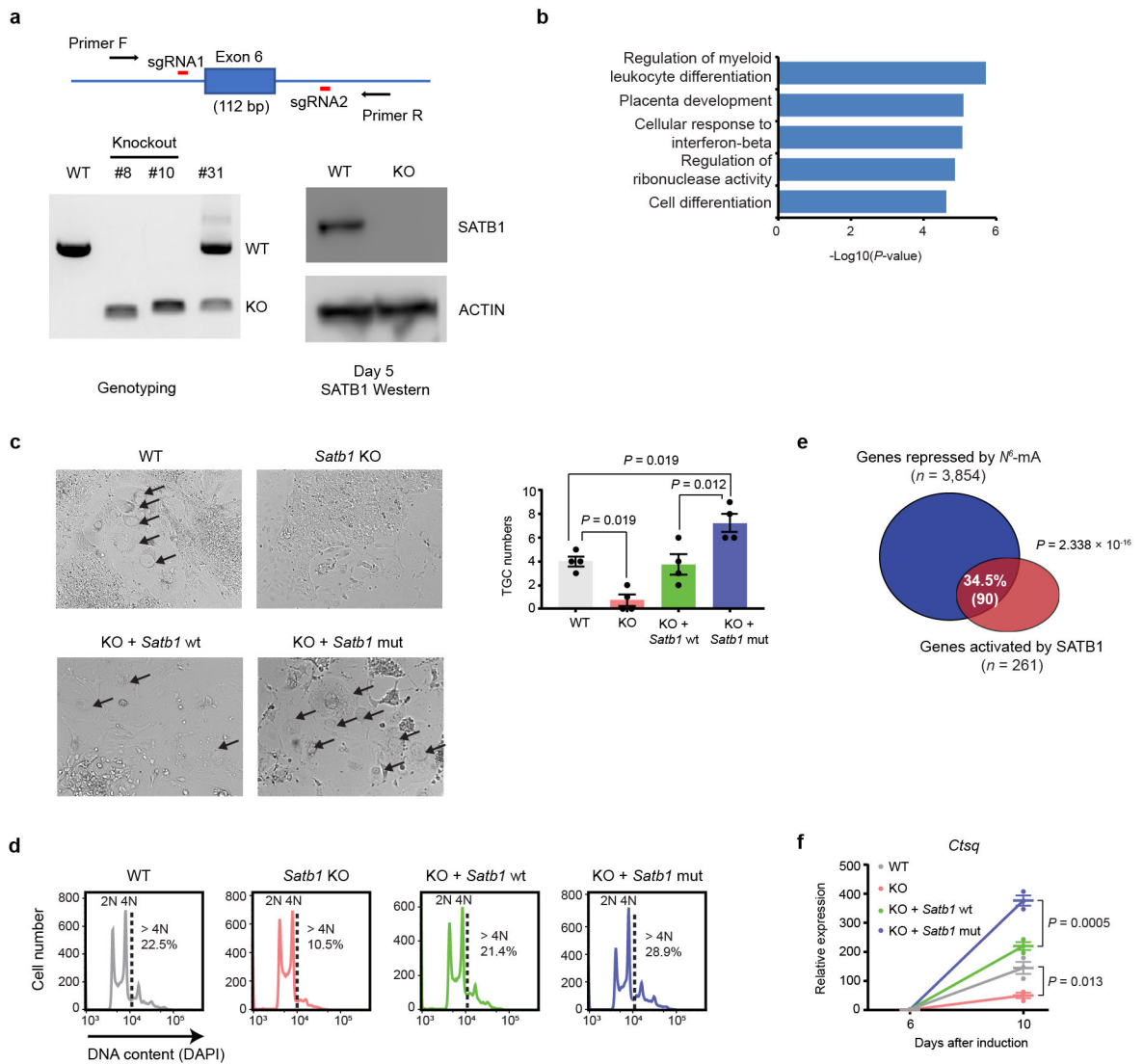
i. *Alkbh1*^{1c/1c} homozygote pups are viable and survive to adulthood. Data from 9 independent litters of *Alkbh1*^{1c/1c} x *Alkbh1*^{1c/1c} crosses.

j. *N*⁶-mA staining of i*Cdx2* cells with indicated nuclease treatment. Scale bar = 15 μ m. The experiments were repeated independently 2 times with similar results.

k. E8.5 *Alkbh1*^{1a/1a} embryos display observable levels of *N*⁶-mA in their trophoblast giant cells; no observable *N*⁶-mA is seen in the trophoblast giant cells of *Alkbh1*^{1c/1c} embryos. *N*⁶-mA staining in placental tissues is not depleted by RNase H treatment. Images show merged DAPI (blue) and anti-*N*⁶-mA (green). Images representative of $n = 2$ independent biological replicates stained through 4 planes. H&E scale bar = 50 μ m, immunofluorescence scale bar = 100 μ m. TGC: Trophoblast Giant Cells, De: Decidua, Ch: Chorion.

l. mRNA-seq of the maternal-fetal interface of E10.5 embryos demonstrates significant gene downregulation of genes involved in renin-angiotensin homeostasis, clotting pathways, syncytiotrophoblast differentiation, and innate immunity. Two-sided binomial tests were performed by DESeq2 and displayed as $-\log_{10}$ (non-adjusted *P* value) from $n = 2$ independent biological replicates per genotype. Detailed statistics for significant differentially expressed genes are available in Supplementary Table 2.

m. *Nrp*, a coding gene overlapping with exon 1 of *Alkbh1*, has low expression in the E12.0 maternal-fetal interface and is similar between *Alkbh1* genotypes. Data shown as mean of 2 biological replicates per genotype.



Extended Data Figure 6: The tolerant SATB1 mutant rescues TGC formation in *Satb1* KO cells.

a. Top: Schematic of the CRISPR–Cas9 approach used to knockout *Satb1*. Exon 6 is deleted in KO cells. **Bottom left:** PCR genotyping indicating the homozygosity of the knockout alleles in two clones. **Bottom right:** Western blotting did not detect any SATB1 protein in the KO cells. Western blots were repeated independently 3 times with similar results. For gel and blot source data, see Supplementary Figure 1.

b. Gene ontology enrichment analysis of down-regulated genes ($n = 399$) in *Satb1* KO cells. Two-sided binomial tests with Bonferroni correction were performed using PANTHER and significant results are displayed as $-\log_{10}(P\text{-value})$.

c. Fewer TGC-LCs were developed in *Satb1* KO cells than in controls. Reconstitution with *N*⁶-mA tolerant *Satb1* mutant (Q390RQ402A) generates more TGC-LCs than does the WT reconstitution. Quantification of TGC cells (*right*) shown as mean \pm s.e.m. of 4 randomly chosen high-power fields. One-way ANOVA ($P < 0.0001$) followed by Tukey's multiple comparisons test. The experiments were repeated independently 2 times with similar results.

- d.** Histograms of cellular DNA content (DAPI) by flow cytometry analysis of rescue experiment at day 10. Dotted line marks threshold used for polyploid (> 4N) cells. The experiments were repeated independently 2 times with similar results.
- e.** The overlap of genes upregulated by *Alkbh1* overexpression and genes down-regulated in *Satb1* KO at day 0. Significance of gene set intersections was determined by the chi-squared test (χ^2 (df = 1, $n = 24,319$ genes) = 67.30).
- f.** RT-qPCR showing that *Ctsq* expression, a TGC-specific marker, is more efficiently rescued by the *Satb1* N⁶-mA tolerant mutant. One-way ANOVA ($P < 0.0001$) followed by Tukey's multiple comparisons test on day 10 data. Mean \pm s.e.m. of 3 biological replicates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank the members of all laboratories for their suggestions and feedback for this work. We thank our funding sources: A.Z.X.: is supported by the Ludwig Family Foundation and R01GM114205; R.V.N.: NIH Medical Scientist Training Program Training Grant T32GM007205; M.H.A.: NSF GRFP (DGE1752134) and the Gruber Foundation Science Fellowship; Y.W.: R35 ES031707. Z.J.: USDA-NIFA: (2019-67016-29863), the Audubon Center for Research of Endangered Species and USDA-NIFA W4171. H.L.: the National Natural Science Foundation of China (91753203 and 31725014) and the National Key R&D Program of China (2016YFA0500700).

References

1. Wu TP et al. DNA methylation on N 6 -adenine in mammalian embryonic stem cells. *Nature* 532, 329–333 (2016). [PubMed: 27027282]
2. Dickinson LA, Joh T, Kohwi Y & Kohwi-Shigematsu T A tissue-specific MAR SAR DNA-binding protein with unusual binding site recognition. *Cell* 70, 631–645 (1992). [PubMed: 1505028]
3. Kohwi-Shigematsu T & Kohwi Y Torsional stress stabilizes extended base unpairing in suppressor sites flanking immunoglobulin heavy chain enhancer. *Biochemistry* 29, 9551–60 (1990). [PubMed: 2125477]
4. Bode J et al. Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science* 255, 195–197 (1992). [PubMed: 1553545]
5. Benham C, Kohwi-Shigematsu T & Bode J Stress-induced duplex DNA destabilization in Scaffold/Matrix attachment regions. *J Mol Biol* 274, 181–196 (1997). [PubMed: 9398526]
6. Bode J et al. Correlations between Scaffold/Matrix Attachment Region (S/MAR) Binding Activity and DNA Duplex Destabilization Energy. *J Mol Biol* 358, 597–613 (2006). [PubMed: 16516920]
7. Zhu S et al. Mapping and characterizing N6-methyladenine in eukaryotic genomes using single-molecule real-time sequencing. *Genome Res* 28, 1067–1078 (2018). [PubMed: 29764913]
8. Yao B et al. DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. *Nat Commun* 8, 1122 (2017). [PubMed: 29066820]
9. Xie Q et al. N6-methyladenine DNA Modification in Glioblastoma. *Cell* 175, 1228–1243.e20 (2018). [PubMed: 30392959]
10. Kouzine F et al. Global Regulation of Promoter Melting in Naive Lymphocytes. *Cell* 153, 988–999 (2013). [PubMed: 23706737]
11. Kouzine F et al. Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. *Cell Syst* 4, 344–356.e7 (2017). [PubMed: 28237796]
12. Yasui D, Miyano M, Cai S, Varga-Weisz P & Kohwi-Shigematsu T SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* 419, 641–645 (2002). [PubMed: 12374985]

13. Kohwi-Shigematsu T et al. SATB1-mediated functional packaging of chromatin into loops. *Methods* 58, 243–254 (2012). [PubMed: 22782115]
14. Cai S, Han H-J & Kohwi-Shigematsu T Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat Genet* 34, 42–51 (2003). [PubMed: 12692553]
15. Weber M et al. Genomic Imprinting Controls Matrix Attachment Regions in the Igf2 Gene. *Mol Cell Biol* 23, 8953–8959 (2003). [PubMed: 14645508]
16. Alvarez JD et al. The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev* 14, 521–535 (2000). [PubMed: 10716941]
17. Fessing MY et al. P63 regulates Satb1 to control tissue-specific chromatin remodeling during development of the epidermis. *J Cell Biol* 194, 825–839 (2011). [PubMed: 21930775]
18. Asanoma K et al. SATB homeobox proteins regulate trophoblast stem cell renewal and differentiation. *J Biol Chem* 287, 2257–2268 (2012). [PubMed: 22123820]
19. Ralston A et al. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* 137, 395–403 (2010). [PubMed: 20081188]
20. Ghosh RP et al. Satb1 integrates DNA binding site geometry and torsional stress to differentially target nucleosome-dense regions. *Nat Commun* 10, 3221 (2019). [PubMed: 31324780]
21. Kitagawa Y et al. Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. *Nat Immunol* 18, 173–183 (2017). [PubMed: 27992401]
22. Savarese F et al. Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. *Genes Dev* 23, 2625–2638 (2009). [PubMed: 19933152]
23. Goolam M & Zernicka-Goetz M The chromatin modifier Satb1 regulates cell fate through Fgf signalling in the early mouse embryo. *Development* 144, 1450–1461 (2017). [PubMed: 28289135]
24. Zhang M et al. Mammalian ALKBH1 serves as an N6-mA demethylase of unpairing DNA. *Cell Res* 30, 197–210 (2020). [PubMed: 32051560]
25. Schiffers S et al. Quantitative LC-MS Provides No Evidence for m 6 dA or m 4 dC in the Genome of Mouse Embryonic Stem Cells and Tissues. *Angew Chemie Int Ed* 56, 11268–11271 (2017).
26. Pan Z et al. Impaired placental trophoblast lineage differentiation in *Alkbh1* $-/-$ mice. *Dev Dyn* 237, 316–327 (2008). [PubMed: 18163532]
27. Nishiyama A et al. Uncovering Early Response of Gene Regulatory Networks in ESCs by Systematic Induction of Transcription Factors. *Cell Stem Cell* 5, 420–433 (2009). [PubMed: 19796622]
28. Huang D et al. The role of Cdx2 as a lineage specific transcriptional repressor for pluripotent network during the first developmental cell lineage segregation. *Sci Rep* 7, 17156 (2017). [PubMed: 29214996]
29. Mirkovitch J, Mirault M-E & Laemmli UK Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell* 39, 223–232 (1984). [PubMed: 6091913]
30. Yamasaki K, Akiba T, Yamasaki T & Harata K Structural basis for recognition of the matrix attachment region of DNA by transcription factor SATB1. *Nucleic Acids Res* 35, 5073–5084 (2007). [PubMed: 17652321]
31. Yamaguchi H, Tateno M & Yamasaki K Solution structure and DNA-binding mode of the matrix attachment region-binding domain of the transcription factor SATB1 that regulates the T-cell maturation. *J Biol Chem* 281, 5319–5327 (2006). [PubMed: 16371359]
32. Zhou C, Liu Y, Li X, Zou J & Zou S DNA N6-methyladenine demethylase ALKBH1 enhances osteogenic differentiation of human MSCs. *Bone Res* 4, 16033 (2016). [PubMed: 27785372]
33. Xiao C, Le et al. N 6 -Methyladenine DNA Modification in the Human Genome. *Mol Cell* 71, 306–318.e7 (2018). [PubMed: 30017583]
34. Tian L-F et al. Structural basis of nucleic acid recognition and 6mA demethylation by human ALKBH1. *Cell Res* 30, 272–275 (2020). [PubMed: 32051559]
35. Ougland R et al. ALKBH1 is a histone H2A dioxygenase involved in neural differentiation. *Stem Cells* 30, 2672–2682 (2012). [PubMed: 22961808]
36. Nelson AC, Mould AW, Bikoff EK & Robertson EJ Single-cell RNA-seq reveals cell type-specific transcriptional signatures at the maternal–foetal interface during pregnancy. *Nat Commun* 7, 11414 (2016). [PubMed: 27108815]

37. Fujimori H et al. The H19 induction triggers trophoblast lineage commitment in mouse ES cells. *Biochem Biophys Res Commun* 436, 313–318 (2013). [PubMed: 23743205]
38. Esquiliano DR, Guo W, Liang L, Dikkes P & Lopez MF Placental Glycogen Stores are Increased in Mice with H19 Null Mutations but not in those with Insulin or IGF Type 1 Receptor Mutations. *Placenta* 30, 693–699 (2009). [PubMed: 19524295]
39. Nordstrand LM et al. Mice lacking Alkbh1 display sex-ratio distortion and unilateral eye defects. *PLoS One* 5, e13827 (2010). [PubMed: 21072209]
40. Skarnes WC et al. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* 474, 337–342 (2011). [PubMed: 21677750]
41. Khan S et al. Glypican-3 (GPC3) expression in human placenta: Localization to the differentiated syncytiotrophoblast. *Histol Histopathol* 16, 71–78 (2001). [PubMed: 11193214]
42. Cuffe JSM et al. The effects of gestational age and maternal hypoxia on the placental renin angiotensin system in the mouse. *Placenta* 35, 953–961 (2014). [PubMed: 25249154]
43. Stepan H, Faber R, Stegemann S, Schultheiss HP & Walther T Expression of C-type natriuretic peptide in human placenta and myometrium in normal pregnancies and pregnancies complicated by intrauterine growth retardation: Preliminary results. *Fetal Diagn Ther* 17, 37–41 (2002). [PubMed: 11803215]
44. Wu T et al. Histone Variant H2A.X Deposition Pattern Serves as a Functional Epigenetic Mark for Distinguishing the Developmental Potentials of iPSCs. *Cell Stem Cell* 15, 281–294 (2014). [PubMed: 25192463]
45. Pathak RU, Srinivasan A & Mishra RK Genome-wide mapping of matrix attachment regions in *Drosophila melanogaster*. *BMC Genomics* 15, 1022 (2014). [PubMed: 25424749]
46. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359 (2012). [PubMed: 22388286]
47. Zang C et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* 25, 1952–1958 (2009). [PubMed: 19505939]
48. Castro-Diaz N et al. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev* 28, 1397–1409 (2014). [PubMed: 24939876]
49. Nix DA, Courdy SJ & Boucher KM Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. *BMC Bioinformatics* 9, 523 (2008). [PubMed: 19061503]
50. Jeong H-H, Yalamanchili HK, Guo C, Shulman JM & Liu Z An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. *Pac Symp Biocomput* 23, 168–179 (2018). [PubMed: 29218879]
51. Corces MR et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* 14, 959–962 (2017). [PubMed: 28846090]
52. Zhang Y et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137 (2008). [PubMed: 18798982]
53. Subramanian A et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545–15550 (2005). [PubMed: 16199517]
54. Mi H, Muruganujan A, Ebert D, Huang X & Thomas PD PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 47, D419–D426 (2019). [PubMed: 30407594]
55. Bradley A et al. The mammalian gene function resource: the international knockout mouse consortium. *Mamm Genome* 23, 580–586 (2012). [PubMed: 22968824]

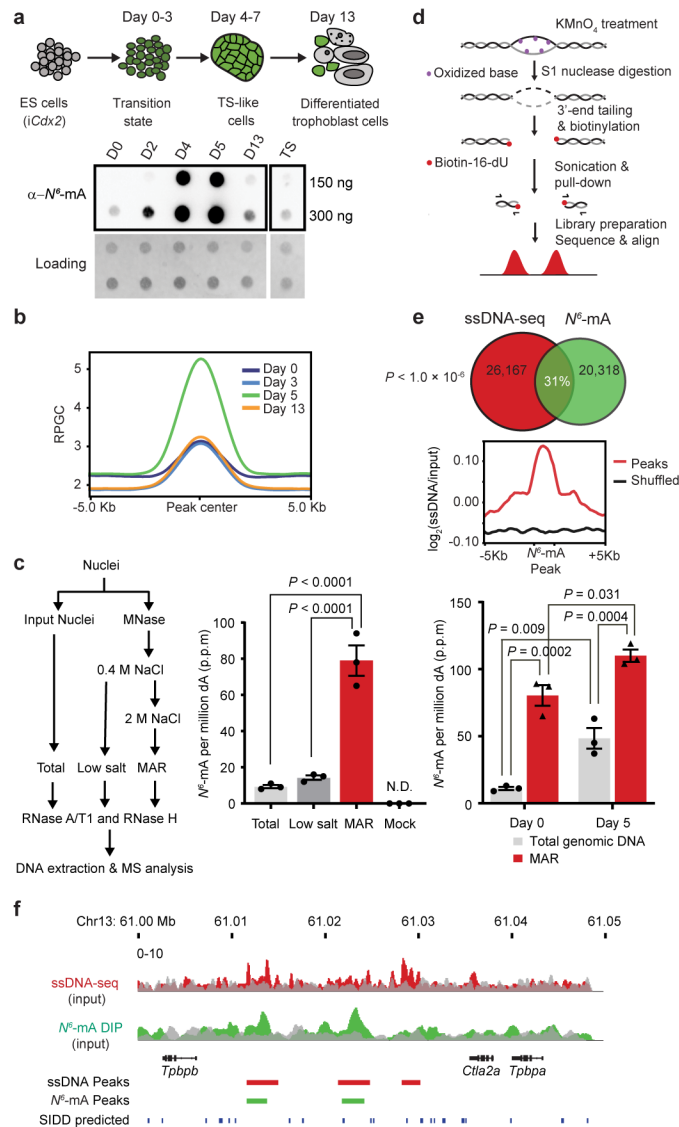


Figure 1. N⁶-mA is upregulated at SIDD regions during TSC development.

a. Top: Schematic of the iCdx2 ES-to-TS cell fate transition system. **Bottom:** DNA dot blotting of N⁶-mA at different time points of differentiation. Experiments were repeated independently three times with similar results. For blot source data, see Supplementary Figure 1.

b. Average N⁶-mA reads per genomic content (RPGC) at different time points centered at day 0 to day 5 differentially increased N⁶-mA peaks and flanking regions. Kb, kilobase.

c. Mass spectrometry (MS) analysis of N⁶-mA. **Left:** Schematic of DNA extraction from MAR. **Middle:** N⁶-mA levels from different chromatin fractions and mock (buffer, nucleases, and Proteinase K). One-way ANOVA ($P < 0.0001$) followed by Tukey's multiple comparisons test. **Right:** MAR DNA N⁶-mA compared to total DNA at differentiation time points. Two-way ANOVA (Day: $P = 0.0005$; Fraction: $P < 0.0001$) followed by Tukey's test. Mean \pm s.e.m. of three biological replicates. dA, deoxyadenosine; p.p.m, parts per million; N.D., not detected.

- d.** Schematic of ssDNA-seq protocol.
- e.** *Top:* Venn diagram showing the percentage of N^6 -mA DIP-seq differentially increased (day 5 vs. day 0) peaks that intersect with ssDNA-seq peaks. Number of peaks per dataset given in diagram; empirical P value computed versus genome random (see Methods).
Bottom: Aggregation of ssDNA-seq signal over N^6 -mA DIP differentially increased peaks or the peaks shuffled to random genomic positions.
- f.** ssDNA-seq and N^6 -mA DIP-seq signal at a representative genomic locus. Input signal is overlaid in grey. Genes, peaks, and predicted SIDD regions are as labeled.

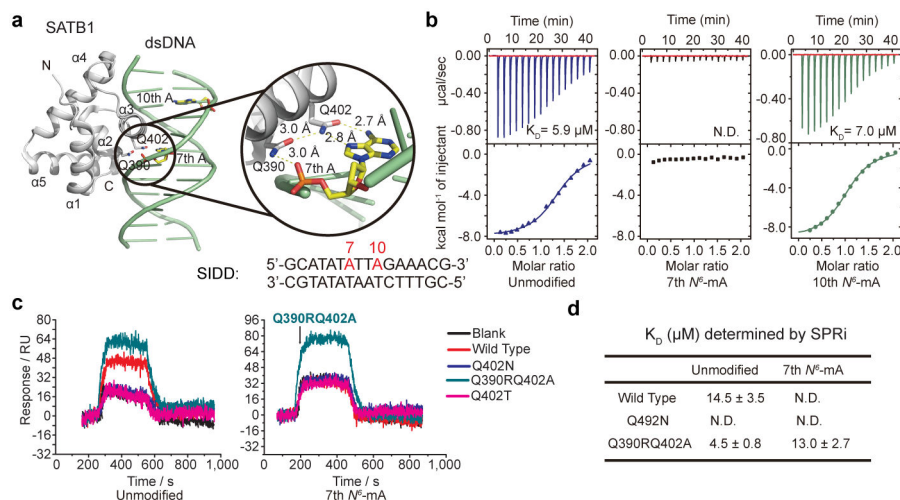


Figure 2. DNA N^6 -mA modification abolishes SATB1 binding with DNA *in vitro*.

a. The structure of SATB1 in complex with unmodified dsDNA (PDB code: 2O49)³⁰. The structure of SATB1 is shown in white and the structure of dsDNA is shown in green as cartoon. The 7th and 10th deoxyadenosines are shown in yellow as sticks.

b. ITC titration curves and fitting curves of SATB1 titrated into unmodified and N^6 -mA modified SIDD dsDNA substrates at the 7th or 10th position, respectively.

c. The SPRi response curves of unmodified and 7th N^6 -mA modified SIDD dsDNA binding to wild type and mutant SATB1 proteins. The immobilized protein concentration was 1 mM. The injected DNA concentration was 2.5 μM .

d. Summary of K_D values determined by SPRi from unmodified and 7th N^6 -mA modified SIDD dsDNA binding to wild type and mutant SATB1 proteins. Mean \pm s.d. of three biological replicates. Data shown in b-d are representative of three independent experiments with similar results. N.D., not detected.

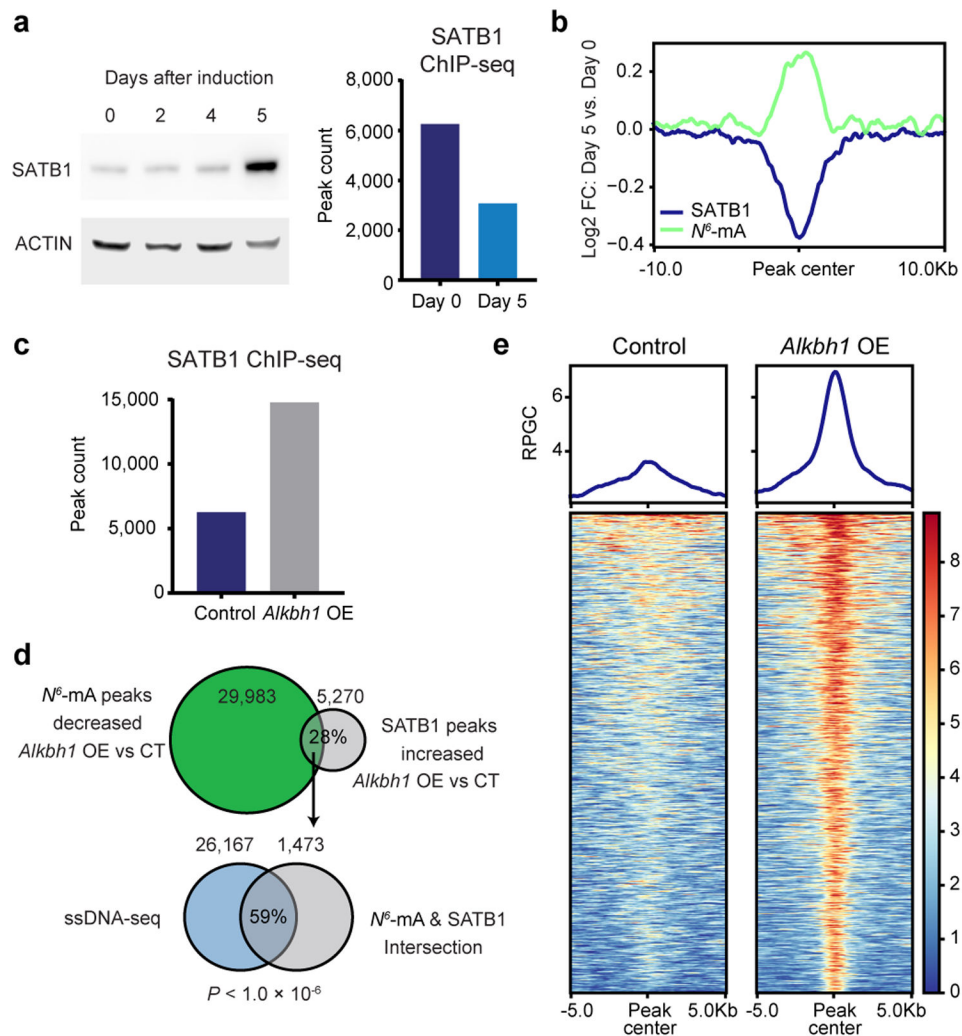


Figure 3. N⁶-mA antagonizes SATB1 binding to the chromatin during TSC development

a. SATB1 binding is inversely correlated with N⁶-mA levels. *Left:* Western blot of SATB1 expression during cell fate transition at indicated time points. The experiments were repeated independently 3 times with similar results. *Right:* Number of SATB1 ChIP-seq peaks at indicated days. For blot source data, see Supplementary Figure 1.

b. Aggregation of differential signals of N⁶-mA DIP and SATB1 ChIP centered at differentially increased N⁶-mA peaks (day 5 vs. day 0).

c. Number of SATB1 ChIP-seq peaks in control or *Alkbh1* OE cells.

d. *Top:* Venn diagram showing the intersection between N⁶-mA differentially decreased peaks and SATB1 ChIP-seq differentially increased peaks in *Alkbh1* OE vs. control (CT) cells. *Bottom:* Overlap between the intersecting regions and ssDNA-seq peaks, not to scale. Number of peaks per dataset given in diagram; empirical *P* value computed versus genome random.

e. Average signal and heat maps of SATB1 ChIP-seq around N⁶-mA peak centers in control or *Alkbh1* OE cells.

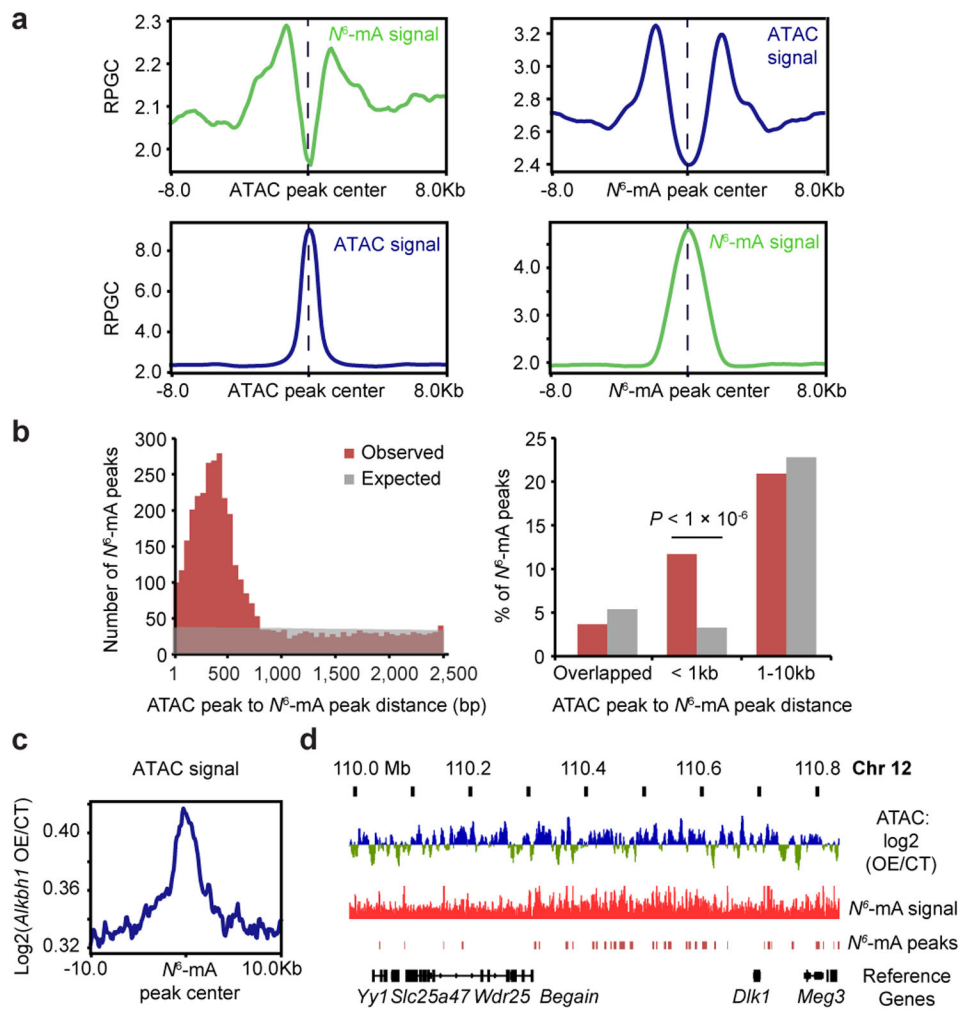


Figure 4. N^6 -mA is enriched at euchromatin boundaries, thereby preventing the spreading of euchromatin.

a. Aggregation profiles of day 5 ATAC-seq inserts at day 5 vs. day 0 differentially increased N^6 -mA peak centers (*left*) and day 5 N^6 -mA DIP-seq signal at ATAC-seq day 5 peak centers (*right*). RPGC, reads per genomic content.

b. Left: Histogram distribution of the distance between ATAC peaks and N^6 -mA peaks at day 5. **Right:** Bar chart showing the percentage of N^6 -mA peaks harboring ATAC peaks within a certain distance at day 5. Empirical P value computed versus genome random.

c. Aggregation profiles showing differential ATAC-seq signals in *Alkbh1*-overexpressing (OE) vs. control (CT) cells at N^6 -mA differentially decreased (OE vs CT) peak centers and flanking regions.

d. Differential *Alkbh1* OE vs CT ATAC-seq signal, day 5 N^6 -mA DIP-seq, and N^6 -mA peaks at the centromeric end of the *Dlk1-Dio3* imprinting locus.

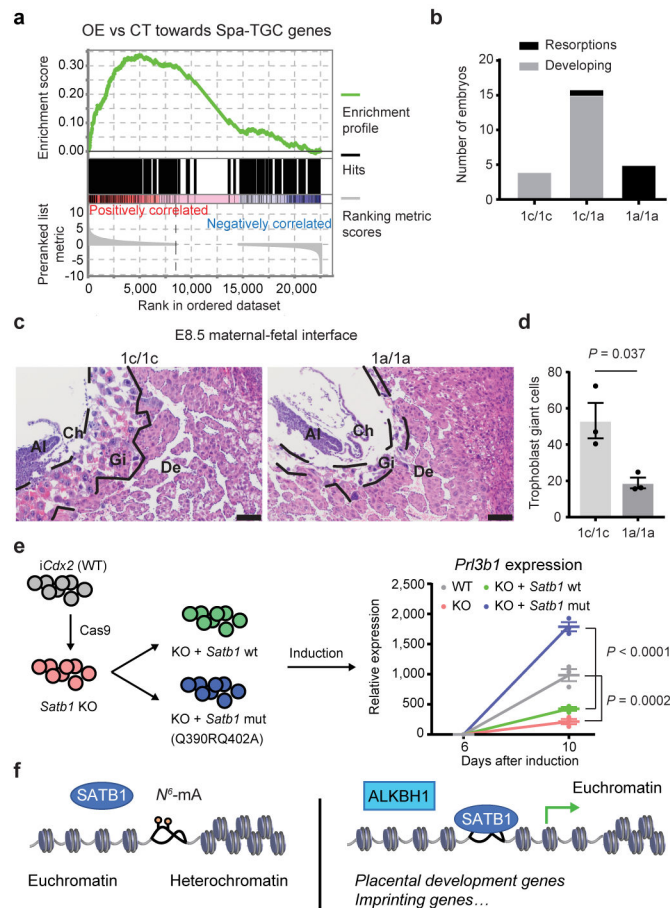


Figure 5. N^6 -mA and SATB1 mediated epigenetic silencing pathway plays critical roles in TE development.

- a.** Gene set enrichment analysis of 273 spiral artery trophoblast giant cell (Spa-TGC) genes reveals significant enrichment ($P < 0.001$) in *Alkbh1* overexpression (OE) cells compared to control (CT). Enrichment score reflects the degree to which Spa-TGC genes are overrepresented at the top or bottom of a ranked list of genes from RNA-seq data, and significance is determined empirically (see Methods).
- b.** Numbers of developing embryos or resorptions from heterozygous *Alkbh1*^{1a/1c} crosses. $n = 25$ total embryos from 4 independent matings.
- c.** H&E staining of the maternal fetal interface shows a great diminishment of TGCs in *Alkbh1*^{1a/1a} placentas. E8.5 *Alkbh1*^{1a/1a} murine placental tissues are depleted of secondary giant cells (Gi), which normally reside between the decidua (De) layer and the chorio-allantoic plate (Ch, Al), as seen in the *Alkbh1*^{1c/1c}. Representative images of 3 litter-matched embryos per genotype from 3 dams. Scale bar = 100 μ m.
- d.** Multi-plane quantification shows a depletion of trophoblast giant cells in *Alkbh1*^{1a/1a} placentas (18.8) compared to *Alkbh1*^{1c/1c} controls (53.3). Two-sided paired t test ($t = 5.06$, $df = 2$). Mean \pm s.e.m. of 5 planes through the maternal fetal interface of 3 litter-matched animals per genotype, from 3 dams.
- e.** *Left:* Schematic of *Satb1* knockout (KO) rescue experiment in *iCdx2* cells using wild type (wt) or N^6 -mA tolerant mutant (mut) *Satb1* overexpression. *Right:* RT-qPCR showing

that the expression of a TGC specific marker gene is more efficiently rescued by the *Satb1* N⁶-mA tolerant mutant. One-way ANOVA ($P < 0.0001$) followed by Tukey's multiple comparisons test on day 10 data. Mean \pm s.e.m. of 3 biological replicates.

f. Proposed model of the function of N⁶-mA, ALKBH1 and SATB1 at euchromatin boundaries.