Article

N^6-methyladenine in DNA antagonizes SATB1 in early development

https://doi.org/10.1038/s41586-020-2500-9

Received: 13 January 2019

Accepted: 4 May 2020

Published online: 15 July 2020

Check for updates

Zheng Li[1,8], Shuai Zhao[2,8], Raman V. Nelakanti[1,8], Kaixuan Lin[1,8], Tao P. Wu[1,7,8], Myles H. Alderman III[1], Cheng Guo[3,4], Pengcheng Wang[3], Min Zhang[2], Wang Min[5], Zongliang Jiang[6], Yinsheng Wang[3], Haitao Li[2]✉ & Andrew Z. Xiao[1]✉

The recent discovery of N^6-methyladenine (N^6-mA) in mammalian genomes suggests that it may serve as an epigenetic regulatory mechanism[1]. However, the biological role of N^6-mA and the molecular pathways that exert its function remain unclear. Here we show that N^6-mA has a key role in changing the epigenetic landscape during cell fate transitions in early development. We found that N^6-mA is upregulated during the development of mouse trophoblast stem cells, specifically at regions of stress-induced DNA double helix destabilization (SIDD)[2–4]. Regions of SIDD are conducive to topological stress-induced unpairing of the double helix and have critical roles in organizing large-scale chromatin structures[3,5,6]. We show that the presence of N^6-mA reduces the in vitro interactions by more than 500-fold between SIDD and SATB1, a crucial chromatin organizer that interacts with SIDD regions. Deposition of N^6-mA also antagonizes SATB1 function in vivo by preventing its binding to chromatin. Concordantly, N^6-mA functions at the boundaries between euchromatin and heterochromatin to restrict the spread of euchromatin. Repression of SIDD–SATB1 interactions mediated by N^6-mA is essential for gene regulation during trophoblast development in cell culture models and in vivo. Overall, our findings demonstrate an unexpected molecular mechanism for N^6-mA function via SATB1, and reveal connections between DNA modification, DNA secondary structures and large chromatin domains in early embryonic development.

Several recent studies have implicated N^6-mA in epigenetic silencing, especially at long interspersed element 1 (LINE-1) endogenous transposons in mouse embryonic stem (mES) cells, in brain tissue under environmental stress, and in human lymphoblastoid cells 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 gene silencing have thus far remained unclear.

N^6-mA is enriched at AT-rich regions in mammalian genomes[1], especially in areas that interact with the nuclear architecture, such as matrix/scaffold attachment regions (M/SARs). M/SARs are liable to undergo DNA unpairing under torsional stress (known as SIDD or as base unpairing regions (BURs))[2–4] during replication, transcription, and recombination[3,5,6]. A study that used genome-wide permanganate and S1 endonuclease mapping (single-stranded (ss)DNA-seq[10]) demonstrated a strong correlation between SIDD regions identified experimentally and those predicted by computational approaches[11]. SIDD regions in M/SARs are crucial for organizing large chromatin domains, such as establishing and maintaining heterochromatin–euchromatin boundaries[12–14] and facilitating long-range interactions[15].

SATB1, a well-known SIDD-regulating protein, is mainly expressed in developing T cells[16], epidermis[17], and trophoblast stem cells (TSCs)[18,19]. SATB1 binds to SIDD and then stabilizes the DNA double helix[2,20], thereby establishing and maintaining large-scale euchromatin and heterochromatin domains[12–14]. For example, SATB1 binds directly to key enhancers in pro- and pre-T cells, thereby activating gene expression in these cells while repressing the mature T cell fate[21]. During early embryogenesis, SATB1 and SATB2 form an intricate network that regulates ES cell differentiation[22]. In addition, there is emerging evidence that *Satb1* is markedly upregulated during extraembryonic tissue development and promotes extraembryonic fates such as trophectoderm and primitive endoderm lineages[18,19,23].

In this work, we show that N^6-mA contributes to TSC development and differentiation by antagonizing SATB1 function at SIDD. Moreover, ALKBH1, the DNA demethylase of N^6-mA, highly prefers SIDD sequences as substrates[24], which further strengthens the connection between N^6-mA and DNA secondary structures during cell fate transitions.

[1]Department of Genetics and Yale Stem Cell Center, Yale School of Medicine, New Haven, CT, USA. [2]MOE 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, China. [3]Department of Chemistry, University of California, Riverside, CA, USA. [4]Cancer Institute, Key Laboratory of Cancer Prevention and Intervention, China National Ministry of Education, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. [5]Interdepartmental Program in Vascular Biology and Therapeutics, Department of Pathology, Yale University School of Medicine, New Haven, CT, USA. [6]School of Animal Sciences, AgCenter, Louisiana State University, Baton Rouge, LA, USA. [7]Present address: Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. [8]These authors contributed equally: Zheng Li, Shuai Zhao, Raman V. Nelakanti, Kaixuan Lin, Tao P. Wu. ✉e-mail: lht@tsinghua.edu.cn; andrew.xiao@yale.edu

N^6-mA at SIDD during TSC development

As N^6-mA is present at low levels (6–7 parts per million (ppm)) in mES cells[1], we searched for conditions under which N^6-mA levels were upregulated. Notably, 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 (cultures with ERK and GSK3b inhibitors, which are 4N negative) but retained under some alternative 2i conditions (4N positive) (Extended Data Fig. 1a), which may explain discrepancies in the literature[25].

As previous studies showed that N^6-mA demethylase *Alkbh1*-deficient mice developed trophectodermal defects[26], we next investigated the role of N^6-mA in the development and differentiation of TSCs. We leveraged a *Cdx2*-inducible expression system[27] (i*Cdx2*) to model TSC development in cell culture, which manifests a well-synchronized and efficient cell fate transition process (Fig. 1a). Consistent with previous work[28], our RNA sequencing (RNA-seq) analysis showed that i*Cdx2* cells go through first a transition state, then a TSC-like cell (TSC-LC) state, before undergoing trophectoderm lineage differentiation (Extended Data Fig. 1b). The N^6-mA level increased transiently at a time window that coincides with the emergence of TSC-LCs, and then tapered off during trophectoderm lineage differentiation (Fig. 1a). Accordingly, the expression of *Alkbh1* inversely correlated with N^6-mA levels during this cell fate transition (Extended Data Fig. 1b).

We then used DNA immunoprecipitation followed by high-throughput sequencing (DIP-seq) to investigate N^6-mA distribution at different stages during cell fate transition. To ensure specificity towards N^6-mA in DNA, samples were first treated with extensive RNase digestion (see Methods). This protocol did not pull down any significant signal from RNA–DNA hybrids ($P = 0.999$) above background (unmodified DNA controls) (Extended Data Fig. 1c). Most of the sequencing reads were mapped (95.55%, with unique mapping) to the mouse 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; see 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, Supplementary Table 1). Notably, an established algorithm[11] predicted that 60–66% of N^6-mA peaks are SIDD (Extended Data Fig. 1i). To confirm these results, we extracted M/SAR DNA using a well-established protocol[29] and quantified N^6-mA levels by mass spectrometry, which can distinguish between methylated and unmethylated dA as well as RNA N^6-methyladenosine (m6A, not detected in the samples) (Fig. 1c, Extended Data Fig. 1j). This approach confirmed a tenfold 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 mES cells (TT2 wild-type), ssDNA-seq reads were enriched at SINE elements, and more signal was observed at the gene bodies and upstream of promoters of expressed versus non-expressed genes, similar to a previous study in B cells[11] (Extended Data Fig. 1k, l). Notably, ssDNA-seq peaks in i*Cdx2* 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 overlapped with ssDNA regions (Fig. 1e). Similar overlap was also observed when all sequencing reads, including unique and non-unique, were included in the analysis[24]. A representative trophoblast gene locus showed N^6-mA directly overlapping with ssDNA peaks (Fig. 1f). Collectively, our genomic and biochemical data demonstrate that N^6-mA is enriched at SIDD regions during the transition from ES cells to TSCs.

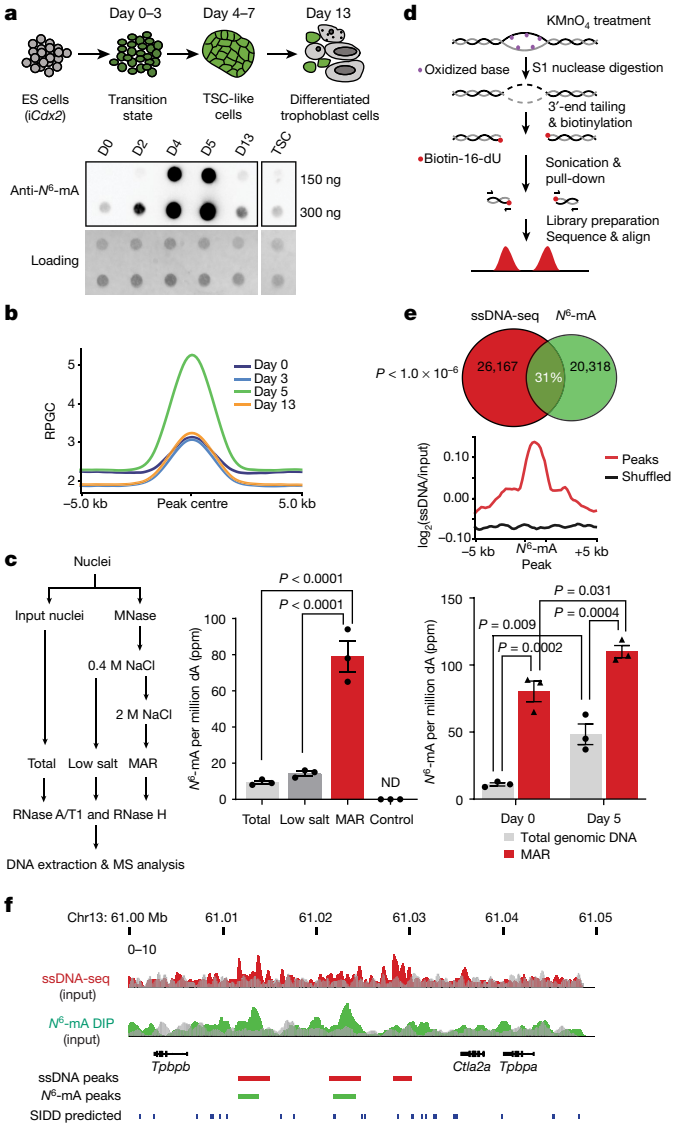
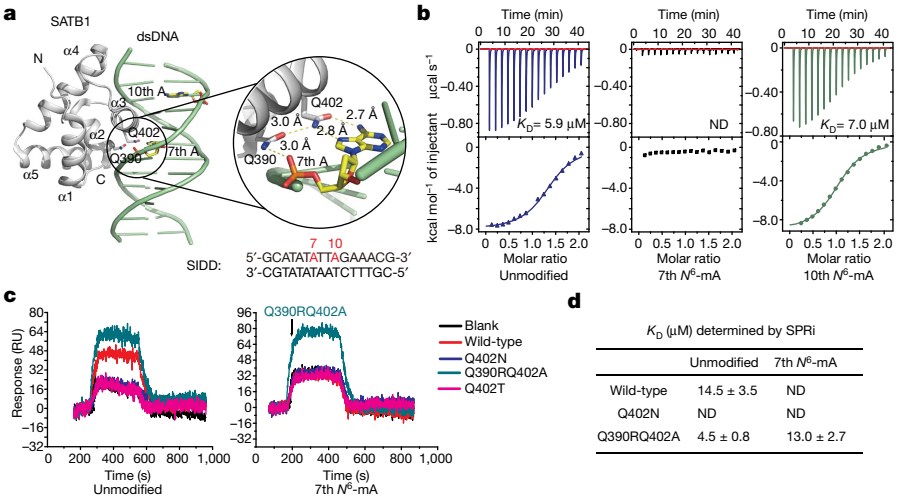


Fig. 1 | N^6-mA is upregulated at SIDD regions during TSC development. **a**, Top, schematic of the i*Cdx2* ES cell-to-TSC fate transition system. Bottom, DNA dot blotting of N^6-mA at different time points during differentiation. D, day. Experiments were repeated independently three times with similar results. For blot source data, see Supplementary Fig. 1. **b**, Average N^6-mA reads per genomic content (RPGC) at different time points, centred at differentially increased N^6-mA peaks (day 5 versus day 0). **c**, Mass spectrometry (MS) analysis of N^6-mA. Left, schematic of matrix attachment region (MAR) DNA extraction. Middle, N^6-mA levels from different chromatin fractions and mock control (buffer, nucleases, and proteinase K). One-way ANOVA ($P < 0.0001$) followed by Tukey's multiple comparisons test. Right, MAR DNA N^6-mA compared to total DNA at time points during differentiation. Two-way ANOVA (day, $P = 0.0005$; fraction, $P < 0.0001$) followed by Tukey's test. Mean ± s.e.m. of three biological replicates. dA, deoxyadenosine; ND, not detected. **d**, Schematic of ssDNA-seq protocol. **e**, Top, Venn diagram showing the percentage of N^6-mA DIP-seq differentially increased peaks (day 5 versus day 0) that intersect with ssDNA-seq peaks. Number of peaks per dataset shown; empirical *P* value computed against genome random (see Methods). Bottom, aggregation of ssDNA-seq signal over N^6-mA DIP-seq differentially increased peaks or the peaks shuffled to random genomic positions. **f**, ssDNA-seq and N^6-mA DIP-seq signals at a representative genomic locus. Input signal is overlaid in grey. Genes, peaks, and predicted SIDD regions are as labelled.



	Unmodified	7th N^6-mA
Wild-type	14.5 ± 3.5	ND
Q402N	ND	ND
Q390RQ402A	4.5 ± 0.8	13.0 ± 2.7

Fig. 2 | DNA N^6-mA modification abolishes binding of SATB1 to 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 seventh and tenth deoxyadenosines are shown in yellow as sticks. **b**, ITC titration curves and fitting curves of SATB1 titrated into unmodified SIDD dsDNA substrates and SIDD dsDNA substrates modified with N^6-mA at the seventh or tenth position. **c**, The SPRi response curves of unmodified and 7th N^6-mA-modified SIDD dsDNA binding to wild-type and mutant SATB1 proteins. Immobilized protein concentration, 1 mM; injected DNA concentration, 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. Mean ± s.d. of three biological replicates. Data shown in **b**–**d** are representative of three independent experiments with similar results.

N^6-mA abolishes SATB1–DNA binding 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 to N^6-mA-modified oligodeoxynucleotides or unmodified controls. The structure of the CUT1 domain, a DNA-binding domain of SATB1[20,30], shows that the α3-helix of SATB1 inserted into the major groove of double-stranded DNA (dsDNA)[30]. Q390 and Q402 formed hydrogen bonds with the seventh deoxyadenosine of the substrates, while the tenth deoxyadenosine (referred to as 7th N^6-mA or 10th N^6-mA, respectively) is located outside the binding pocket[31] (Fig. 2a). We quantitatively determined the binding affinities between SATB1 and SIDD sequences with N^6-mA-modified oligos at the seventh and tenth positions using isothermal titration calorimetry (ITC; Fig. 2b). Of note, the binding affinities were determined using 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 reports[31] (K_D = 5.9 μM). Notably, this interaction was abolished by the 7th N^6-mA (K_D undetectable), whereas the 10th N^6-mA had no obvious effects (K_D = 7.0 μM).

We next sought to engineer SATB1 mutants that could tolerate the 7th N^6-mA, which would be helpful for elucidating the function of N^6-mA in vivo. The surface plasmon resonance imaging (SPRi) results revealed that the Q390RQ402A mutant bound similarly to both unmodified and 7th N^6-mA SIDD sequences, whereas the Q402N and Q402T mutants did not bind either sequence, presumably owing to the loss of coordinated interactions (Fig. 2c, d, Extended Data Fig. 2a). Biolayer interferometry (BLI) confirmed these results (Extended Data Fig. 2b, c).

In addition, the DNA-binding properties of SATB1 are also influenced by sequence specificity (Extended Data Fig. 2d, e). A few other known AT-rich DNA-binding proteins tested did not display N^6-mA specificity. ARID3A recognized neither the unmodified nor the 7th N^6-mA substrates, whereas FOXM1 and FOXD3 bound 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 effect of N^6-mA on interactions between SATB1 and chromatin. Consistent with previous results[18,19], we found marked upregulation of SATB1 during the cell fate transition from ES cells to a TSC-LC fate, especially at maximal N^6-mA levels. Paradoxically, despite this global increase, SATB1 chromatin immunoprecipitation followed by high-throughput sequencing (ChIP–seq) demonstrated a 50% reduction in SATB1 peaks at days 5–6, when N^6-mA levels are highest (Fig. 3a).

Further analysis revealed that binding of SATB1 to chromatin is inversely correlated with N^6-mA deposition, and that SATB1 sites are excluded from N^6-mA sites (Extended Data Fig. 3a, b). The differentially decreased SATB1 peaks (day 5 versus day 0) overlapped significantly ($P < 1 \times 10^{-6}$) with both the gained N^6-mA sites at day 5 and the ssDNA-seq peaks (Extended Data Fig. 3c). Concordantly, aggregation profiling also showed that SATB1 ChIP–seq signals decreased at peak centres where N^6-mA signals increased during the cell fate transition (day 5 versus day 0; Fig. 3b).

We further interrogated N^6-mA function by overexpressing *Alkbh1*, which encodes the DNA N^6-mA demethylase[1,9,24,32–34]. Consistent with several previous studies[1,9,35], overexpressed ALKBH1 was localized primarily to the nucleus, but not the mitochondria, in ES cells and TSC-LCs (Extended Data Fig. 3d). As expected, overexpression of *Alkbh1* 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 ($P < 1 \times 10^{-6}$) occurred at SIDD identified by ssDNA-seq (Extended Data Fig. 3f). Upon overexpression of *Alkbh1*, the number of SATB1 ChIP–seq peaks more than doubled (Fig. 3c), whereas overall SATB1 levels remained unchanged (Extended Data Fig. 3g). The increased SATB1 peaks in cells overexpressing *Alkbh1* intersected significantly with differentially decreased N^6-mA peaks ($P < 1 \times 10^{-6}$, *Alkbh1* overexpression versus control); more than 50% of those intersected peaks (SATB1 increased and N^6-mA decreased) overlapped with ssDNA peaks (Fig. 3d). There was also a conspicuous increase in the SATB1 ChIP–seq signal at differentially decreased N^6-mA sites in *Alkbh1*-overexpressing cells compared to controls (Fig. 3e). In addition, the increased SATB1 peaks were enriched on young LINE-1s, but not other classes of endogenous transposons, and overlapped with ssDNA regions (Extended Data Fig. 3h, i, Supplementary Table 1). Thus, N^6-mA antagonizes the binding of SATB1 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 assay for transposase-accessible

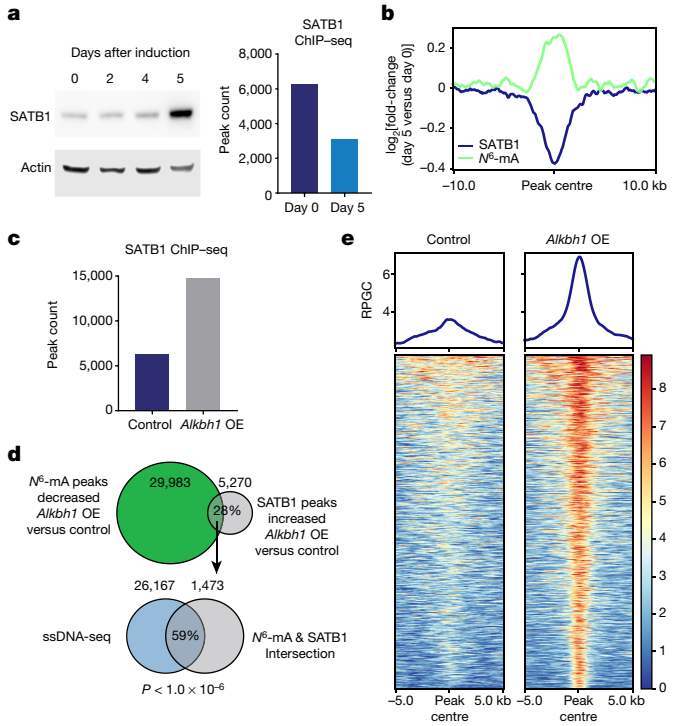


Fig. 3 | N^6-mA antagonizes binding of SATB1 to the chromatin during TSC development. a, SATB1 binding is inversely correlated with N^6-mA levels. Left, western blot of SATB1 expression during cell fate transition at indicated time points. The experiments were repeated independently three times with similar results. Right, number of SATB1 ChIP–seq peaks at indicated days. For blot source data, see Supplementary Fig. 1. **b**, Aggregation of differential signals of N^6-mA DIP and SATB1 ChIP centred at differentially increased N^6-mA peaks (day 5 versus day 0). **c**, Number of SATB1 ChIP–seq peaks in control or *Alkbh1*-overexpressing (OE) cells. **d**, Top, Venn diagram showing the intersection between N^6-mA differentially decreased peaks and SATB1 ChIP–seq differentially increased peaks in *Alkbh1*-overexpressing versus control 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 N^6-mA peak centres in control or *Alkbh1*-overexpressing cells.

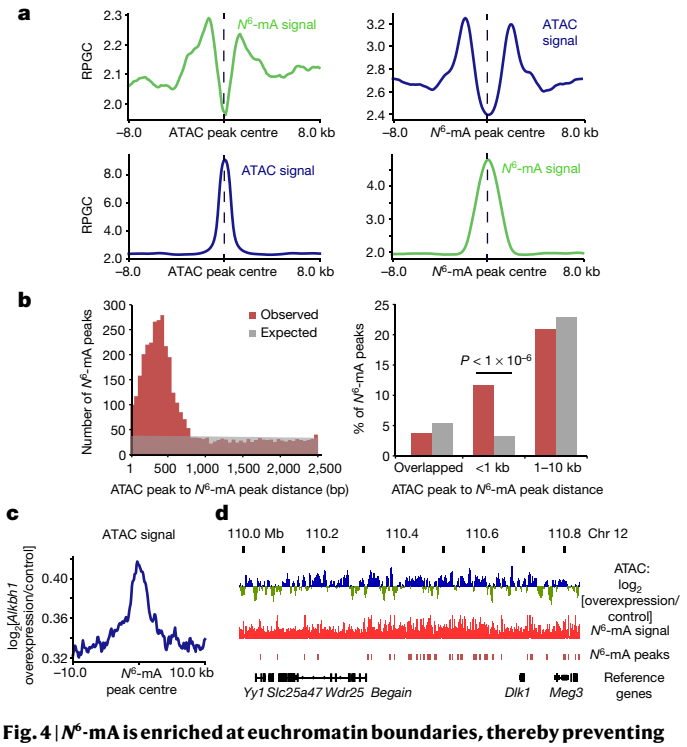


Fig. 4 | N^6-mA is enriched at euchromatin boundaries, thereby preventing the spreading of euchromatin. a, Aggregation profiles of day 5 ATAC-seq inserts at day 5 versus day 0 differentially increased N^6-mA peak centres (left) and day 5 N^6-mA DIP-seq signal at ATAC-seq day 5 peak centres (right). **b**, 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 harbouring 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 versus control cells at N^6-mA differentially decreased (overexpression versus control) peak centres and flanking regions. **d**, Differential *Alkbh1* overexpression versus control 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.

chromatin using sequencing (ATAC-seq) to interrogate chromatin accessibility in i*Cdx2* cells (Fig. 4, Extended Data Fig. 4). Although ATAC-seq and N^6-mA peaks are mutually exclusive at each other's peak centres (Fig. 4a), further analysis revealed a striking 'peak and valley' distribution pattern. N^6-mA signals are maximal at the boundaries of the ATAC-seq inserts, mostly 0.5–1.0 kb from the centres of the ATAC-seq peaks; similarly, ATAC-seq signals are maximal at the boundaries of the N^6-mA peaks (Fig. 4a). 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 (approximately 0.5–1 kb; Fig. 4b). Moreover, when *Alkbh1* was overexpressed, chromatin accessibility increased at the original (control) N^6-mA peaks and spread to much larger chromatin domains beyond the boundaries of N^6-mA peaks, such as at the *Dlk1–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.

N^6-mA is vital in early embryogenesis

Gene set enrichment analysis (GSEA) demonstrated that LINE-1 and genes involved in trophoblast differentiation were significantly upregulated in *Alkbh1*-overexpressing cells, with specificity for genes that are upregulated in spiral artery trophoblast giant cells (SpA-TGCs)[36], but not in the other trophectoderm lineages (Fig. 5a, Extended Data Fig. 5a, b), as confirmed by quantitative PCR with reverse transcription (RT–qPCR) (Extended Data Fig. 5c). Furthermore, the majority of imprinted genes, such as *H19*, which have critical roles in TSC and placental development[37,38], were upregulated when N^6-mA was removed by overexpression of *Alkbh1*; maternally expressed genes appeared to be more severely affected than paternally expressed genes (Extended Data Fig. 5d). We then confirmed the accelerated differentiation of TGC-like cells (TGC-LCs) by using 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*-knockout mice using the 'knockout-first' strategy[40] 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 concerns for off-target or secondary mutations; $Alkbh1^{1c/1c}$ mice were viable, fertile and indistinguishable from wild-type mice ($n = 52$ pups, 9 litters;

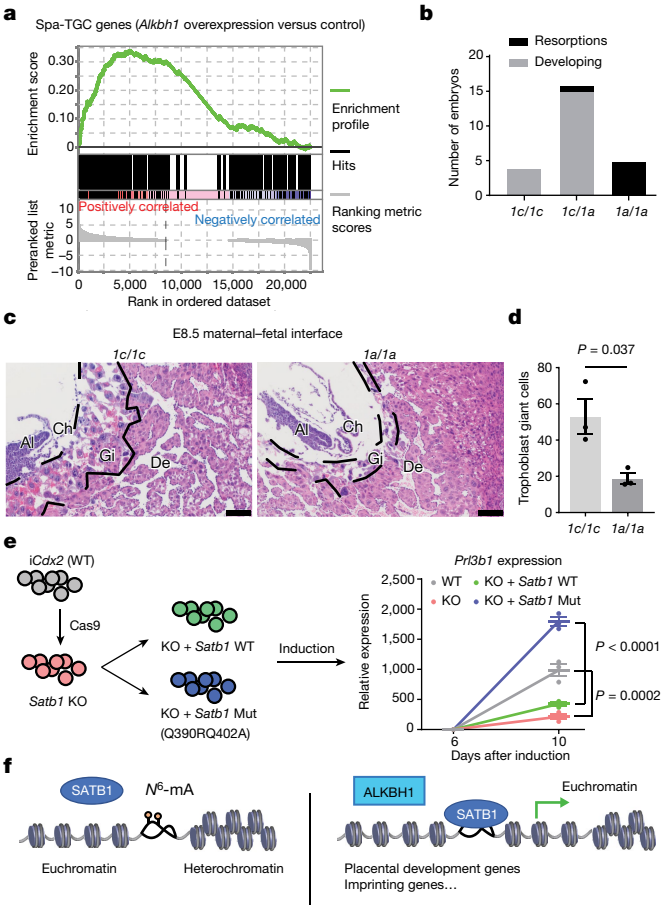


Fig. 5 | Epigenetic silencing pathway mediated by N^6-mA and SATB1 has critical roles in trophectoderm development. a, GSEA of 273 spiral artery trophoblast giant cell (Spa-TGC) genes reveals significant enrichment ($P < 0.001$) in *Alkbh1*-overexpressing cells compared to control cells. Enrichment score 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 four independent matings. **c**, Haematoxylin and eosin (H&E) staining of the maternal–fetal interface shows strong diminishing of TGCs in *Alkbh1*$^{1a/1a}$ placentas. *Alkbh1*$^{1a/1a}$ mouse placental tissues at E8.5 are depleted of secondary giant cells (Gi), which normally reside between the decidual (De) layer and the chorio-allantoic plate (Ch, Al), as seen in the *Alkbh1*$^{1c/1c}$ sample. Representative images of three litter-matched E8.5 or E12.5 embryos per genotype from three dams. Scale bars, 100 µm. **d**, Multi-plane quantification shows 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 ($t = 5.06$, degrees of freedom (df) = 2). Mean ± s.e.m. of five planes through the maternal–fetal interface of three litter-matched mice per genotype, from three dams. **e**, Left, schematic of *Satb1* knockout (KO) rescue experiment in i*Cdx2* cells overexpressing wild-type (WT) or N^6-mA tolerant mutant (Mut) *Satb1*. Right, RT–qPCR showing that the expression of a TGC-specific marker gene (*Prl3b1*) is more efficiently rescued by the *Satb1* N^6-mA tolerant mutant. One-way ANOVA ($P < 0.0001$) followed by Tukey's multiple comparisons test on data from day 10. Mean ± s.e.m. of three biological replicates. **f**, Proposed model of the function of N^6-mA, ALKBH1 and SATB1 at euchromatin boundaries.

Extended Data Fig. 5g, i). Furthermore, we leveraged an immunofluorescence approach[9] to interrogate N^6-mA levels in i*Cdx2* cells and embryos at embryonic day 8.5 (E8.5) (Extended Data Fig. 5j, k). Digestion with S1 nuclease, micrococcal nuclease, or DNase I (in addition to RNase A/T1) greatly reduced or completely abolished the signal, whereas treatment with RNase H did not affect the signal, indicating that the signal comes exclusively from DNA. Moreover, examination of the cohort of embryos from a timed mating experiment demonstrated that *Alkbh1*$^{1a/1a}$ embryos developed at expected Mendelian frequencies, but were 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 that deletion of *Alkbh1* induced a significant reduction in TGCs (Fig. 5c, d). RNA-seq of the maternal–fetal interface at E10.5 (Extended Data Fig. 5l, Supplementary Table 2) showed that a few key factors involved in trophoblast development were significantly reduced, including *Gpc3*, which is expressed by differentiating human syncytiotrophoblasts[41]. 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 finding that overexpression of *Alkbh1* facilitates TSC differentiation in cell culture.

As our results showed that N^6-mA antagonizes the interaction of SATB1 with chromatin, we next used a CRISPR–Cas9 approach to knock out *Satb1* in the i*Cdx2* system (Extended Data Fig. 6a). RNA-seq demonstrated that trophectoderm genes were inefficiently induced in *Satb1*-knockout cells (Extended Data Fig. 6b). Consequently, *Satb1* deficiency impairs differentiation of TSC-LCs into TGC-LCs, as assayed by flow cytometry and morphology (Extended Data Fig. 6c, d). Thirty-four per cent of genes activated by *Satb1* are also repressed by N^6-mA, as identified in *Alkbh1*-overexpressing cells (Extended Data Fig. 6e).

We next leveraged the aforementioned N^6-mA-tolerant SATB1 mutant (Q390RQ402A) (Fig. 2) to further elucidate the epistasis between *Satb1* and N^6-mA (Fig. 5e). We reconstituted *Satb1*-knockout cells with wild-type *Satb1* or the N^6-mA-tolerant *Satb1* mutant and showed that reconstitution with the tolerant mutant generated more polyploid TGC-LCs at a faster rate than with the wild-type gene (Extended Data Fig. 6c, d). Concordantly, the tolerant mutant rescued the expression of the downregulated trophectoderm lineage markers in the *Satb1*-knockout cells more efficiently than did the wild-type *Satb1* (Fig. 5e, Extended Data Fig. 6f). These results further support the notion that N^6-mA modulates the ES cell-to-TSC fate transition by antagonizing SATB1 function.

Discussion

In this work, we have shown how N^6-mA helps to regulate 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 ES cells from adopting a TSC fate. These findings are corroborated by genetic studies of *Alkbh1*-knockout mice. Notably, the DNA N^6-mA demethylase ALKBH1 also prefers the unpairing SIDD sequences[24]. These findings together 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 an unexpected role for N^6-mA as an antagonist of SATB1 function in vitro and in vivo. Strikingly, binding of SATB1 to SIDD sequences is completely abolished in the presence of N^6-mA (K_D undetectable); this is unusual for epigenetic effector binding proteins, which usually manifest several-fold differences in binding. Consistent with the long-appreciated role of SATB1, N^6-mA is specifically enriched at the boundaries between euchromatin and heterochromatin, 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 is important for TSC differentiation[37]. In parallel, this pathway may regulate gene expression by controlling the chromatin landscape at long ranges, as indicated by the changes in chromatin accessibility found here; this also warrants further research in the future.

In summary, we have shown that N^6-mA contributes to epigenetic regulation by antagonizing the function of SATB1. This study sheds new light on the mechanisms by which epigenetic modification and DNA secondary structures regulate chromatin structure and gene expression in early development.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-2500-9.

1. Wu, T. P. et al. DNA methylation on N^6-adenine in mammalian embryonic stem cells. *Nature* **532**, 329–333 (2016).
2. Dickinson, L. A., 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).
3. Kohwi-Shigematsu, T. & Kohwi, Y. Torsional stress stabilizes extended base unpairing in suppressor sites flanking immunoglobulin heavy chain enhancer. *Biochemistry* **29**, 9551–9560 (1990).
4. Bode, J. et al. Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science* **255**, 195–197 (1992).
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).
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).
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).
8. Yao, B. et al. DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. *Nat. Commun.* **8**, 1122 (2017).
9. Xie, Q. et al. N^6-methyladenine DNA modification in glioblastoma. *Cell* **175**, 1228–1243 (2018).
10. Kouzine, F. et al. Global regulation of promoter melting in naive lymphocytes. *Cell* **153**, 988–999 (2013).
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 (2017).
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).
13. Kohwi-Shigematsu, T. et al. SATB1-mediated functional packaging of chromatin into loops. *Methods* **58**, 243–254 (2012).
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).
15. Weber, M. et al. Genomic imprinting controls matrix attachment regions in the Igf2 gene. *Mol. Cell. Biol.* **23**, 8953–8959 (2003).
16. Alvarez, J. D. 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).
17. Fessing, M. Y. et al. p63 regulates Satb1 to control tissue-specific chromatin remodeling during development of the epidermis. *J. Cell Biol.* **194**, 825–839 (2011).
18. Asanoma, K. et al. SATB homeobox proteins regulate trophoblast stem cell renewal and differentiation. *J. Biol. Chem.* **287**, 2257–2268 (2012).
19. Ralston, A. et al. Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* **137**, 395–403 (2010).
20. Ghosh, R. P. et al. Satb1 integrates DNA binding site geometry and torsional stress to differentially target nucleosome-dense regions. *Nat. Commun.* **10**, 3221 (2019).
21. Kitagawa, Y. et al. Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. *Nat. Immunol.* **18**, 173–183 (2017).
22. Savarese, F. et al. Satb1 and Satb2 regulate embryonic stem cell differentiation and Nanog expression. *Genes Dev.* **23**, 2625–2638 (2009).
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).
24. Zhang, M. et al. Mammalian ALKBH1 serves as an N^6-mA demethylase of unpairing DNA. *Cell Res.* **30**, 197–210 (2020).
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. Chem. 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).
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).
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).
29. Mirkovitch, J., Mirault, M.-E. & Laemmli, U. K. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell* **39**, 223–232 (1984).
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).
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).
32. Zhou, C., Liu, Y., Li, X., Zou, J. & Zou, S. DNA N^6-methyladenine demethylase ALKBH1 enhances osteogenic differentiation of human MSCs. *Bone Res.* **4**, 16033 (2016).
33. Xiao, C. L. et al. N^6-methyladenine DNA modification in the human genome. *Mol. Cell* **71**, 306–318 (2018).
34. Tian, L.-F. et al. Structural basis of nucleic acid recognition and 6mA demethylation by human ALKBH1. *Cell Res.* **30**, 272–275 (2020).
35. Ougland, R. et al. ALKBH1 is a histone H2A dioxygenase involved in neural differentiation. *Stem Cells* **30**, 2672–2682 (2012).
36. Nelson, A. C., Mould, A. W., Bikoff, E. K. & Robertson, E. J. Single-cell RNA-seq reveals cell type-specific transcriptional signatures at the maternal–foetal interface during pregnancy. *Nat. Commun.* **7**, 11414 (2016).
37. Fujimori, H. et al. The H19 induction triggers trophoblast lineage commitment in mouse ES cells. *Biochem. Biophys. Res. Commun.* **436**, 313–318 (2013).
38. Esquiliano, D. R., Guo, W., Liang, L., Dikkes, P. & Lopez, M. F. 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).
39. Nordstrand, L. M. et al. Mice lacking Alkbh1 display sex-ratio distortion and unilateral eye defects. *PLoS One* **5**, e13827 (2010).
40. Skarnes, W. C. et al. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).
41. Khan, S. et al. Glypican-3 (GPC3) expression in human placenta: localization to the differentiated syncytiotrophoblast. *Histol. Histopathol.* **16**, 71–78 (2001).
42. Cuffe, J. S. M. et al. The effects of gestational age and maternal hypoxia on the placental renin angiotensin system in the mouse. *Placenta* **35**, 953–961 (2014).
43. Stepan, H., Faber, R., Stegemann, S., Schultheiss, H. P. & 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).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

i*Cdx2* ES cell differentiation

The Tet-inducible-*Cdx2*-Flag ES cell line (i*Cdx2* cells) was obtained from the NIA mES cell bank[27]. i*Cdx2* cells were maintained on feeder cells in DMEM supplemented with 15% fetal bovine serum (Gibco), 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 μg/ml doxycycline (Sigma). This line and all i*Cdx2*-derived lines were tested for mycoplasma contamination prior to experimentation. When starting differentiation[44], i*Cdx2* cells were selectively seeded on gelatin-coated plates after stepwise elimination of the feeder mouse embryonic fibroblasts, then *Cdx2* overexpression was induced by removing doxycycline. The cells were split at day 1 after doxycycline withdrawal and cell samples were collected on the following days.

Mouse ES cells in 2i/L culture 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 medium with glutamine (Invitrogen) supplemented with B27 (Invitrogen), 1× Pen/Strep (Invitrogen), 1,000 units of mLIF (Milipore), 1.5 μM CGP77675 (Tocris) and 3 μM CHIR99021 (Tocris). PKCi/L medium contains a 1:1 mixture of DMEM/F12 supplemented with N2 (Invitrogen) and neurobasal medium with glutamine (Invitrogen) supplemented with B27 (Invitrogen), PenStrep (Invitrogen), 1,000 units of mLIF (Milipore) and 5 μM Gö6976 (Tocris).

Generation of *Satb1*-knockout cell line

Two guide RNAs targeted to flank 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). Twenty-four hours after transfection, 400 μg/ml hygromycin was added to the medium for 2 days. ES cells were then seeded at low density to obtain single-cell-derived colonies. Then, 72 ES cell colonies were randomly picked up and screened by PCR genotyping (Extended Data Fig. 6a). PCR screening primer sequences can be found in Supplementary Table 3.

Establishment of *Satb1*-overexpressing cell line

The wild-type and mutant (Q390RQ402A) human *Satb1-myc* DNA sequences 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 the 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 A/T1 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 h or overnight. The remaining steps were performed as described in the kit, and the DNA was eluted with ddH_2O. Furthermore, to remove trace RNA contamination, eluted DNA was treated with once more with RNase A/T1 for 2 h and then RNase H (NEB M0297S) overnight, followed by phenol-chloroform extraction and ethanol precipitation.

N^6-mA dot blotting

DNA samples were denatured at 95 °C for 5 min, cooled 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 PBS, 0.1% Tween 20 (PBS-T)) for 1 h at room temperature, then incubated with N^6-mA antibody (1:1,000, Synaptic Systems 202-003) at 4 °C overnight. Membranes were washed with PBS-T three times and then incubated with HRP-linked secondary anti-rabbit IgG antibody (1:5,000, Cell Signaling 7074S) for 1 h at room temperature. After three washes with PBS-T, the membrane signal was detected with SuperSignal West Dura Extended Duration Substrate kit (Thermo Fisher 34076).

MAR DNA extraction

The MAR DNA extraction was performed as described[29,45] with modifications. In brief, the cell pellets were resuspended with buffer A (10 mM HEPES, PH 7.0, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, and proteinase inhibitor cocktail (Roche)) and treated with 0.1% Triton X-100 for 7 min on ice. After a wash with buffer A, the cells were resuspended with buffer A and MNase buffer (NEB). Five microlitres of MNase was added into the cell suspension and incubated at 37 °C for 5 min, 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 min, 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 min, the cell pellets were resuspended with buffer A and treated with RNase A/T1 for 2 h and then RNase H at 37 °C, then Proteinase K was added and incubated at 56 °C for 2 h. 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 three times and fixed with cold 70% ethanol overnight at 4 °C. After washing with cold PBS three times, cells were resuspended and stained with 0.5 μg/ml DAPI for 10 min, then washed with cold PBS two more times. Samples were analysed by a BD LSR II.

Liquid chromatography with tandem mass spectrometry analysis of N^6-methyl-2′-deoxyadenosine

We added 0.1 U nuclease P1, 0.25 nmol *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 $ZnCl_2$ to 1 μg DNA. 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 alkaline phosphatase, 2.5×10^{-4} U phosphodiesterase 1 and 4 μl 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 1.0 M formic acid. Then 200 pmol ^{15}N-labelled 2′-deoxyadenosine and 25 fmol D_3-labelled 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 ddH_2O, to which 90 μl acetonitrile was added. The solution was subsequently centrifuged, and the supernatant was dried and re-dissolved in 50 μl acetonitrile/ddH_2O (95:5, v/v) for liquid chromatography with tandem mass spectrometry (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) H_2O 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 used, and the flow rate was 100 μl/min. The injection volume was 25 μl. High-energy collisional dissociation was used and the transitions of *m/z* 252.1→136.0 (2′-deoxyadenosine), *m/z* 257.1→141.0 ([$^{15}N_5$]2′-deoxyadenosine), *m/z* 266.1→150.0 (N^6-methyl-2′-deoxyadenosine) and *m/z* 269.1→153.0 ([D_3]N^6-methyl-2′-deoxyadenosine) were monitored for quantifications.

Protein expression and purification

The cDNAs 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. In brief, 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 μl/s for 300 s association and then flowing running buffer for 300 s dissociation, followed by 200 s regeneration buffer at 3 μl/s. Binding data were collected and analysed 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 rpm vibration at 25 °C.

Isothermal titration calorimetry

For ITC measurement, synthetic DNA oligos and the purified proteins were extensively dialysed 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 DIP sequencing and analysis

N^6-mA DIP was performed as described[1]. In brief, 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 ssDNA 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 adaptor 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 Bowtie2 (2.3.1, default settings)[46]. To identify uniquely mapped N^6-mA peaks before 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 generated very similar results (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. LINE-1 transposons were grouped by evolutionary age into young: < 1.5 million years (Myr; L1Md_A, L1MdT, L1Md_Gf); middle: > 1.5 Myr & < 6 Myr (L1Md_F, L1VL); old > 6 Myr (Lx, L1_Mus) as previously described[48]. Intersections were performed by USeq[49]. 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 μg sonicated genomic DNA. The spiked-in oligos were heated at 95 °C for 10 min then slowly cooled down in a PCR machine and tested by agarose gel electrophoresis. Then DIP was performed according to the denaturing DNA IP protocol (Active Motif 55010). After denaturing under 95 °C for 10 min, DNA was quickly transferred on ice for 5 min, then 5 μg 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 h at 4 °C. After washing 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 or CFX384 system (BioRAD).

ssDNA-seq

ssDNA-seq was performed as previously described with minor modifications[11]. ES cells (8×10^7) were cultured in feeder-free conditions. Low-salt buffer (15 mM Tris-HCl, pH 7.5, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.5 mM EGTA, and 300 mM sucrose) at 37 °C was added for 5 min. Cells were then treated with 100 mM $KMnO_4$ 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 μg/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 h 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 of 2–10 kb, the 50 U S1-treated samples (treated and blank) were chosen for further processing. Next, 35 μg 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 1× 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 tenfold 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 2 × 100 bp in one lane of an Illumina HiSeq4000.

Sequencing reads were filtered and pre-analysed 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 $\log_2$ 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), with window size 200 bp and gap size 600 bp.

Analysis of ssDNA enrichment in transposable elements

Raw fastq reads were analysed 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 index[50]. Standalone aggregation profiles were constructed from bigWigs without filtering for uniquely mapped reads over RepeatMasker annotations of full-length (longer than 6 kb) young LINE-1 elements (L1Md_A, Tf, and Gf). Heatmaps used only uniquely mapped data over the same annotated LINE-1 regions.

Chromatin immunoprecipitation

Cells (10^7) were washed once with DPBS and then fixed in 5 ml DMEM medium with 1% formaldehyde for 10 min at room temperature with rotation. Fixation was terminated with addition of 100 mM glycine followed by 5 min rotation at room temperature. Cells were then scraped and spun down at 4 °C at 1,000 rpm for 5 min, followed by two washes with 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 1,400*g* for 5 min and resuspended in 1 ml lysis buffer 2 (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 1,400g for 5 min and resuspended in 200–300 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. Ten microlitres of 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 h at 4 °C, samples were extensively washed three 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 A/T1 mix and incubated at 37 °C for 1 h, followed by the addition of 3 μl proteinase K and incubation for 2 h 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-analysed 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 N^6-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 > twofold enrichment of intersections compared to the mean of genome random.

ATAC-seq

ATAC-seq was performed as described[51]. 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 $MgCl_2$, 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 × DPBS, 0.01% Tween-20, 1 × TD buffer (Nextera DNA Library Prep Kit, Illumina) and transposase (100 nM final concentration, Illumina). Transposition reaction was incubated at 37 °C for 30 min 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 five 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 USeq.

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 and 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) 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. GSEA on RNA-seq data was conducted through the GSEA Preranked tool[53] based on cuffdiff output and the target gene sets (SpA-TGCs or Prdm1-TGCs[36]). 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 14[54]. Part of the data analysis was done by in-house customized scripts in R. Supplementary 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 or CFX384 system (BioRAD).

Mouse breeding

Animal experiments were performed with approval from Yale University's Institutional Animal Care and Use Committee and in accordance with all relevant ethical regulations and guidelines of Yale University and the NIH. C57BL/6N-*Alkbh1*$^{tm1a(EUCOMM)Hmgu}$/Ieg (*Alkbh1*$^{tm1a/wt}$) males and females were purchased from the EMMA repository (derived from a C57BL/6N background ES cell 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. Intra-litter 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 more than five generations before performing subsequent experiments.

Timed matings were performed by synchronizing the oestrus cycle of 12-week-old female *Alkbh1*^{tm1a} heterozygotes via exposure to male soiled bedding, and subsequently placing two 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 E0.5. Pregnant animals were euthanized using CO_2 per university guidelines, and embryos and 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 N^6-mA immunofluorescence 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 was performed by YPTS.

N^6-mA immunofluorescence 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 hydrolysed by 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-N^6-methyladenine monoclonal antibody (Cell Signaling Technologies 56593, lot 1) overnight at 4 °C. After rinsing in PBS-T, slides were incubated with 1:1,000 goat-anti-rabbit IgG Alexa Fluor 488 (Invitrogen A11008) at room temperature for 60 min. Slides were rinsed in PBS-T for 5 min, incubated in 1:1,000 DAPI for 3 min and rinsed in PBS three times. After immunofluorescence 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* or 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 h, 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 min. Slides were mounted with Fluormount-G.

N^6-mA staining in cultured cells was performed as described[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 min, 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 h, 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 h. DNA was stained with DAPI for 5 min. 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 2N 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 2N HCl.

Microscopy

Immunofluorescence microscopy was performed within 7 days of staining using (1) an Olympus IX81 Fluoresence 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 10× 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. Five sections spanning 128 µm were chosen for TGC quantification from representative interfaces, and TGCs between the chorioallantoic plate (E8.5)/labyrinth (E12.5) and decidua were counted manually. Quantitative analysis was performed on three litter-matched sets of embryos from three different litters by individuals blinded to the genotype of the sample.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequencing data generated for this study have been deposited in the Gene Expression Omnibus database with accession numbers GSE126077 and GSE126227. All other data, materials, and custom code are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

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).
45. Pathak, R. U., Srinivasan, A. & Mishra, R. K. Genome-wide mapping of matrix attachment regions in *Drosophila melanogaster*. *BMC Genomics* **15**, 1022 (2014).
46. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
47. Zang, C. et al. A clustering approach for identification of enriched domains from histone modification ChIP-seq data. *Bioinformatics* **25**, 1952–1958 (2009).
48. Castro-Diaz, N. et al. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* **28**, 1397–1409 (2014).
49. Nix, D. A., Courdy, S. J. & Boucher, K. M. Empirical methods for controlling false positives and estimating confidence in ChIP-seq peaks. *BMC Bioinformatics* **9**, 523 (2008).
50. Jeong, H.-H., Yalamanchili, H. K., Guo, C., Shulman, J. M. & Liu, Z. An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. *Pac. Symp. Biocomput.* **23**, 168–179 (2018).
51. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
52. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
53. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
54. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).
55. Bradley, A. et al. The mammalian gene function resource: the International Knockout Mouse Consortium. *Mamm. Genome* **23**, 580–586 (2012).

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. by NIH Medical Scientist Training Program Training Grant T32GM007205; M.H.A.III by NSF GRFP (DGE1752134) and the Gruber Foundation Science Fellowship; Y.W. by R35 ES031707; Z.J. by USDA-NIFA (2019-67016-29863), the Audubon Center for Research of Endangered Species and USDA-NIFA W4171; and H.L. by the National Natural Science Foundation of China (91753203 and 31725014) and the National Key R&D Program of China (2016YFA0500700).

Author contributions Z.L., R.V.N. and T.P.W. designed the cell biology, genetics and genomic experiments and analysed the data. T.P.W. established the i*Cdx2* 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 analysed the data. K.L. and T.P.W. analysed the sequencing data. M.H.A.III and R.V.N. generated the *Alkbh1*-knockout mice and characterized the phenotype. C.G., P.W. and Y.W. performed the MS experiments and analysed 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 manuscript.

Competing interests The authors declare no competing interests.

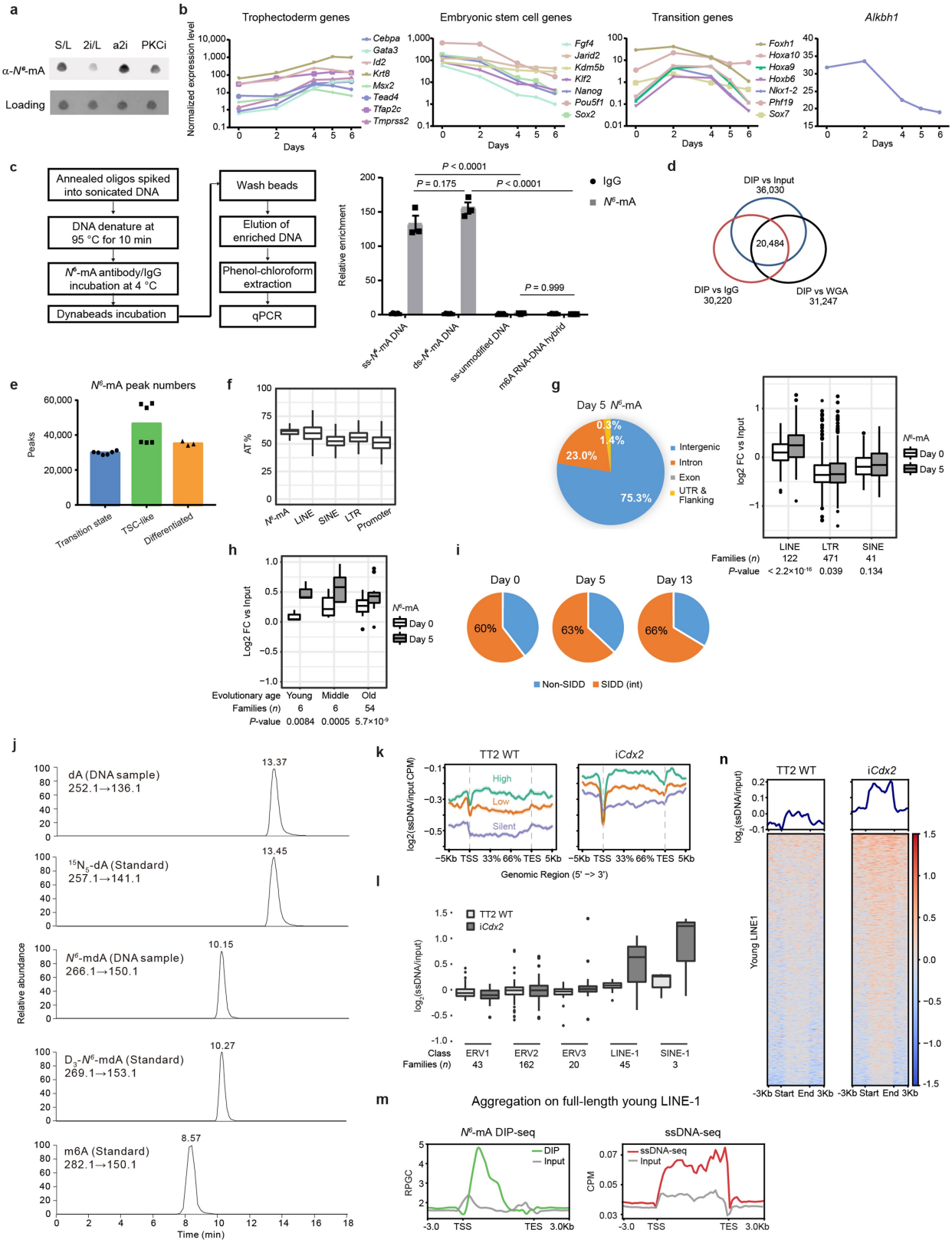
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-2500-9.

Correspondence and requests for materials should be addressed to H.L. or A.Z.X.

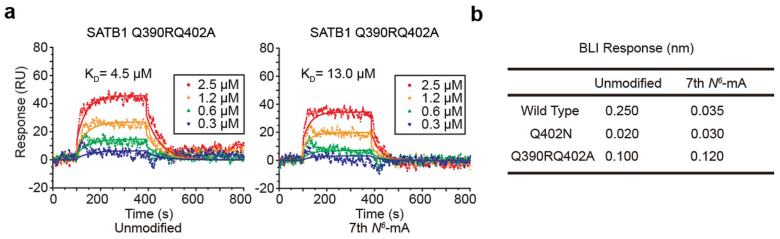
Peer review information *Nature* thanks Miguel Branco, Arne Klungland and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.



Extended Data Fig. 1 | See next page for caption.

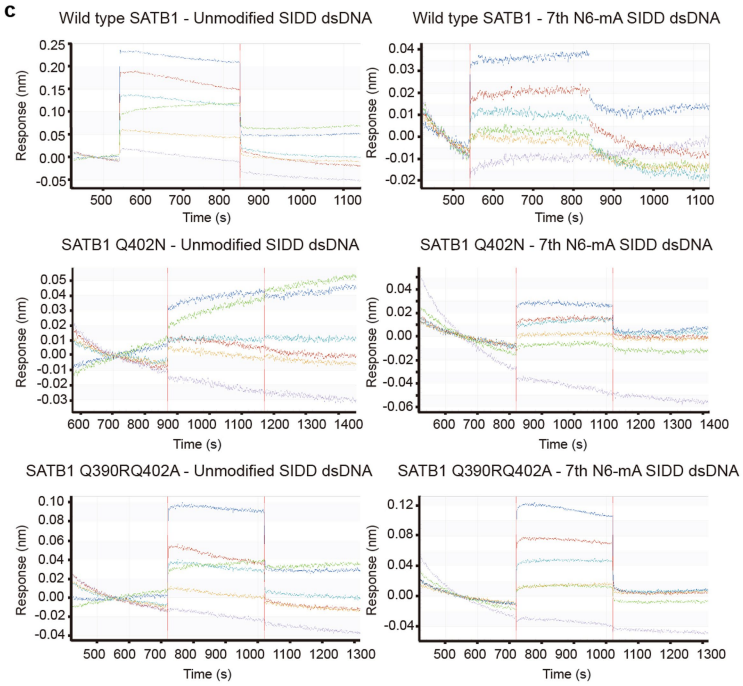
Extended Data Fig. 1 | N^6-mA upregulation during TSC development. a. Dot blotting of N^6-mA in ES cells cultured in various conditions. Methylene blue staining was performed as a DNA loading control. S/L, serum and LIF; 2i/L, Gsk3β, ERK inhibitor, and LIF; a2i, Gsk3β and Src inhibitor; PKCi, PKC inhibitor. The experiments were repeated independently three times with similar results. For blot source data, see Supplementary Fig. 1. **b**. RNA-seq analysis revealed a transition state with unique gene expression signatures before 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 ± s.e.m. of three 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-N6mA 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; TSC-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) × 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: Log_2 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 versus day 0 are provided. Raw data are in Supplementary Table 1. **h**. Log_2 FC enrichment of N^6-mA DIP-seq over input reads for different LINE-1 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 versus 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 labelled 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 three times with similar results. **k**. Average log_2 FC signal of ssDNA-seq in TT2 wild type (WT) and i*Cdx2* ES at highly expressed (FPKM > = 10), low (0.5 < = FPKM <10), and silent genes (FPKM <0.5). Kb, kilobase; TSS, transcription start site; TES, transcription end site. **l**. Log_2 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 i*Cdx2* ES cells at young LINE-1 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 × IQR when outliers that fall outside that range are present.

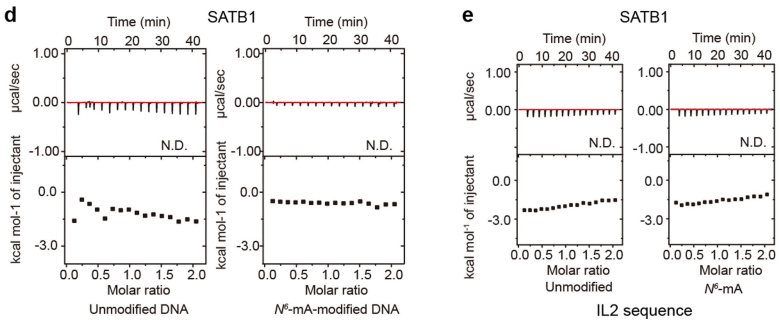


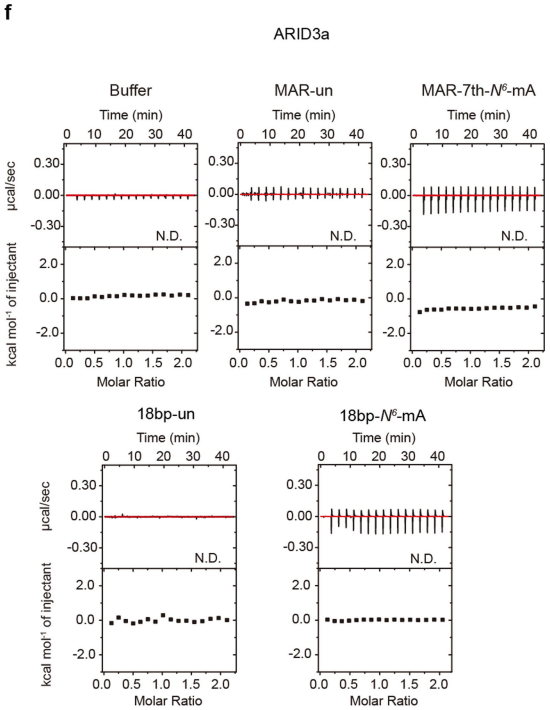
b

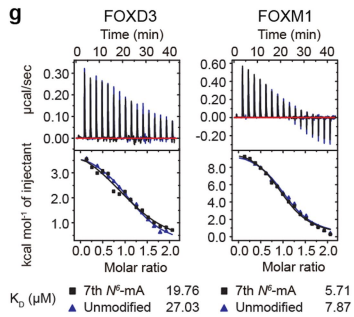
BLI Response (nm)

	Unmodified	7th N^6-mA
Wild Type	0.250	0.035
Q402N	0.020	0.030
Q390RQ402A	0.100	0.120

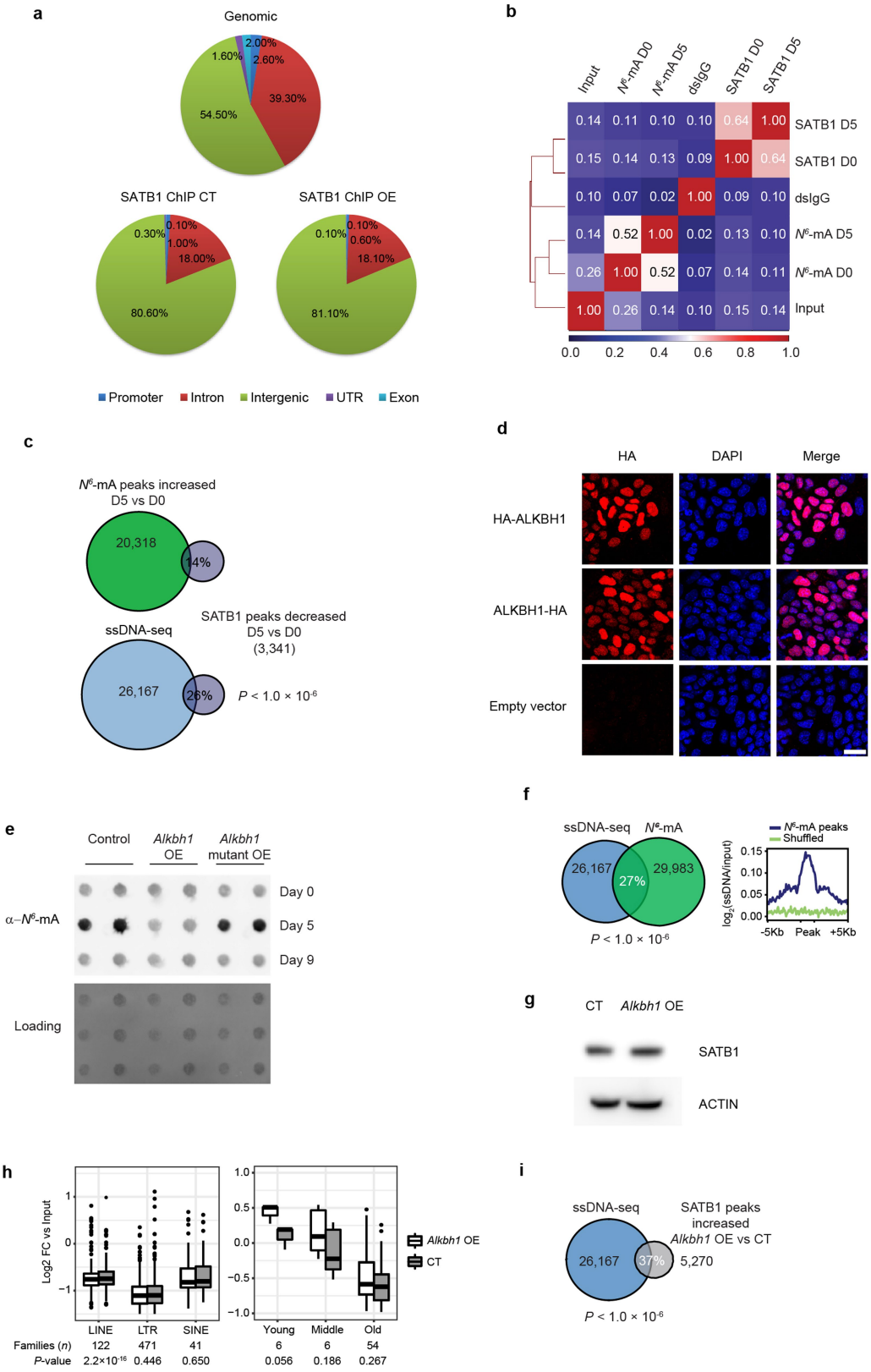






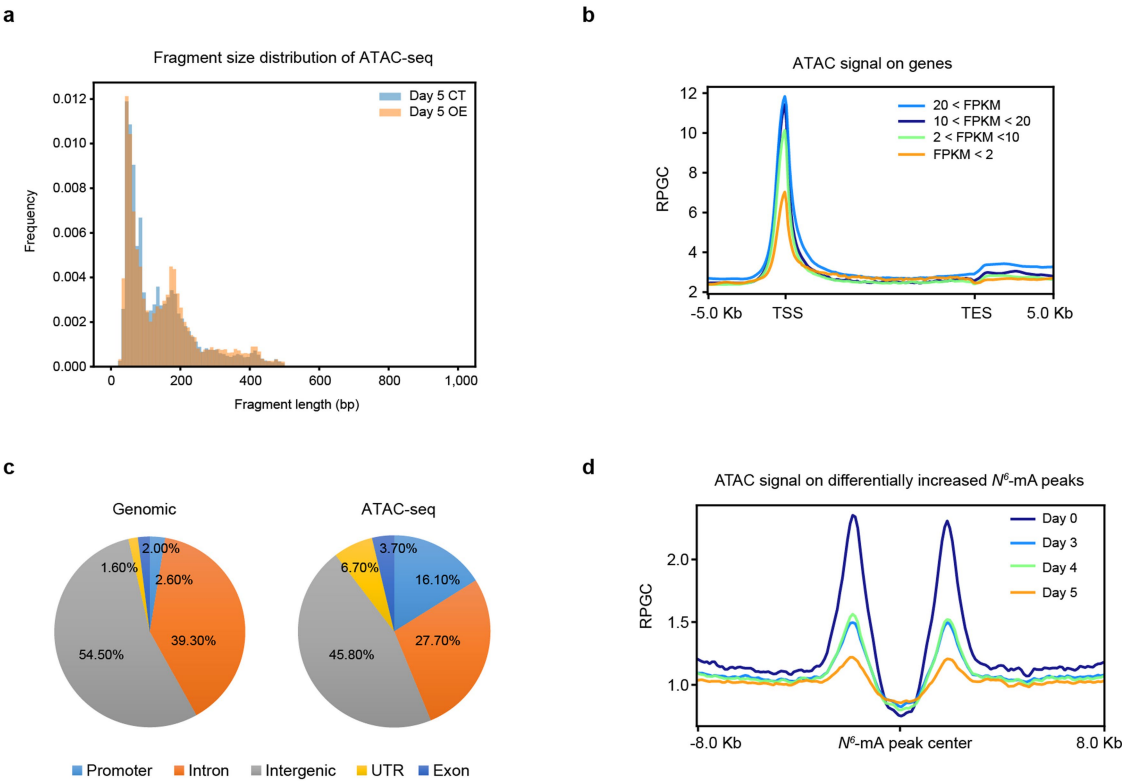


Extended Data Fig. 2 | Summary data and BLI binding curves of SATB1 and mutants, and ITC titration curves and fitting curves of control DNA sequences and proteins. a. The SPRi fitting curves of unmodified and 7th N^6-mA modified SIDD dsDNA binding to SATB1Q390RQ402A. RU, response units. **b**. Summary of BLI response signals from unmodified and 7th N^6-mA modified SIDD dsDNA binding to wild type and mutant SATB1 proteins. **c**. Original BLI binding curves of SATB1 and mutants with unmodified, and N^6-mA modified SIDD dsDNA substrates at the 7th position respectively. **d**. ITC titration curves and fitting curves of SATB1 titrated into unmodified, and N^6-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 N^6-mA modified SIDD dsDNA substrates derived from IL-2 sequences. The sequence is 5′-AGATGAAAAGAATAAATGTT TAGATTTGTTGATTAAA-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 three independent experiments with similar results.

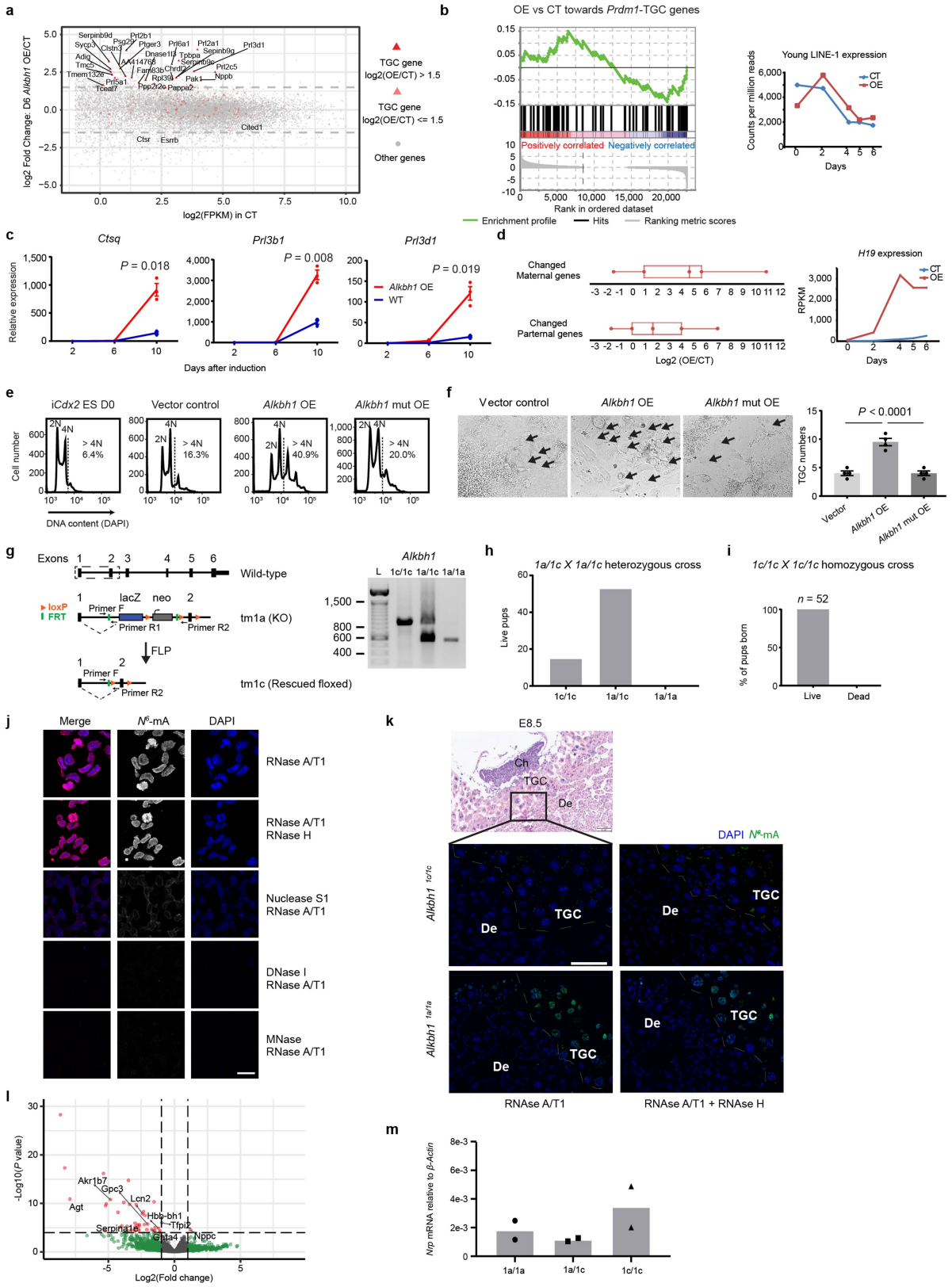


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 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 versus 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 i*Cdx2* 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 µm. The experiment was repeated independently three 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 three times with similar results. **f**. Left: Venn diagram showing the percentage of N^6-mA DIP-seq differentially decreased *Alkbh1* OE versus 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 versus 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 three times with similar results. **h**. Top: Log_2 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: Log_2 fold-change (FC) enrichment of Satb1 ChIP-seq over input reads for different LINE-1 elements grouped by evolutionary age (young: < 1.5 Myr; middle: > 1.5 Myr & < 6 Myr; old > 6 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 × IQR when outliers that fall outside that range are present. **i**. Venn diagram showing the percentage of SATB1 differentially increased *Alkbh1* OE versus 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 Fig. 1.

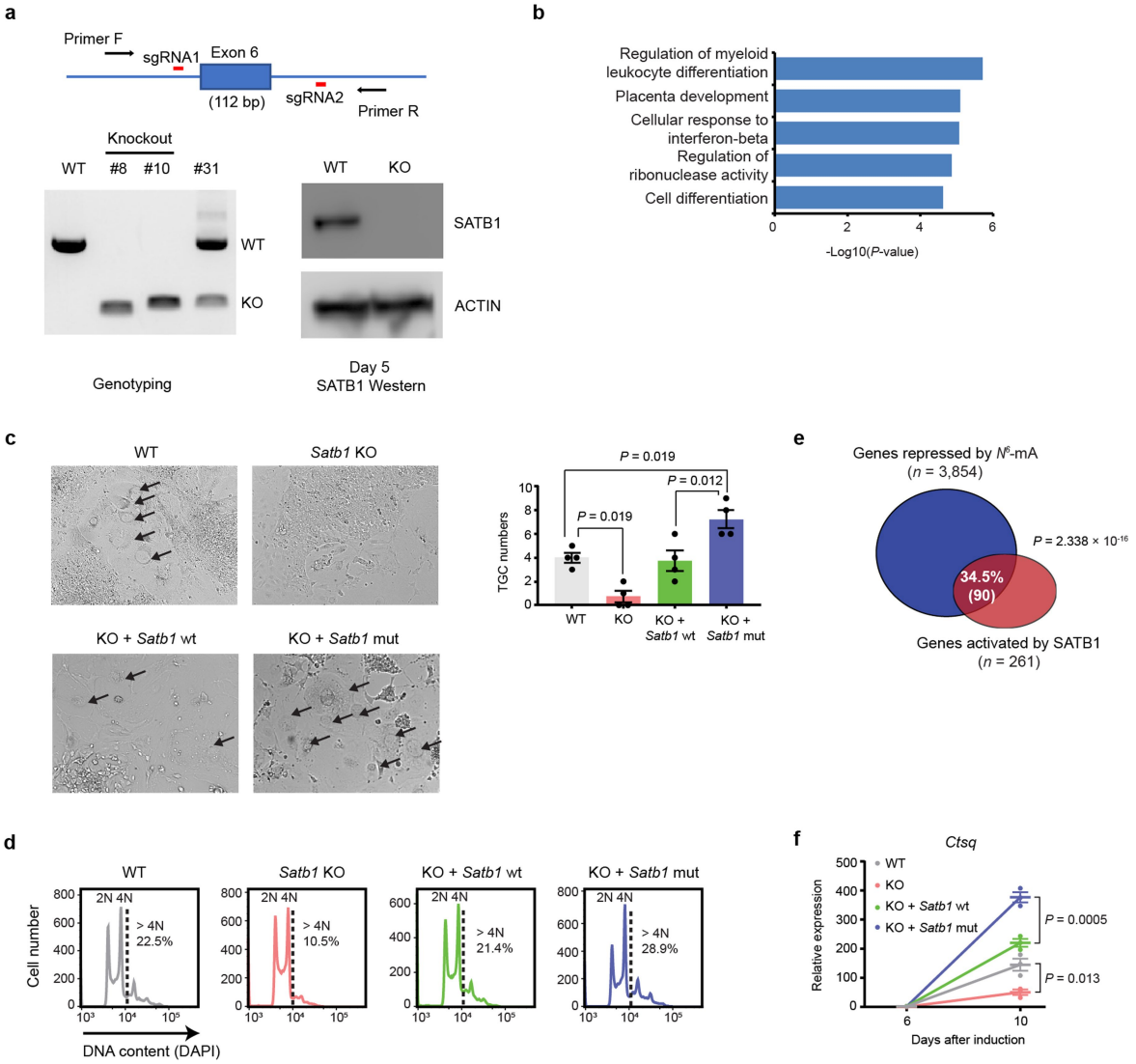


Extended Data Fig. 4 | ATAC-seq signal is diminished at N^6-mA DIP-seq peaks. a. Fragment size distribution of ATAC-seq inserts, which culminates around 100 bp, is unchanged in vector control (CT) and *Alkbh1*-overexpressing (OE) cells at day 5. **b**. ATAC-seq signal in control cells at day 5 is enriched 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 centred at N^6-mA DIP-seq differentially increased peaks (day 5 versus day 0). The signal diminishes at regions that gain N^6-mA signals (see Fig. 1b) during cell fate transition.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 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) versus 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 $\log_2$ 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 are downregulated 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 time points and genes; adjusted *P* values that reached significance are provided. Mean ± s.e.m. of three 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 twice with similar results. **f**. More TGC-LCs formed in *Alkbh1* OE cells than in WT cells. Black arrows indicate the TGC-LCs. Quantification of TGC cells shown as mean ± s.e.m. of four randomly chosen high-power fields. One-way ANOVA ($P < 0.0001$) followed by Tukey's multiple comparisons test. The experiments were repeated independently twice 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 Fig. 1. **h**. Crosses of $Alkbh1^{1a/1c}$ heterozygotes generate no 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 nine 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 µm. The experiments were repeated independently twice 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 four planes. H&E scale bar, 50 µm; immunofluorescence scale bar, 100 µm. TGC, trophoblast giant cells; De, decidua; Ch, chorion. **l**. mRNA-seq of the maternal-fetal interface of E10.5 $Alkbh1^{1a/1a}$ embryos demonstrates significant downregulation of genes involved in renin-angiotensin homeostasis, clotting pathways, syncytiotrophoblast differentiation, and innate immunity compared to litter-matched controls. 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 two biological replicates per genotype.



Extended Data Fig. 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 three times with similar results. For gel and blot source data, see Supplementary Fig. 1. **b**. Gene ontology enrichment analysis of downregulated 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*-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 ± s.e.m. of four randomly chosen high-power fields. One-way ANOVA ($P < 0.0001$) followed by Tukey's multiple comparisons test. The experiments were repeated independently twice 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 twice with similar results. **e**. The overlap of genes upregulated by *Alkbh1* overexpression and genes downregulated in *Satb1* KO at day 0. Significance of gene set intersections was determined by the chi-squared test (X^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* N^6-mA tolerant mutant. One-way ANOVA ($P < 0.0001$) followed by Tukey's multiple comparisons test on day 10 data. Mean ± s.e.m. of three biological replicates.

nature research

Corresponding author(s): Andrew Xiao

Last updated by author(s): Apr 22, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Plexera SPR Data Analysis Module, MetaMorph Software (Molecular Devices, v7.7.0), LAS AF software suite (Leica Microsystems, v2.6.0), CellSens (Olympus, v1.17)
Data analysis	Bowtie2 (v2.3.1), STAR (v2.5.2b), SAMtools (v1.9), BEDTools (v2.29.2), SICER1.1, Cufflinks (v2.2.1), USeq, Picard (v2.9.0), MACS2 (v2.1.1), deepTools (v3.1.3), SalmonTE (v0.8.2), HTSeq (v0.11.2), DESeq2, Pysam (v0.14.1), GSEAPreranked, PANTHER (v14), CEAS (v1.0.2), Origin 7.0, GraphPad Prism 7.0, FlowJo (v10.5.3), R, Python, Perl

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data generated in 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 experiments are provided, and uncropped images of gels and blots are in the supplement. All other data, materials, and custom code are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed. A sample size of three independent samples were chosen for cell culture and pulldown experiments to be able to do statistical testing. Sequencing experiments had various sample sizes depending on the method. N6-mA DIP had three samples per time point and various controls. ChIP-seq had one sample per condition or time point, as well as an input control. ATAC-seq had one sample per condition or time point. ssDNA-seq had one sample and input control per cell line. RNA-seq of cell lines had one sample per condition or time point; RNA-seq of animal samples had two per genotype. Animal experiments had various sample sizes. For breeding experiments, many embryos/pups from multiple crosses were analyzed to have sufficient power to determine changes in Mendelian distributions. Staining and histological quantification was performed on 2-3 biological samples with multiple sections to account for variability between section planes.
Data exclusions	None.
Replication	We repeated most experiments three times, as noted in the figure legends. All attempts gave similar results.
Randomization	Not relevant. This is time-series dynamics research where we compared KO vs. WT conditions or OE vs. WT.
Blinding	Not relevant. This is time-series dynamics research where we compared KO vs. WT conditions or OE vs. WT.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-m6A (Synaptic Systems, Cat#202003, Dot blot: 1:1000, IP: 5 µg/sample, ICC: 1:400), anti-SATB1 (Abcam, Cat#Ab109122, Clone:EPR3951, WB: 1:1000), anti-HA (Cell Signaling Technology, Cat#3724S, Clone:C29F4, 1:500), anti-β-Actin HRP (Santa Cruz Biotechnology, Cat#sc-47778 HRP, Clone:C4, 1:1000), anti-N6-methyladenosine monoclonal antibody (Cell Signaling Technology, Cat#56593, Clone:D9D9W, Lot#1, IHC: 1:400), rabbit IgG (Active Motif, Cat#55010), anti-rabbit IgG-HRP (Cell Signaling Technology, Cat#7074S, WB/Dot blot: 1:5000), anti-rabbit IgG Alexa Fluor 488 (Invitrogen, Cat#A11008, 1:1000), anti-Rabbit IgG Alexa Fluor Plus 594 (Invitrogen, Cat#A32740, 1:1000)
Validation	Primary antibody validation: anti-m6A (Synaptic Systems): Validated by manufacturer for use in mouse for dot blot, immunoprecipitation (IP) and immunocytochemistry (ICC/IF). We further validated ICC in this manuscript with various control samples (Extended Data Fig. 5j). It has also been validated for these uses in independent previous studies by other researchers. anti-SATB1: Validated by manufacturer for use in mouse for WB, IP. Further validated in our manuscript for WB with a negative control (Extended Data Fig. 6a) and for chromatin-IP by a published study (PMID: 25847946). anti-HA: Validated by manufacturer for cells expressing HA-tagged proteins for IF. anti-β-Actin HRP: Validated by manufacturer for use in mouse for WB. anti-N6-methyladenosine (Cell Signaling Technology): Validated by manufacturer for specificity against N6-methyladenosine.

All the secondary antibodies were validated by the manufacturers for use in WB/dot blot or ICC where applicable.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	293T cell line (CRL-3216) was purchased from ATCC , TT2 ES cell line (AES0014) was obtained from RIKEN BioResource Center and Cdx2 ES cell line was bought from NIA mouse ES cell bank.
Authentication	The Cdx2 ES line was authenticated by WB detection of Flag-Cdx2 after withdrawal of doxycycline and detection of the transgene in RNA-seq data. The remaining cell lines were not authenticated after receipt from the cell banks.
Mycoplasma contamination	Cell lines were negative for mycoplasma.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mus musculus, C57BL/6N-Alkbh1<tm1a(EUCOMM)Hmgu/Ieg>, Sex: M,F, Age: E8.5-17.5, 6-18 weeks Mus musculus, C57BL/6J (Jackson Laboratory), Sex: M, Age: 6-12 weeks (Breeding only) Mus musculus, 129S4/SvJaeSor-Gt(ROSA)26Sor<tm1(FLP1)Dym/J> (Jackson Laboratory), Sex: M,F, Age: 6-12 weeks (Breeding only)
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	All animal work was pre-approved by the Yale University IACUC.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	ssDNA-seq: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126077 ATAC-seq, ChIP-seq, RNA-seq, MeDIP-seq: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE126227
Files in database submission	GSE126077 ssDNA-seq: peaks_cdx2ES-w200-g600.bed.gz peaks_TT2-w200-g600.bed.gz peaks_cdx2ES-w100-g200.bed.gz cdx2ES_input_R1.fastq.gz cdx2ES_input_R2.fastq.gz cdx2ES_S1_R1.fastq.gz cdx2ES_S1_R2.fastq.gz TT2_input_R1.fastq.gz TT2_input_R2.fastq.gz TT2_S1_R1.fastq.gz TT2_S1_R2.fastq.gz GSE126227 RNA-seq raw fastq files: Cdx2_ES_ALKBH1CTRL_D0_RNAseq Cdx2_ES_ALKBH1CTRL_D2_RNAseq Cdx2_ES_ALKBH1CTRL_D4_RNAseq Cdx2_ES_ALKBH1CTRL_D5_RNAseq Cdx2_ES_ALKBH1CTRL_D6_RNAseq Cdx2_ES_ALKBH1OE_D0_RNAseq Cdx2_ES_ALKBH1OE_D2_RNAseq Cdx2_ES_ALKBH1OE_D4_RNAseq Cdx2_ES_ALKBH1OE_D5_RNAseq Cdx2_ES_ALKBH1OE_D6_RNAseq

Cdx2_ES_SATB1CTRL_D0_RNAseq
Cdx2_ES_SATB1CTRL_D4_RNAseq
Cdx2_ES_SATB1KO_clone1_D0_RNAseq
Cdx2_ES_SATB1KO_clone1_D4_RNAseq
Cdx2_ES_SATB1KO_clone2_D0_RNAseq
Cdx2_ES_SATB1KO_clone2_D4_RNAseq
WT-MFI-E10.5-1
WT-MFI-E10.5-2
Alkbh1KO-MFI-E10.5-1
Alkbh1KO-MFI-E10.5-2

RNA-seq processed files:
Alkbh1_WT_OE_gene_expression_cuffdiff.txt.gz
Satb1_CT_KO_gene_expression_cuffdiff.txt.gz

ATAC-seq raw fastq files:
D0_WT_ATAC
D0_OE_ATAC
D3_WT_ATAC
D3_OE_ATAC
D4_WT_ATAC
D4_OE_ATAC
D5_WT_ATAC
D5_OE_ATAC

ATAC-seq processed files:
D5_WT_ATAC.narrowPeak.gz
D5_OE_ATAC.narrowPeak.gz
D0_WT_ATAC.bw.gz
D0_OE_ATAC.bw.gz
D3_WT_ATAC.bw.gz
D3_OE_ATAC.bw.gz
D4_WT_ATAC.bw.gz
D4_OE_ATAC.bw.gz
D5_WT_ATAC.bw.gz
D5_OE_ATAC.bw.gz

MeDIP-seq raw fastq files:
Cdx2_ES_D0_N6mA_MeDIP
Cdx2_ES_D3_N6mA_MeDIP
Cdx2_ES_D4_N6mA_MeDIP
Cdx2_ES_D5_N6mA_MeDIP
Cdx2_ES_D7_N6mA_MeDIP
Cdx2_ES_D13_N6mA_MeDIP
Cdx2_ES_D0_N6mA_MeDIP_rep2
Cdx2_ES_D3_N6mA_MeDIP_rep2
Cdx2_ES_D4_N6mA_MeDIP_rep2
Cdx2_ES_D5_N6mA_MeDIP_rep2
Cdx2_ES_D7_N6mA_MeDIP_rep2
Cdx2_ES_D13_N6mA_MeDIP_rep2
Cdx2_ES_D0_N6mA_MeDIP_rep3
Cdx2_ES_D3_N6mA_MeDIP_rep3
Cdx2_ES_D4_N6mA_MeDIP_rep3
Cdx2_ES_D5_N6mA_MeDIP_rep3
Cdx2_ES_D7_N6mA_MeDIP_rep3
Cdx2_ES_D13_N6mA_MeDIP_rep3
ssIN
dsIN
ssDIP
dsDIP
ssWGAIP
dsWGAIP
IgG

MeDIP-seq processed file:
N6mA_D5vsD0_W200-G600-increased-peaks-FDR0.01.bed.gz
DIP_D0_q20uniquePeaks-W200-G600-FDR0.01-island.bed.gz
DIP_D3_q20uniquePeaks-W200-G600-FDR0.01-island.bed.gz

DIP_D5_q20uniquePeaks-W200-G600-FDR0.01-island.bed.gz
DIP_D7_q20uniquePeaks-W200-G600-FDR0.01-island.bed.gz
DIP_D13_q20uniquePeaks-W200-G600-FDR0.01-island.bed.gz
DIP_D0-W200-G600-FDR0.01-island.bed.gz
DIP_D13-W200-G600-FDR0.01-island.bed.gz
DIP_D5-W200-G600-FDR0.01-island.bed.gz
DIP_D3-W200-G600-FDR0.01-island.bed.gz
DIP_D7-W200-G600-FDR0.01-island.bed.gz

ChIP-seq raw fastq files:
Satb1_ChIP_WT
Satb1_ChIP_ALKBH1_OE
Satb1_ChIP_D5_WT
Input

ChIP-seq processed files:
Satb1_OEChIP.bw.gz
Satb1_WTChIP.bw.gz
Satb1_Alkbh1OE_q20-W200-G600-FDR0.01-island.bed.gz
Satb1_Alkbh1OEvsWT_q20-W200-G600-increased-islands.bed.gz
Satb1_D5WT_q20-W200-G600-FDR0.01-island.bed.gz
Satb1_WT_D5vsD0_q20-W200-G600-decreased-islands.bed.gz
Satb1_WT_q20-W200-G600-FDR0.01-island.bed.gz

Genome browser session
(e.g. UCSC)

Methodology

Replicates

N6-mA MeDIP: three replicates
Satb1 ChIP: one replicate
ATAC-seq: one replicate per time point
RNA-seq (Cdx2_ES): one replicate
RNA-seq (maternal fetal interface): two biological replicates per genotype
ssDNA-seq: one sample and one input control per cell line

Sequencing depth

N6-mA MeDIP: samples were processed as 100bp paired-end sequences with about 20-30 million reads per sample. Control samples on single-strand/double-strand inputs/IPs, whole genome amplifications were sequenced for 10-20 million reads. IgG was sequenced for 1 million reads.
RNA-seq: ALKBH1 samples were sequenced as 50bp single-end, SATB1 as 100bp paired-end, and mouse maternal-fetal interface (MFI) samples as 200bp paired-end with 15-50 million reads.
Satb1 ChIP-seq: samples were sequenced as 50bp single-end with 15-35 million reads.
ATAC-seq: samples were sequenced as 100bp paired-end with 40-65 million reads.
ssDNA-seq: samples were sequenced as 100bp paired-end with 110-170 million reads for samples and 30-60 million reads for input controls.

Antibodies

DIP-seq: anti-N6-mA (Synaptic Systems, Cat#202003, Polyclonal, 5 µg/sample)
ChIP-seq: anti-SATB1 (Abcam, Cat#Ab109122, Clone# EPR3951, 10 µL/sample)
ATAC-seq, RNA-seq, ssDNA-seq: no antibodies used

Peak calling parameters

DIP-seq: reads were mapped to mouse genome mm9 using Bowtie2. Peaks were determined by SICER v 1.1 using single-strand DNA as input. Window size 200, gap 600, FDR 0.01 and otherwise default parameters were used during peak calling.
ssDNA-seq: Sequences were mapped to mm9 using Bowtie2, PCR duplicates removed by Picard, and analyzed by SICERpy, a python-based implementation of the SICER peak caller. Peaks were determined using an FDR cutoff of 0.01, window 200, gap 600 and otherwise default parameters (sicer.py -t sample.bam -c input.bam -w 200 -g 3).

Data quality

DIP-seq: Specificity of MeDIP experiments was determined by antibody pull down against whole genome amplification and IgG controls. High overlap was observed among replicates. Uniquely mapped N6-mA peaks were identified by keeping reads with map quality (MapQ) scores above 20 and with properly oriented read mates. Peak quality was controlled with FDR < 0.01.
ssDNA-seq: Uniquely mapped reads were kept with map quality (MapQ) scores above 10. Peak quality was controlled with FDR < 0.01.
ATAC-seq: Peak quality was controlled with q<0.05. Peaks overlapped with Encode blacklist regions were removed for heatmap generation.

Software

RNA-seq: STAR, Cufflinks, HTSeq, DESeq2
ATAC-seq: Bowtie2, MACS2, deepTools, USeq
DIP-seq and ChIPseq: Bowtie2, BEDTools, SAMtools, SICERv1.1, USeq, deepTools
ssDNA-seq: Bowtie2, Picard, BEDTools, SAMtools, SICERpy, USeq, deepTools, SalmonTE.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were collected using trypsin treatment, 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.
Instrument	BD LSR II
Software	FlowJo
Cell population abundance	Sorting was not used.
Gating strategy	SSC/FSC gating strategy (SSC-A/FSC-A--- SSC-W/SSC-H--- FSC-W/FSC-H) allows single cell analysis. The DAPI positive population is isolated from negative population (control samples without DAPI staining).

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.