LETTER

The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape

Daniel N. Weinberg[1,12], Simon Papillon-Cavanagh[2,12], Haifen Chen[2], Yuan Yue[3], Xiao Chen[4,5], Kartik N. Rajagopalan[6], Cynthia Horth[2], John T. McGuire[4,5], Xinjing Xu[4,5], Hamid Nikbakht[2], Agata E. Lemiesz[1], Dylan M. Marchione[7,8], Matthew R. Marunde[9], Matthew J. Meiners[9], Marcus A. Cheek[9], Michael-Christopher Keogh[9], Eric Bareke[2], Anissa Djedid[2], Ashot S. Harutyunyan[2], Nada Jabado[2,10,11], Benjamin A. Garcia[7,8], Haitao Li[3], C. David Allis[1,13]*, Jacek Majewski[2,13]* & Chao Lu[4,5,13]*

Enzymes that catalyse CpG methylation in DNA, including the DNA methyltransferases 1 (DNMT1), 3A (DNMT3A) and 3B (DNMT3B), are indispensable for mammalian tissue development and homeostasis[1–4]. They are also implicated in human developmental disorders and cancers[5–8], supporting the critical role of DNA methylation in the specification and maintenance of cell fate. Previous studies have suggested that post-translational modifications of histones are involved in specifying patterns of DNA methyltransferase localization and DNA methylation at promoters and actively transcribed gene bodies[9–11]. However, the mechanisms that control the establishment and maintenance of intergenic DNA methylation remain poorly understood. Tatton–Brown–Rahman syndrome (TBRS) is a childhood overgrowth disorder that is defined by germline mutations in *DNMT3A*. TBRS shares clinical features with Sotos syndrome (which is caused by haploinsufficiency of *NSD1*, a histone methyltransferase that catalyses the dimethylation of histone H3 at K36 (H3K36me2)[8,12,13]), which suggests that there is a mechanistic link between these two diseases. Here we report that NSD1-mediated H3K36me2 is required for the recruitment of DNMT3A and maintenance of DNA methylation at intergenic regions. Genome-wide analysis shows that the binding and activity of DNMT3A colocalize with H3K36me2 at non-coding regions of euchromatin. Genetic ablation of *Nsd1* and its paralogue *Nsd2* in mouse cells results in a redistribution of DNMT3A to H3K36me3-modified gene bodies and a reduction in the methylation of intergenic DNA. Blood samples from patients with Sotos syndrome and *NSD1*-mutant tumours also exhibit hypomethylation of intergenic DNA. The PWWP domain of DNMT3A shows dual recognition of H3K36me2 and H3K36me3 in vitro, with a higher binding affinity towards H3K36me2 that is abrogated by TBRS-derived missense mutations. Together, our study reveals a *trans*-chromatin regulatory pathway that connects aberrant intergenic CpG methylation to human neoplastic and developmental overgrowth.

To characterize the role of post-translational modifications of histones in regulating DNA methyltransferase (DNMT) targeting and CpG methylation, we profiled the genome-wide distribution patterns of histone post-translational modifications (H3K36me3, H3K36me2, H3K27me3, H3K27ac (acetylation of H3 at K27), H3K9me3 and H3K4me1), DNMT3A and DNMT3B using chromatin immunoprecipitation followed by sequencing (ChIP–seq) in C3H10T1/2 mouse mesenchymal stem cells (MSCs). We also measured levels of CpG methylation at base-pair resolution using whole-genome bisulfite

sequencing (WGBS) at high coverage (840 million reads; 45× coverage). Enriched regions of H3K36me2 and H3K36me3 were in close proximity to one another, and together defined domains greater than 1 Mb in size that were largely exclusive from domains demarcated by H3K9me3 and H3K27me3 (Fig. 1a). H3K36me2 levels positively correlated with CpG methylation, and on the basis of this histone post-translational modification, together with H3K27me3 and H3K9me3, the genome could be partitioned into regions of high (75%; H3K36me2), intermediate (50%; H3K27me3) and low (30%; H3K9me3) levels of CpG methylation (Fig. 1a, Extended Data Fig. 1a, b). Furthermore, DNMT3A and DNMT3B ChIP–seq reads were predominantly observed in H3K36me2 and H3K36me3 domains (Fig. 1a, Extended Data Fig. 1b), which suggests that the high levels of CpG methylation in these regions are—at least in part—a result of the preferential recruitment of DNMT3A and DNMT3B. Additional chromatin features that were associated with these domains of high levels of H3K36 and CpG methylation included higher levels of H3K27ac and H3K4me1, as well as increased gene expression—consistent with previous reports that linked H3K36me2 and H3K36me3 to active gene transcription[14,15] (Fig. 1b, Extended Data Fig. 1c). Moreover, the regions with high levels of H3K36 and CpG methylation corresponded to compartments and topologically associated domains derived from chromatin conformation capture (Hi-C) studies of mouse myoblasts[16] (Extended Data Fig. 1d), suggesting that these modifications together define transcriptionally active euchromatin that is spatially segregated from constitutive and facultative heterochromatin.

Our observations are consistent with previous reports that suggest that H3K36me3 mediates the targeting of DNMT3B activity[10,11]. However, as this interaction is confined to gene bodies, we hypothesized that an additional chromatin *trans*-regulatory pathway may act in parallel to facilitate CpG methylation at euchromatic regions. On closer examination, we observed that, whereas H3K36me3 exhibited characteristic enrichment within gene bodies, H3K36me2 showed a more-diffuse distribution that encompassed both genic and intergenic regions (Fig. 1a, Extended Data Fig. 2a, b). Within actively transcribed genes, H3K36me2 covered areas downstream of the transcriptional start site through the first intron, followed by a marked switch to H3K36me3 after the first splice junction (Extended Data Fig. 2a–c). Notably, whereas DNMT3B was enriched within gene bodies that contained H3K36me3 (Extended Data Fig. 2d), the localization of DNMT3A mimicked the distribution of H3K36me2 and spanned broad intergenic regions that did not exhibit discernible levels of H3K36me3 (Fig. 1a, Extended Data Fig. 2a). The contrast between the targeting

[1]Laboratory of Chromatin Biology and Epigenetics, The Rockefeller University, New York, NY, USA. [2]Department of Human Genetics, McGill University, Montreal, Quebec, Canada. [3]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. [4]Department of Genetics and Development, Columbia University Irving Medical Center, New York, NY, USA. [5]Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY, USA. [6]Division of Pulmonary, Allergy, and Critical Care Medicine, Columbia University Irving Medical Center, New York, NY, USA. [7]Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [8]Penn Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [9]EpiCypher Inc, Durham, NC, USA. [10]Department of Pediatrics, McGill University, Montreal, Quebec, Canada. [11]Research Institute of the McGill University Health Center, Montreal, Quebec, Canada. [12]These authors contributed equally: Daniel N. Weinberg, Simon Papillon-Cavanagh. [13]These authors jointly supervised this work: C. David Allis, Jacek Majewski, Chao Lu. *e-mail: alliscd@rockefeller.edu; jacek.majewski@mcgill.ca; cl3684@cumc.columbia.edu
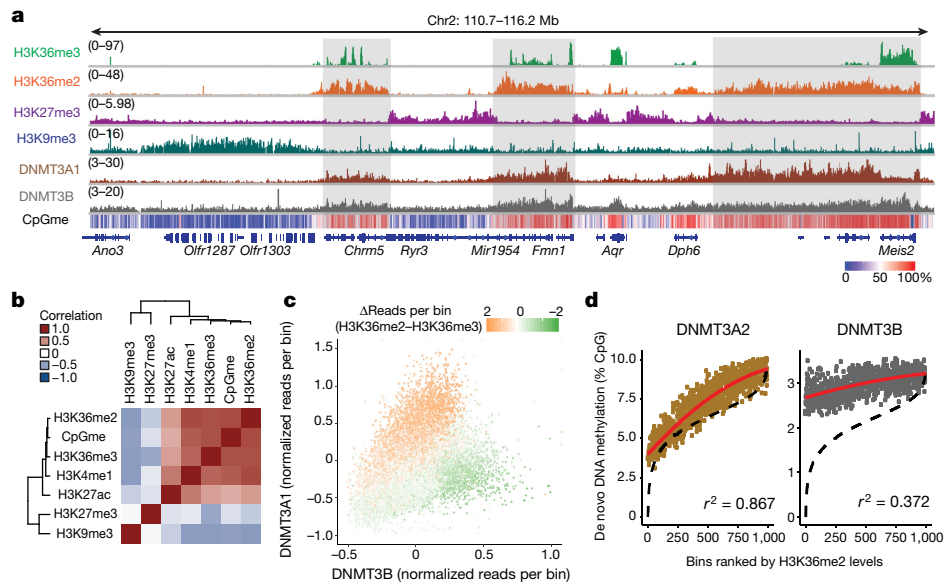
Fig. 1 | Genome-wide colocalization of DNMT3A, CpG methylation and H3K36me2. a, Genome browser representation of ChIP–seq normalized reads for H3K36me3, H3K36me2, H3K27me3, H3K9me3, DNMT3A1 and DNMT3B in mouse MSCs at chromosome 2 (Chr2): 110.7–116.2 Mb. The levels of CpG methylation (CpGme) are depicted as a heat map (blue, low; white, intermediate; red, high) and genes (from the RefSeq database) are annotated at the bottom. The shaded areas indicate H3K36me2-enriched genomic regions. For H3K36me3, H3K36me2 and DNMT3A, data are representative of two independent ChIP–seq experiments; for H3K27me3, H3K9me3 and DNMT3B, ChIP–seq was performed once and an independent ChIP was performed in which genomic regions of selective enrichment and depletion were confirmed by quantitative PCR (qPCR). WGBS was performed once. **b**, Heat map showing pairwise Pearson correlation of 10-kb bins ($n = 246,285$) between profiles of DNMT3A and DNMT3B could be reconciled by analysing the genome-wide presence of H3K36me2 and H3K36me3. DNMT3A was selectively enriched over DNMT3B at genomic regions with high levels of H3K36me2 compared to H3K36me3 (Fig. 1c).

H3K36me2, H3K36me3, H3K27me3, H3K27ac, H3K9me3, H3K4me1 and CpG methylation. **c**, ChIP–seq normalized reads per 100-kb bin for DNMT3A1 (y axis) and DNMT3B (x axis) in mouse MSCs ($n = 25,687$). Bins (dots) are colour-coded on the basis of differences between H3K36me3 and H3K36me2 ChIP–seq reads, to show selective enrichment for H3K36me2 (orange) or H3K36me3 (green). **d**, De novo methylation per bin by DNMT3A2 (brown) or DNMT3B (grey) after reintroduction into *Dnmt1*, *Dnmt3a* and *Dnmt3b* triple-knockout mouse ES cells relative to H3K36me2 ($n = 1,000$). To generate bins, 1-kb genomic tiles ($n = 2,462,755$) were ranked by H3K36me2 enrichment in ES cells and grouped into 1,000 bins, ordered by rank (2,463 tiles per group). The dashed lines indicate H3K36me2 enrichment per bin. Goodness of fit was computed using a quadratic model (red lines).

profiles of DNMT3A and DNMT3B could be reconciled by analysing the genome-wide presence of H3K36me2 and H3K36me3. DNMT3A was selectively enriched over DNMT3B at genomic regions with high levels of H3K36me2 compared to H3K36me3 (Fig. 1c).

To confirm our findings in another context, we next determined whether H3K36me2 had a similarly important role in establishing the genomic methylation landscape in mouse embryonic stem (ES) cells. We stably expressed haemagglutinin (HA)-tagged DNMT3A2 (the predominant isoform of DNMT3A) in parental ES cells and cells in which *Dnmt3a* had been genetically ablated using single-guide (sg) RNA (sgDnmt3a). When expressed at close to physiological levels, ChIP–seq of HA-tagged DNMT3A revealed that patterns of genome-wide localization were similar between parental and sgDnmt3a cells (Extended Data Fig. 3a, b). We compared these ChIP–seq profiles with the ChIP–seq profiles of H3K36me2 and H3K36me3, and observed a selective co-enrichment between H3K36me2–DNMT3A and H3K36me3–DNMT3B—although in general the distribution patterns of DNMT3A and DNMT3B were more positively correlated in ES cells than MSCs (Extended Data Fig. 3c, d). Notably, the de novo methylation activity of DNMT3A—which was measured by the levels of CpG methylation following the reintroduction of DNMT3A2 into *Dnmt1*, *Dnmt3a* and *Dnmt3b* triple-knockout ES cells (which lack virtually all CpG methylation)[10]—tracked with levels of H3K36me2 (Fig. 1d). This contrasts with a previous observation that the de novo activity of DNMT3B correlates with H3K36me3[10] (replicated in Extended Data Fig. 3e), and suggests that intergenic targeting of DNMT3A and gene-body targeting of DNMT3B together facilitate the establishment of CpG methylation in euchromatin.

Enzymes of the NSD family of histone methyltransferases have previously been shown to catalyse H3K36me2 at intergenic regions[17]. In mouse ES cells, NSD1 is the major NSD enzyme expressed, whereas in MSCs both NSD1 and NSD2 are expressed. We therefore used sgRNA to genetically ablate both *Nsd1* and *Nsd2* (sgNsd1/2) in MSCs or *Nsd1* alone (sgNsd1) in ES cells using CRISPR–Cas9 (Extended Data Fig. 4a, b). *Setd2* was disrupted separately as a control (sgSetd2) (Extended Data Fig. 4a). sgNsd1/2 and sgSetd2 MSCs showed marked and specific depletion of H3K36me2 and H3K36me3, respectively, which suggests that enzymes of the NSD family and SETD2 function distinctly at the chromatin level (Extended Data Fig. 4c–e). ChIP–seq with spike-in *Drosophila* chromatin as an exogenous reference control[18] was used to quantitatively analyse genome-wide changes between parental cells and those depleted in H3K36me2 or H3K36me3 (Extended Data Fig. 4d, f). We observed a global reduction in H3K36me2 in sgNsd1/2 MSCs and sgNsd1 ES cells, most prominently at intergenic regions (Extended Data Fig. 4g, h). By contrast, levels of H3K36me3 were unperturbed and reductions in genic H3K36me2 were relatively modest (Extended Data Fig. 4d–h), potentially owing to the activity of other H3K36 methyltransferases such as NSD3 or ASH1L.

We next profiled the genome-wide binding patterns of HA-tagged DNMT3A1 (for MSCs) and DNMT3A2 (for ES cells) and DNA methylation in H3K36me2-depleted mouse cells (Extended Data Fig. 5a, b). The recruitment of DNMT3A to intergenic regions was abrogated by deletion of enzymes of the NSD family in both ES cells and MSCs (Extended Data Fig. 5c, d). Genome-wide analysis revealed that loss of DNMT3A targeting at H3K36me2-depleted regions coincided with reduced CpG methylation (Fig. 2a, c). As predicted by the colocalization between decreased H3K36me2 and decreased CpG methylation (Extended Data Fig. 5e, f) and the disproportional intergenic enrichment of H3K36me2, CpG hypomethylation in sgNsd1/2 MSCs and sgNsd1 ES cells predominantly affected intergenic regions (Fig. 2b, d). Reintroduction of wild-type NSD1, but not a catalytic mutant form (C2023A), was able to rescue global and intergenic levels of H3K36me2 (Fig. 2e, Extended Data Figs. 4b, 5g). Notably, intergenic localization of DNMT3A could be recovered only through expression of wild-type
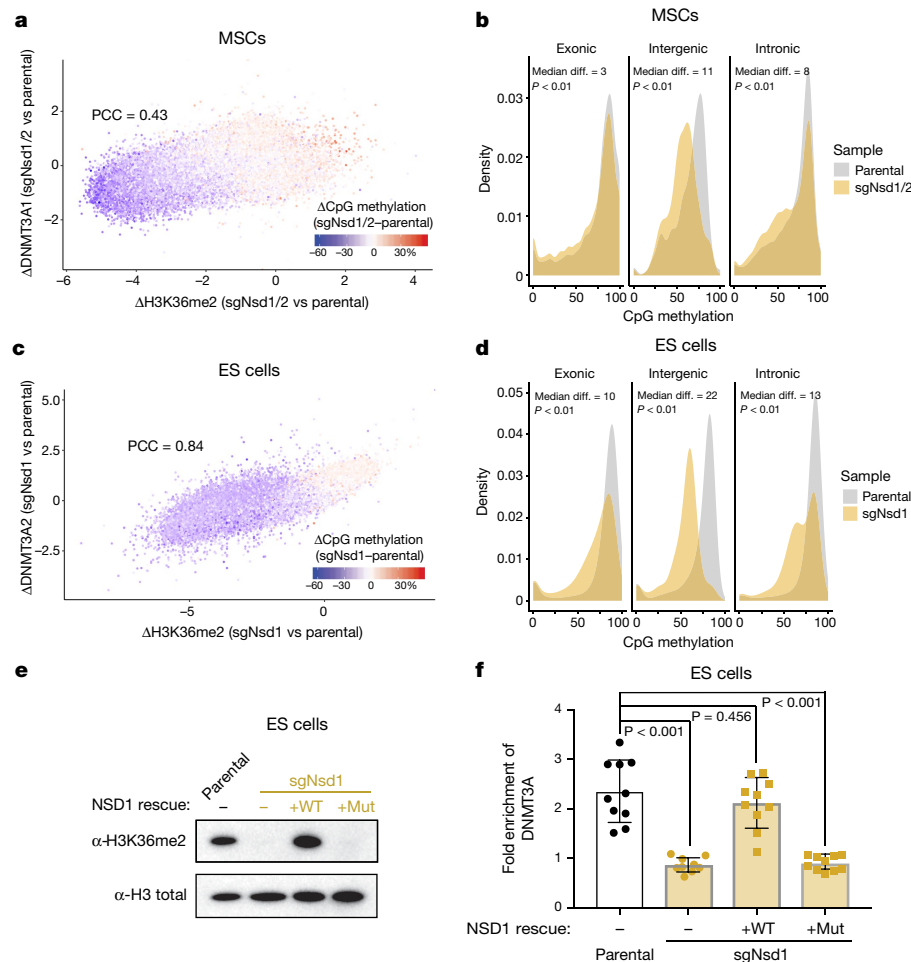
Fig. 2 | NSD1-mediated H3K36me2 is required for intergenic DNMT3A localization and CpG methylation. **a**, Difference in ChIP–seq normalized reads of DNMT3A1 between parental and sgNsd1/2 mouse MSCs relative to that of H3K36me2 for 10-kb non-overlapping bins ($n = 246,285$). Bins (dots) are colour-coded on the basis of the change in average CpG methylation, to show the loss (blue) or gain (red) of CpG methylation in sgNsd1/2 cells. Pearson's correlation coefficient (PCC) is indicated. **b**, Histograms for CpG methylation at intergenic ($n = 1,165$), exonic ($n = 13,601$) and intronic ($n = 12,364$) regions for parental (grey) and sgNsd1/2 (orange) MSCs. P values were determined by Wilcoxon's rank-sum test (two-sided). **c**, Difference in ChIP–seq normalized reads of DNMT3A2 between parental and sgNsd1 mouse ES cells relative to of H3K36me2 for 10-kb non-overlapping bins ($n = 246,285$). Bins (dots) are colour-coded on the basis of the change in average CpG methylation, to show the loss (blue) or gain (red) of CpG methylation in sgNsd1 cells.

Pearson's correlation coefficient is indicated. **d**, Histograms for CpG methylation at intergenic ($n = 1,165$), exonic ($n = 13,601$) and intronic ($n = 12,364$) regions for parental (grey) and sgNsd1 (orange) ES cells. P values were determined by Wilcoxon's rank-sum test (two-sided). **e**, Immunoblots of lysates generated from parental and sgNsd1 ES cells for H3K36me2. Total H3 was used as a loading control. sgNsd1 cells were rescued with ectopic expression of wild-type (WT) NSD1 or the C2023A catalytic mutant (Mut). Data are representative of two independent experiments. **f**, Fold enrichment of DNMT3A at various H3K36me2-enriched versus H3K36me2-depleted intergenic regions in parental (black) and sgNsd1 (orange) ES cells rescued with ectopic expression of wild-type or C2023A-mutant NSD1, measured by ChIP–qPCR. Each data point represents a genomic locus ($n = 10$). Data are mean ± s.d. P values were determined by one-way analysis of variance (ANOVA). For gel source data, see Supplementary Fig. 1.

NSD1 (Fig. 2f), which indicates that there is a specific requirement for NSD1-catalysed H3K36me2 in the recruitment of DNMT3A to non-coding euchromatic regions. Thus, we conclude that NSD1 has an essential and specific role in directing DNMT3A to intergenic regions and maintaining CpG methylation at these sites.

We next investigated the mechanisms that underlie the specificity of DNMT3A recruitment by H3K36me2. Both DNMT3A and DNMT3B contain PWWP 'reader' domains that can interact with H3K36 methylation in vitro[10,19]. To test whether the PWWP domain of DNMT3A shows specificity towards different valences of H3K36 methylation, we examined interactions between the purified DNMT3A PWWP domain (DNMT3A$_{PWWP}$) and a panel of semi-synthetic nucleosomes (Extended Data Fig. 6a). DNMT3A$_{PWWP}$ bound with highest affinity to nucleosomes modified with H3K36me2, followed by those modified with H3K36me3, but did not bind to nucleosomes modified with H3K36me1 or any valence at H3K4, H3K9, H3K27 or H4K20 (Fig. 3a, Extended Data Fig. 6b). Quantitative isothermal titration calorimetry

(ITC) assays with either H3.1 or H3.3 K36-modified peptides and DNMT3A$_{PWWP}$ further supported its preferential recognition of H3K36me2 and H3K36me3 (Extended Data Fig. 6c, d). These results indicate that DNMT3A$_{PWWP}$ recognizes both methylation states but exhibits greater affinity for H3K36me2.

We speculated that such a binding preference, combined with the relative abundance of H3K36me2 (Extended Data Fig. 4e), contributes to the favoured enrichment of DNMT3A at intergenic regions. Thus, we tested the hypothesis that DNMT3A will colocalize with H3K36me3 when H3K36me2 is depleted. Indeed, we noted a marked redistribution of DNMT3A to gene bodies in sgNsd1/2 MSCs (Fig. 3b). This could not be explained through the interaction of DNMT3A with residual genic H3K36me2, as DNMT3A now exhibited a binding profile reminiscent of DNMT3B, in which signal is enriched towards the 3′ end of the gene body (Extended Data Fig. 7a). Furthermore, we found that DNMT3A was depleted at regions of high H3K36me3 in parental MSCs, but enriched at these same regions in sgNsd1/2 cells (Fig. 3b, c).
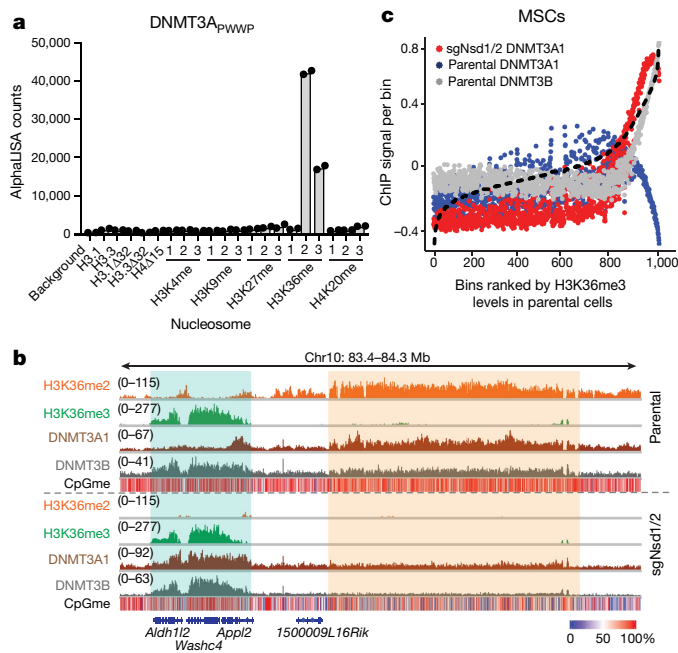
Fig. 3 | Preferential recognition of H3K36me2 and H3K36me3 by the PWWP domain of DNMT3A facilitates its localization. a, AlphaLISA counts for the interaction of GST-tagged DNMT3A$_{PWWP}$ with semi-synthetic modified nucleosomes. Data are from two replicates. Tail-less nucleosome controls include deletion of residues 1–32 in H3.1 (H3.1Δ32), residues 1–32 in H3.3 (H3.3Δ32) and residues 1–15 in H4 (H4Δ15). **b**, Genome browser representation of ChIP–seq normalized reads for H3K36me2, H3K36me3, DNMT3A1 and DNMT3B in parental and sgNsd1/2 mouse MSCs at Chr10: 83.4–84.3 Mb. The levels of CpG methylation (CpGme) are depicted as a heat map (blue, low; white, intermediate; red, high) and RefSeq genes are annotated at the bottom. The shaded areas indicate H3K36me2-enriched intergenic regions (orange) and H3K36me3-enriched genic regions (green) in parental cells. After depletion of H3K36me2, DNMT3A is retargeted to H3K36me3-enriched gene bodies (for example, *Aldh1l2* and *Appl2*). For H3K36me2, H3K36me3 and DNMT3A1 in parental cells, data are representative of two independent ChIP–seq experiments. For DNMT3B in parental cells, ChIP–seq was performed once and an independent ChIP was performed in which genomic regions of selective enrichment and depletion were confirmed by qPCR. DNMT3A1 and DNMT3B ChIP–seq in sgNsd1/2 cells and WGBS in both lines were performed once. **c**, ChIP–seq normalized reads per bin for DNMT3A1 (blue) and DNMT3B (grey) in parental MSCs and DNMT3A1 (red) in sgNsd1/2 MSCs relative to H3K36me3. To generate bins, 1-kb genomic tiles (*n* = 2,462,755) were ranked by H3K36me3 enrichment in MSCs and grouped into 1,000 bins, ordered by rank (2,463 tiles per group). The dashed line indicates H3K36me3 enrichment per bin.

Disruption of SETD2 catalytic activity in sgNsd1/2 cells (that is, triple-knockout cells) impaired the localization of DNMT3A to gene bodies, reflecting a requirement for H3K36me3 in genomic retargeting of DNMT3A (Extended Data Fig. 7b, c). Introduction of a point mutation (D333A) in the DNMT3A PWWP domain that impairs its binding to H3K36me2 and H3K36me3[19] abrogated the relocalization of DNMT3A to H3K36me3-enriched regions in sgNsd1/2 cells (Extended Data Fig. 7d, e). Together, these results suggest that preferential recognition of H3K36me2 and H3K36me3 by the PWWP domain guides the localization of DNMT3A across the cellular chromatin landscape.

We asked whether a similar redistribution mechanism may occur for DNMT3B. As expected from its localization pattern, the PWWP domain of DNMT3B preferentially binds to H3K36me3-modified recombinant nucleosomes in vitro (Extended Data Fig. 8a, b). Loss of gene-body localization for DNMT3B in H3K36me3-depleted sgSetd2 mouse MSCs was accompanied by a relatively modest increase in retargeting to H3K36me2-enriched intergenic regions (Extended Data

Fig. 8c, d). Overall, these data suggest that the ratio of H3K36me2 to H3K36me3 is a key determinant for specifying the targeting patterns of DNMT3A and, to a lesser extent, DNMT3B.

It has previously been reported that a subset of head and neck squamous cell carcinomas (HNSCCs) that are depleted in H3K36me2 as a result of genetic or biochemical inactivation of NSD1 show a DNA hypomethylation signature[20]. Similarly, a significant decrease in DNA methylation was observed in blood samples from patients with Sotos syndrome, which is defined by germline *NSD1* haploinsufficiency, compared with control samples[21]. To explore the relevance of our findings in the context of human disease, we profiled H3K36me2 and CpG methylation in patient-derived HNSCC cell lines that carry wild-type *NSD1* (Cal27 and Fadu) or mutant *NSD1* (SCC-4 and SKN-3). Despite the heterogeneous genetic backgrounds of these cell lines, DNA hypomethylation in *NSD1*-mutant cell lines tracked closely with genome-wide reductions in H3K36me2 compared with *NSD1*-wild-type cell lines, and occurred primarily at intergenic regions (Fig. 4a, b). We further analysed publicly available DNA methylation arrays of samples from patients with HNSCCs or Sotos syndrome. Probes that were hypomethylated in NSD1-inactivated compared to NSD1-wild-type HNSCC tumours were significantly enriched for intergenic regions (Fig. 4c). Patients with Sotos syndrome also exhibited an enrichment of hypomethylated probes at intergenic regions compared to healthy control individuals (Fig. 4c).

Missense mutations in *DNMT3A* cause TBRS—a developmental disorder that shares many clinical features with Sotos syndrome, including skeletal overgrowth, facial dysmorphism and intellectual disability[8,12,22]. We characterized point mutations within DNMT3A$_{PWWP}$ that are associated with TBRS (W297del, I310N and Y365C) (Fig. 4d). Recombinantly expressed DNMT3A$_{PWWP}$ carrying these mutations showed reduced nucleosome binding in vitro (Fig. 4e, Extended Data Fig. 9a), which suggests that the mutations may impair the recruitment of DNMT3A to chromatin. Indeed, W297del and I310N mutations reduced the association of DNMT3A with bulk chromatin (Extended Data Fig. 9b), and this was accompanied by a reduction in protein levels (consistent with a recent study[23]) (Extended Data Fig. 9c). Further examination of chromatin-bound wild-type, W297del-, I310N- or Y365C-mutant DNMT3A revealed a substantial decrease in the levels of H3K36me2-modified nucleosomes in complex with mutant DNMT3A (Extended Data Fig. 9d). Consistent with this, ChIP–seq analysis showed that the I310N mutation abrogated targeting of DNMT3A to H3K36me2 on a genome-wide level (Extended Data Fig. 9e, f). Thus, impaired recruitment of DNMT3A and reduced CpG methylation at H3K36me2-enriched intergenic euchromatic regions appear to be common features that are associated with TRBS and Sotos syndrome. In support of this notion, unsupervised hierarchical clustering of published DNA methylation array profiles[21,24] indicated that patients with Sotos syndrome—but not with Weaver syndrome, which is associated with mutations in *EZH2*[22]—have a DNA methylome that is similar to that of patients with TBRS (Extended Data Fig. 9g).

Our findings provide an enhanced framework for understanding how DNA methylation landscapes are established and maintained in the euchromatic genome. Preferential targeting of DNMT3A and DNMT3B, guided by their PWWP chromatin reader domains, directs CpG methylation to H3K36me2-enriched intergenic regions and H3K36me3-enriched gene bodies, respectively (Extended Data Fig. 9h). Simultaneously, the presence of H3K4me3 at active promoters disrupts the interaction of the ADD (ATRX–DNMT3–DNMT3L) domain of DNMT3A and DNMT3B with histone H3, thereby protecting these regions from ectopic methylation[9]. Depletion of DNA methylation at promoters is further ensured by the recruitment of several proteins that contain a CXXC domain and preferentially recognize CpG-dense promoters; these proteins include enzymes of the TET family (which are involved in active DNA demethylation) and the H3K36 demethylase KDM2B[25,26]. We propose that CpG methylation is therefore established across broad euchromatic regions through the combined actions of DNMT3A and DNMT3B, whereas focal sites of
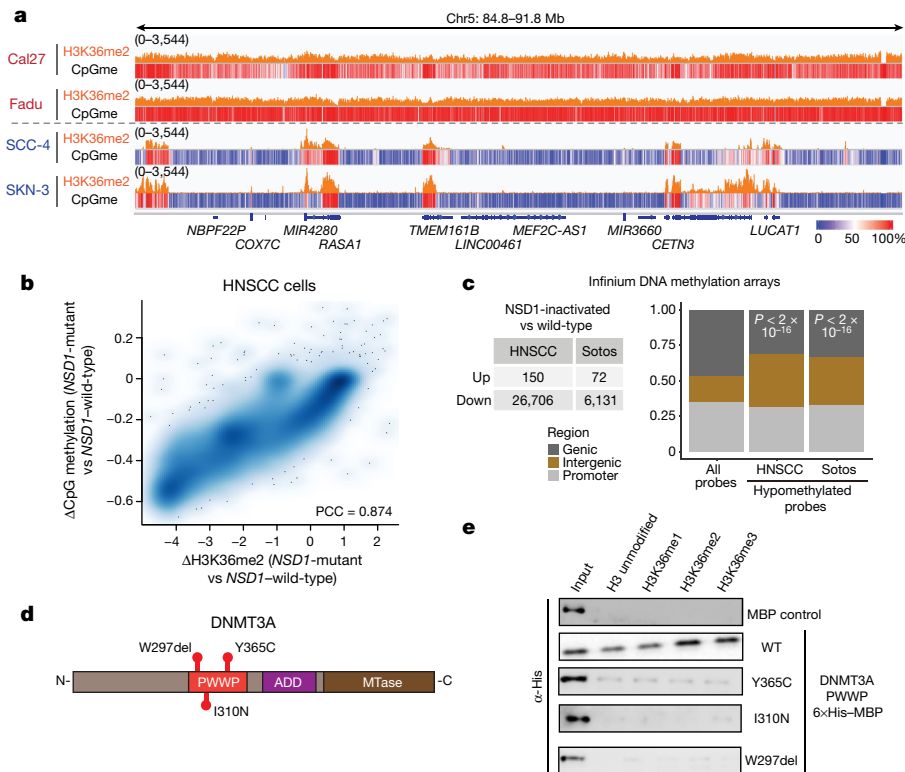
Fig. 4 | Impaired intergenic DNMT3A localization and CpG methylation in neoplastic and developmental overgrowth. a, Genome browser representation of ChIP–seq normalized reads for H3K36me2 in *NSD1*-wild-type (Cal27, Fadu) and *NSD1*-mutant (SCC-4, SKN-3) HNSCC cell lines at Chr5: 84.8–91.8 Mb. The levels of CpG methylation (CpGme) are depicted as a heat map (blue, low; white, intermediate; red, high) and RefSeq genes are annotated at the bottom. For H3K36me2 in Fadu and SKN-3 lines, data are representative of two independent ChIP–seq experiments. H3K36me2 ChIP–seq in Cal27 and SCC-4 lines and WGBS in all lines were performed once. **b**, Per cent change in averaged CpG methylation between *NSD1* wildtype and mutant HNSCC cell lines relative to changes in ChIP–seq normalized reads of H3K36me2 for 100-kb non-overlapping bins (*n* = 28,395). Pearson's correlation coefficient is indicated. **c**, Left, table summarizing the number of

upregulated and downregulated Infinium 450K DNA methylation array probes between NSD1-inactivated and NSD1-wild-type HNSCC tumour samples, and between samples from patients with Sotos syndrome and healthy control individuals. Right, bar graph showing intergenic enrichment of DNA hypomethylated probes relative to all probes (*n* = 370,898). *P* values were determined by chi-squared test. **d**, Schematic of conserved structural domains in DNMT3A with the indicated TBRS-associated mutations. **e**, Immunoblots of recombinant histidine and maltose-binding-protein (His–MBP)-tagged wild-type and mutant (Y365C, I310N or W297del) DNMT3A PWWP domains bound to H3K36-modified recombinant nucleosomes following the in vitro pull-down assay. Data are representative of two independent experiments. For gel source data, see Supplementary Fig. 1.

transcriptional initiation are spared. Our findings, however, do not exclude the possibility that DNMT1 activity is also regulated by H3K36 methylation. Structural characterization of DNMT3A and DNMT3B bound to H3K36me2 and H3K36me3 will be necessary to determine the basis of selective recognition by their PWWP domains.

Our studies demonstrate that intergenic DNA hypomethylation that results from a dysfunctional NSD1–DNMT3A *trans*-regulatory pathway represents a mechanistic link between two phenotypically overlapping human overgrowth syndromes (Extended Data Fig. 9h). This pathway is also likely to facilitate neoplastic tissue overgrowth. In mouse models of squamous cell carcinomas, loss of *Nsd1* or *Dnmt3a* promotes tumour development[27,28]. Somatic mutations in *DNMT3A* are common in acute myeloid leukaemia, and recurrent translocations that fuse *NSD1* to *NUP98* and lead to aberrant localization of NSD1 are also observed in patients with this malignancy[5,29]. Notably, germline mutations in additional epigenetic regulators, including *EZH2* and histone H1, cause developmental disorders that are similar to Sotos syndrome and TRBS[25]. Furthermore, gain-of-function mutations in *DNMT3A* that result in ectopic DNA hypermethylation at polycomb repressive domains have recently been found to cause delayed growth and microcephaly in both mice and humans[23,30]. Thus, future work is warranted to investigate the complex chromatin regulatory network that governs the DNA methylation landscape across broad euchromatic regions in the control of cellular differentiation and growth.

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/s41586-019-1534-3.

1. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
2. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
3. Wu, H. et al. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* **329**, 444–448 (2010).
4. Challen, G. A. et al. Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44**, 23–31 (2012).
5. Ley, T. J. et al. DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).
6. Klein, C. J. et al. Mutations in *DNMT1* cause hereditary sensory neuropathy with dementia and hearing loss. *Nat. Genet.* **43**, 595–600 (2011).
7. Xu, G. L. et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187–191 (1999).
8. Tatton-Brown, K. et al. Mutations in the DNA methyltransferase gene *DNMT3A* cause an overgrowth syndrome with intellectual disability. *Nat. Genet.* **46**, 385–388 (2014).

9. Ooi, S. K. T. et al. DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448**, 714–717 (2007).
10. Baubec, T. et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243–247 (2015).
11. Morselli, M. et al. In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. *eLife* **4**, e06205 (2015).
12. Kurotaki, N. et al. Haploinsufficiency of *NSD1* causes Sotos syndrome. *Nat. Genet.* **30**, 365–366 (2002).
13. Rayasam, G. V. et al. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J.* **22**, 3153–3163 (2003).
14. Rao, B., Shibata, Y., Strahl, B. D. & Lieb, J. D. Dimethylation of histone H3 at lysine 36 demarcates regulatory and nonregulatory chromatin genome-wide. *Mol. Cell. Biol.* **25**, 9447–9459 (2005).
15. Xiao, T. et al. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev.* **17**, 654–663 (2003).
16. Doynova, M. D., Markworth, J. F., Cameron-Smith, D., Vickers, M. H. & O'Sullivan, J. M. Linkages between changes in the 3D organization of the genome and transcription during myotube differentiation in vitro. *Skelet. Muscle* **7**, 5 (2017).
17. Kuo, A. J. et al. NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Mol. Cell* **44**, 609–620 (2011).
18. Orlando, D. A. et al. Quantitative ChIP-seq normalization reveals global modulation of the epigenome. *Cell Rep.* **9**, 1163–1170 (2014).
19. Dhayalan, A. et al. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* **285**, 26114–26120 (2010).
20. Papillon-Cavanagh, S. et al. Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. *Nat. Genet.* **49**, 180–185 (2017).
21. Choufani, S. et al. *NSD1* mutations generate a genome-wide DNA methylation signature. *Nat. Commun.* **6**, 10207 (2015).
22. Tatton-Brown, K. et al. Mutations in epigenetic regulation genes are a major cause of overgrowth with intellectual disability. *Am. J. Hum. Genet.* **100**, 725–736 (2017).
23. Heyn, P. et al. Gain-of-function *DNMT3A* mutations cause microcephalic dwarfism and hypermethylation of Polycomb-regulated regions. *Nat. Genet.* **51**, 96–105 (2019).
24. Jeffries, A. R. et al. Growth disrupting mutations in epigenetic regulatory molecules are associated with abnormalities of epigenetic aging. *Genome Res.* **29**, 1057–1066 (2019).
25. Williams, K., Christensen, J. & Helin, K. DNA methylation: TET proteins—guardians of CpG islands? *EMBO Rep.* **13**, 28–35 (2011).
26. Boulard, M., Edwards, J. R. & Bestor, T. H. FBXL10 protects Polycomb-bound genes from hypermethylation. *Nat. Genet.* **47**, 479–485 (2015).
27. Quintana, R. M. et al. A transposon-based analysis of gene mutations related to skin cancer development. *J. Invest. Dermatol.* **133**, 239–248 (2013).
28. Rinaldi, L. et al. Loss of Dnmt3a and Dnmt3b does not affect epidermal homeostasis but promotes squamous transformation through PPAR-γ. *eLife* **6**, e21697 (2017).
29. Jaju, R. J. et al. A novel gene, *NSD1*, is fused to *NUP98* in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia. *Blood* **98**, 1264–1267 (2001).
30. Sendžikaitė, G., Hanna, C. W., Stewart-Morgan, K. R., Ivanova, E. & Kelsey, G. A DNMT3A PWWP mutation leads to methylation of bivalent chromatin and growth retardation in mice. *Nat. Commun.* **10**, 1884 (2019).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Plasmid construction and lentivirus production for cell culture. sgRNAs directed against mouse *Dnmt3a*, *Nsd1*, *Nsd2* and *Setd2* were cloned into px458 (Addgene 48138; a gift from F. Zhang). Mouse *Dnmt3a* and *Dnmt3b* cDNA sequences from Horizon Dharmacon were cloned into pCDH-EF1-MCS-Neo and PiggyBac (pCAGGS-IRES-Neo; a gift from H. Niwa, Institute of Molecular Embryology and Genetics, Kumamoto) with an N-terminal Flag-HA epitope tag using Gibson assembly (NEB). Mouse *Nsd1* cDNA from Horizon Dharmacon was cloned into pPB-CAG-3×Flag-empty-pgk-hph (Addgene 48754; a gift from A. Smith) using Gibson assembly. Standard site-directed mutagenesis techniques were used to generate C2023A in NSD1 and the TRBS missense mutations W297del, I310N and Y365C that correspond to mouse DNMT3A residues W293, I306 and Y361 respectively. To produce lentivirus, 293T cells were transfected with the lentiviral vector and helper plasmids (psPAX2, pVSVG). Supernatant containing lentivirus was collected and filtered 48 h later for transduction.

Cell culture, CRISPR–Cas9 gene editing and generation of stable cell lines. 293T and C3H10T1/2 cells (ATCC) were cultured in Dulbecco's modified Eagle medium (DMEM; Invitrogen) with 10% fetal bovine serum (FBS; Sigma). Fadu (ATCC), SKN-3 (JCRB cell bank), Cal27 (ATCC) and SCC-4 (ATCC) cells were cultured in DMEM:F12 medium (Invitrogen) with 10% FBS. V6.5 mouse ES cells (C57BL/6 × 129S4/SvJae F_1) were maintained on gelatin-coated plates in KnockOut DMEM (Gibco) supplemented with 15% ES-cell-qualified FBS (Gemini), 0.1 mM 2-mercapoethanol, 2mM L-glutamine (Life Technologies) and LIF. *Drosophila* S2 cells were cultured in Schneider's *Drosophila* medium (Invitrogen) containing 10% heat-inactivated FBS. All cell lines tested negative for mycoplasma contamination. To generate knockout lines, mouse ES cells and C3H10T1/2 cells were transfected with sgRNA-containing px458 using Xfect mouse ES cell transfection reagent (Takara) or Lipofectamine 2000 (Invitrogen) and incubated for 48 h. Single GFP+ cells were then sorted into 96-well plates. Clones were expanded, screened for global reduction of H3K36me2 or H3K36me3 by immunoblot and individually verified by Sanger sequencing of the target loci (Supplementary Table 1). To generate transgenic C3H10T1/2 lines expressing epitope-tagged DNMT3A or DNMT3B, cells were transduced with concentrated lentivirus as previously described[20]. Transduced cells were grown under G418 selection (1,000 μg ml^{-1}) 48 h after transduction and selected for 1 week before being collected for immunoblot or ChIP–seq. To generate transgenic mouse ES cell lines, around $5 × 10^6$ cells were electroporated with PiggyBac expression vectors plus transposase (pBase) in a 3:1 ratio using the Amaxa ESC Nucleofector kit (VPH-1001, program A-023, Lonza). Cells were plated on gelatin-coated plates and grown under G418 (500 μg ml^{-1}) or hygromycin (100 μg ml^{-1}) selection 48 h after electroporation for at least 2 passages before being collected for immunoblot or ChIP–seq analysis.

Immunoblotting. Fractionated or whole-cell lysates were resolved by SDS–PAGE, transferred to a nitrocellulose or PVDF membrane, blocked in 5% non-fat milk in PBS plus 0.5% Tween-20, probed with primary antibodies and detected with horseradish peroxidase-conjugated anti-rabbit or anti-mouse secondary antibodies (GE Healthcare). Primary antibodies were: anti-H3K36me2 (Cell Signaling Technology, 2901), anti-H3K36me3 (Active Motif, 61101), anti-NSD1 (Abbexa, abx135901), anti-NSD2 (Millipore Sigma, MABE191), anti-SETD2 (Abcam, ab31358), anti-DNMT3A (Abcam, ab188470), anti-vinculin (Cell Signaling Technology, 13901), anti-His (ZSGB-Bio,TA-02), anti-lamin B1 (Cell Signaling Technology, 12586), anti-β-tubulin (Cell Signaling Technology, 2128), anti-β-actin (Abcam, ab8224), anti-H3 (Abcam, ab1791) and anti-HA (Biolegend, 901501). The specificities of anti-H3K36me3 and anti-H3K36me2 antibodies were validated using *Setd2* knockout and *Nsd1/2* double-knockout cell lines.

Histone acid extraction, histone derivatization and analysis of post-translational modifications by nano-LC–MS. Cells were lysed in nuclear isolation buffer (15 mM Tris pH 7.5, 60 mM KCl, 15 mM NaCl, 5 mM MgCl$_2$, 1 mM CaCl$_2$, 250 mM sucrose, 10 mM sodium butyrate, 1 mM DTT, 500 μM AEBSF and 5 nM microcystin) containing 0.3% NP-40 alternative on ice for 5 min. Nuclei were pelleted and resuspended in 0.2 M H$_2$SO$_4$, followed by 1.5 h rotation at 4 °C. After centrifugation, supernatants were collected and proteins were precipitated in 33% TCA overnight on ice, washed with acetone and resuspended in deionized water. Acid-extracted histones (5–10 μg) were resuspended in 100 mM ammonium bicarbonate (pH 8), derivatized using propionic anhydride and digested with trypsin as previously described[31]. After a second round of propionylation, the resulting histone peptides were desalted using C18 Stage Tips, dried using a centrifugal evaporator and reconstituted using 0.1% formic acid in preparation for liquid chromatography–mass spectrometry (LC–MS) analysis. Nanoflow liquid chromatography was performed using a Thermo Fisher Scientific. Easy nLC 1000 equipped with a 75 μm × 20-cm col-

umn packed in-house using Reprosil-Pur C18-AQ (3 μm; Dr. Maisch). Buffer A was 0.1% formic acid and Buffer B was 0.1% formic acid in 80% acetonitrile. Peptides were resolved using a two-step linear gradient from 5% B to 33% B over 45 min, then from 33% B to 90% B over 10 min at a flow rate of 30 nl min^{-1}. The HPLC was coupled online to an Orbitrap Elite mass spectrometer operating in the positive mode using a Nanospray Flex Ion Source (Thermo Fisher Scientific) at 2.3 kV. Two full mass spectrometry scans (*m/z* 300–1,100) were acquired in the Orbitrap mass analyser with a resolution of 120,000 (at 200 *m/z*) every 8 data-independent acquisition tandem mass spectrometry (MS/MS) events, using isolation windows of 50 *m/z* each (for example, 300–350, 350–400…650–700). MS/MS spectra were acquired in the ion trap operating in normal mode. Fragmentation was performed using collision-induced dissociation in the ion trap mass analyser with a normalized collision energy of 35. The automatic gain control target and maximum injection time were $5 × 10^5$ and 50 ms for the full mass spectrometry scan, and $3 × 10^4$ and 50 ms for the MS/MS scan, respectively. Raw files were analysed using EpiProfile 2.0[32]. The area for each modification state of a peptide was normalized against the total signal for that peptide to give the relative abundance of the histone modification.

dCypher nucleosome-binding assays. A total of 5 μl of 250 nM GST-tagged DNMT3A (Active Motif, 31541) was incubated with 5 μl of 10 nM biotinylated nucleosomes (EpiCypher, 16-9001) for 30 min at room temperature in binding buffer (20 mM HEPES pH 7.5, 250 mM NaCl, 0.01% BSA, 0.01% NP-40, 1 mM DTT) in a 384-well plate. A mixture of 10 μl of 2.5 μg ml^{-1} glutathione acceptor beads (PerkinElmer, AL109M) and 5 μg ml^{-1} streptavidin donor beads (PerkinElmer, 6760002) was prepared in bead buffer (20 mM Hepes pH 7.5, 250 mM NaCl, 0.01% BSA, 0.01% NP-40) and added to each well. The plate was incubated at room temperature in subdued lighting for 60 min and the AlphaLISA signal was measured on a PerkinElmer 2104 EnVision (680-nm laser excitation, 570-nm emission filter ± 50-nm bandwidth). Each binding interaction was performed in duplicate.

Nucleosome pull-down assays. Biotinylated nucleosomes (1 μg) (Epicypher; 16-0006, 16-0322, 16-0319, 16-0320) were immobilized on 10 μl BSA-blocked Dynabeads MyOne Streptavidin C1 (Invitrogen, 65001) in 20 mM Tris pH 7.5, 100 mM NaCl, 0.1% NP40 and 1 mM DTT buffer plus protease inhibitors (Selleck, K4000) for 30 min on ice. The beads were then incubated with 1 μM or 10 μM His-tagged DNMT3A$_{PWWP}$ or DNMT3B$_{PWWP}$ proteins at 4 °C for 4 h. After 5 washes with 1 ml binding buffer, the protein-bound beads were pooled and the pull-down products were analysed by western blotting with anti-His antibody (ZSGB-Bio, TA-02).

The recombinant DNMT3A$_{PWWP}$ and DNMT3B$_{PWWP}$ proteins for nucleosome pull-down assays were prepared using the following procedures. The DNMT3A$_{PWWP}$ domains (residues 278–427) of wild-type and mutant human DNMT3A were cloned into a modified pRSFDuet vector with an N-terminal 6×His–MBP tag. Recombinant proteins were overexpressed in *Escherichia coli* strain BL21 (DE3) induced by 0.2 mM isopropyl-1-thio-D-galactopyranoside (IPTG) at 16 °C overnight. Cells were collected and resuspended in buffer containing 20 mM Tris pH 8.0, 300 mM NaCl and 5% glycerol. After cell lysis and centrifugation, proteins were purified by Dextrin Beads 6FF (Smart-Lifesciences, SA026100) and eluted with 20 mM maltose elution buffer. The eluted proteins were further purified by gel filtration Superdex 200 10/300 GL (GE Healthcare). Human DNMT3B$_{PWWP}$ (residues 206–355) was cloned into a pET28b vector and then expressed at 16 °C overnight. The protein was purified using a HisTrap column and Superdex 75 10/300 (GE Healthcare). All gel-filtration buffer contained 20 mM Tris-HCl pH 7.5 and 100 mM NaCl for the nucleosome pull-down assay. Protein purity was analysed by SDS–PAGE and western blot. Protein concentrations were determined by UV spectroscopic measurement at 280 nm.

Recombinant DNMT3A$_{PWWP}$ expression and isothermal titration calorimetry assay. The PWWP domain (residues 278–427) of human DNMT3A (DNMT3A$_{PWWP}$) was sub-cloned into a modified pET28b vector with an N-terminal 6×His-GB1 tag. Recombinant protein was overexpressed in *E. coli* BL21 (DE3) induced by 0.2 mM IPTG at 16 °C overnight. Cells were collected and resuspended in buffer containing 20 mM Tris pH 7.5, 300 mM NaCl, 10 mM imidazole and 5% glycerol. After cell lysis and centrifugation, the His-GB1-tagged DNMT3A$_{PWWP}$ supernatant was purified using a HisTrap column (GE Healthcare). After PreScission protease digestion to cleave the 6×His-GB1 tag, DNMT3A$_{PWWP}$ was further purified on a HiTrap Heparin HP Column (GE Healthcare). The protein sample was finally polished over a Superdex75 10/300 GL (GE Healthcare) column in buffer containing 20 mM Tris pH 7.5 and 150 mM NaCl.

Isothermal titration calorimetry (ITC) was performed at 25 °C with a MicroCal PEAQ-ITC (Malvern Panalytical). DNMT3A$_{PWWP}$ was exchanged to ITC buffer containing 150 mM NaCl and 20 mM HEPES-Na pH 7.4 by gel filtration. Lyophilized peptide was directly dissolved in ITC buffer for titration. Concentrations of DNMT3A$_{PWWP}$ and H3.3$_{1-42}$ or H3.1$_{1-42}$ peptides were 0.2 mM and 2 mM, respectively. Protein concentration was determined by the UV absorb-

ance at 280 nm. Peptides were quantified by weighing 20 mg and then dissolved in water. Peptide solutions were further aliquoted and freeze-dried for individual use. Acquired ITC data were analysed by Origin 8.0 (GE Healthcare) using the 'one set of binding sites' fitting model.

Chromatin fractionation. C3H10T1/2 cells were washed with PBS and lysed in Buffer A (10 mM HEPES, 10 mM KCl, 1.5 mM MgCl$_2$, 0.34 M sucrose, 10% glycerol, 0.5 mM PMSF and 0.1% Triton X-100) on ice for 8 min. Centrifugation was carried out at 1,300g at 4 °C for 5 min and the supernatant was collected (cytosolic fraction). The nuclei pellet was further lysed in Buffer B (3 mM EDTA, 0.2 mM EGTA and 0.2 mM PMSF) on ice for 30 min and centrifugation was performed (1,700g at 4 °C for 5 min) to obtain the supernatant (nuclear soluble fraction). Cytosolic and nuclear soluble fractions were combined to make the soluble fraction. The insoluble pellet was lysed in SDS sample loading buffer, boiled and sonicated to yield the chromatin fraction. The protein concentration of each fraction was measured and equal amounts analysed by immunoblot.

Chromatin immunoprecipitation. Cross-linking ChIP in mouse MSCs and ES cells was performed using approximately 2 × 10^7 cells per immunoprecipitation. Before fixation, the medium was aspirated and cells were washed once with PBS. Cells were cross-linked directly on the plate using 1% paraformaldehyde for 5 min at room temperature with gentle shaking and glycine added to quench (final concentration 125 mM, incubated for 5 min at room temperature). Cells were washed once with cold PBS, scraped off the plates and pelleted. To obtain a soluble chromatin extract, cells were resuspended in 1 ml LB1 (50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100 and 1× complete protease inhibitor) and incubated rotating at 4 °C for 10 min. Samples were centrifuged, resuspended in 1 ml LB2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 1× complete protease inhibitor) and incubated rotating at 4 °C for 10 min. Finally, samples were centrifuged, resuspended in 1 ml LB3 (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% sodium deoxycholate, 0.5% N-lauroylsarcosine, 1% Triton X-100 and 1× complete protease inhibitor) and homogenized by passing two times through a 27-gauge needle. Chromatin extracts were sonicated for 8 min (anti-HA ChIP) or 12 min (anti-histone post-translational modification (PTM) ChIP) using a Covaris E220 focused ultrasonicator at a peak power of 140, a duty factor of 5 and 200 cycles per burst. For ChIP with reference exogenous genome (ChIP-Rx) for histone PTMs, after centrifugation soluble chromatin was spiked-in with soluble chromatin from *Drosophila* S2 cells that was similarly prepared and equivalent to 5–10% of the chromatin from mouse or human cells. The lysates were incubated with 100 μl Pierce anti-HA beads (Thermo Fisher Scientific, 88836) or with anti-H3K4me1 (Abcam, ab8895), anti-H3K9me3 (Abcam, ab8898), anti-H3K27ac (Active Motif, 39133), anti-H3K27me3 (Cell Signaling Technology, 9733), anti-H3K36me2 (Cell Signaling Technology, 2901) or anti-H3K36me3 (Active Motif, 61101) antibodies bound to 75 μl protein A or protein G Dynabeads (Invitrogen) and incubated overnight at 4 °C with 5% kept as input DNA. Magnetic beads were sequentially washed with low-salt buffer (150 mM NaCl, 0.1% SDS, 1% Triton X-100, 1 mM EDTA and 50 mM Tris-HCl), high-salt buffer (500 mM NaCl, 0.1% SDS, 1% Triton X-100, 1 mM EDTA and 50 mM Tris-HCl), LiCl buffer (150 mM LiCl, 0.5% sodium deoxycholate, 0.1% SDS, 1% Nonidet P-40, 1 mM EDTA and 50 mM Tris-HCl) and TE buffer (1 mM EDTA and 10 mM Tris-HCl). For ChIP–seq, beads were resuspended in elution buffer (1% SDS, 50 mM Tris-HCl pH 8.0, 10 mM EDTA and 200 mM NaCl) and incubated for 30 min at 65 °C. After centrifugation, the eluate was reverse-cross-linked overnight at 65 °C. The eluate was then treated with RNaseA for 1 h at 37 °C and with Proteinase K (Roche) for 1 h at 55 °C and the DNA was recovered using a Qiagen PCR purification kit. For ChIP immunoblotting, TE washes were replaced with TBS 0.1% Tween-20, and the beads were resuspended in Laemmli buffer and boiled to de-cross-link proteins before separation by SDS–PAGE. For HNSCC lines, sonication was performed on a BioRuptor UCD-300 for 60 cycles at 4 °C (10 s on followed by 20 s off, centrifuging every 15 cycles), and ChIP-Rx for H3K36me2 (Cell Signaling Technology, 2901) using 2% S2 chromatin spike-in was performed on a Diagenode SX-8G IP-Star Compact using the Diagenode Automated Ideal ChIP–seq kit according to the manufacturer's instructions.

ChIP–qPCR and ChIP–seq. ChIP–qPCR was performed using the Applied Biosystems StepOnePlus system and SYBR green dye. Primers are listed in Supplementary Table 2. Fold enrichment was calculated by dividing the enrichment (per cent input) at individual genomic loci by the averaged signal from multiple negative control regions (for example, H3K36me2-depleted intergenic loci). For ChIP–seq, library preparation was carried out using KAPA HTP Illumina or KAPA Hyper Prep library preparation reagents, following the manufacturer's protocol. ChIP libraries were sequenced using the Illumina HiSeq 2000, 2500 or 4000 platform with 50-bp single reads.

RNA sequencing. Total RNA was extracted from cell pellets using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen) according to the manufacturer's instructions. Library preparation was performed with ribosomal RNA depletion

according to the manufacturer's instructions (Epicentre) to achieve greater coverage of mRNA and other long non-coding transcripts. Paired-end sequencing (125 bp) was performed on the Illumina HiSeq 2500 or 4000 platform.

WGBS. Whole-genome-sequencing libraries were generated from 1,000 ng of genomic DNA spiked with 0.1% (w/w) unmethylated λ DNA (Roche Diagnostics) previously fragmented to peak sizes of 300–400 bp using the Covaris focused-ultrasonicator E210. Fragment size was controlled on a Bioanalyzer High Sensitivity DNA Chip (Agilent), and the KAPA High Throughput Library Preparation Kit (KAPA Biosystems) or NxSeq AmpFREE Low DNA Library Kit (Lucigen) applied. End repair of the generated double-stranded DNA with 3′ or 5′ overhangs, adenylation of 3′-ends, adaptor ligation and clean-up steps were carried out as per the manufacturer's recommendations. The cleaned-up ligation product was then analysed on a Bioanalyzer High Sensitivity DNA Chip (Agilent). Samples were then bisulfite-converted using the EZ-DNA Methylation Gold Kit (Zymo Research) according to the manufacturer's protocol. DNA was amplified by nine cycles of PCR using the Kapa Hifi Uracil+ DNA polymerase (KAPA Biosystems) according to the manufacturer's protocol. The amplified libraries were purified using Ampure XP Beads (Beckman Coulter), validated on Bioanalyzer High Sensitivity DNA Chips and quantified by PicoGreen. Sequencing of the WGBS libraries was performed on the Illumina HiSeqX system using 150-bp paired-end sequencing.

Analysis of ChIP–seq data. Raw sequencing reads were aligned using BWA[33] (v.0.7.17) with default parameters to the hg19, mm10 and dm6 reference genomes for human, mouse and *Drosophila* data, respectively. Aligned reads were summed into bins according to different annotations (CpG islands, promoters, genic regions, 1-kb, 10-kb and 100-kb bins) using BEDtools[34] (v.2.22.1). Annotations of CpG islands and RefSeq transcripts for the hg19 and mm10 genomes were downloaded from the UCSC Table Browser. Promoters were defined as 5-kb regions centred on the RefSeq transcriptional start site. For genic regions, we used the longest isoform. ChIP–seq coverage tracks were generated and visualized using the Integrated Genomics Viewer[35,36] v.2.3. The ChIP enrichment was computed using ChIP-Rx, which uses spike-in *Drosophila* chromatin as internal control[18]. For each ChIP–seq profile, the ChIP-Rx ratio (denoted as Rx) was calculated as follows:

$$Rx = \frac{SP/SP_{dmel}}{IP/IP_{dmel}},$$

in which SP and SP$_{dmel}$ are the percentages of reads in the target sample that mapped to the mouse (or human) and spike-in *Drosophila* genomes, respectively (and in a similar manner, IP and IP$_{dmel}$ indicate the input percentages). Tracks were normalized by dividing by the total sequencing depth and multiplying by the ChIP-Rx ratio. The signal was multiplied by 10^{10} in all samples. For analysis of ChIP–seq data, binned reads were divided by total sequencing depth and normalized over input (using log$_2$). De novo methylation plots were generated as previously described [10].

To call broad domains of H3K36me2, H3K27me3 and H3K9me3, ChIP signals (after input normalization) were segmented to identify changes in signal average and variance using PELT[37,38]. Segments were merged within a distance of 10 bp, and the segments greater than 500 kb were outputted as the large domains. BEDtools was used to aggregate the signals of CpG methylation, gene expression and ChIP within these broad domains.

Gene-body plots were generated using ngs.plot.r[39] with a fragment length of 300 bp and flanking length of 10 kb. For gene coordinates, we used a BED file of the top 50% most-expressed genes, sorted by distance between the transcriptional start site and first internal exon (relative to total gene length when appropriate).

Annotated genomic regions within H3K36me2 domains were identified by a custom PERL script. ChIP–seq reads for H3K36me2 and H3K36me3 in these regions were observed by featureCounts[40]. Expected ChIP–seq reads with equivalent numbers of total reads for H3K36me2 and H3K36me3 in the same regions were simulated by the 'shuffle' function of BEDtools.

Analysis of RNA-sequencing data. Raw sequencing reads were aligned using STAR[41] (v.2.5.3a) to the reference genomes mentioned above. Gene-level read counts were obtained from featureCounts[40] (v.1.5.3) using the USCS gene annotation and normalized to total sequencing depth.

Analysis of WGBS data. Raw reads were aligned to mouse or human genome builds (UCSC mm10 or hg19) using BWA[33] (v.0.6.1) after converting the reference genome to bisulfite mode. Low-quality sequences at the 3′ ends were trimmed. For the overlapping paired-end reads, we clipped the 3′ end of one of them to avoid double counting on both forward and reverse strands. After alignment, we filtered duplicated or poorly mapping reads (more than 2% mismatches or aberrant insert size). To call methylation of individual CpGs, we used SAMtools[42] (v.0.1.18) mpileup. We filtered CpGs that had less than 5× coverage, that overlapped with SNPs from dbSNPs (build 137) or that were located within the ENCODE DAC blacklisted regions or Duke excluded regions[43]. For visualization in the Integrated Genomics Viewer, the coordinates of individual CpGs were artificially extended to

the midpoint between their neighbours, as previously described[44]. Partially methylated domains were called in a similar fashion[45]. Average methylation levels were computed in 10-kb bins and neighbouring bins with methylation levels of over 70% were merged into putative partially methylated domains. Only regions greater than 1 Mb were considered to be true partially methylated domains.

DNA methylation clustering. Illumina HumanMethylation450 BeadChip methylation beta values of samples from patients with TBRS, Sotos syndrome and Weaver syndrome, together with healthy control individuals, were obtained from the Gene Expression Omnibus (GEO) database (accession numbers: GSE74432 and GSE128801). After removing probes that are on the sex chromosomes, unsupervised hierarchical clustering was performed using the 1,000 most-variable sites between samples from patients with TBRS and control samples. Clustering was performed using the pheatmap R function (Pretty Heatmaps v1.0.10) with parameters: clustering_method = 'complete', clustering_distance_cols = 'euclidean' and cutree_cols = 2.

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

Data availability

The WGBS, ChIP–seq and RNA-sequencing data have been deposited in the GEO under accession number GSE118785.

Code availability

The source code for bioinformatic analysis is available on request.

31. Sidoli, S., Bhanu, N. V., Karch, K. R., Wang, X. & Garcia, B. A. Complete workflow for analysis of histone post-translational modifications using bottom-up mass spectrometry: from histone extraction to data analysis. *J. Vis. Exp.* **111**, e54112 (2016).
32. Yuan, Z. F. et al. EpiProfile 2.0: a computational platform for processing epi-proteomics mass spectrometry data. *J. Proteome Res.* **17**, 2533–2541 (2018).
33. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
34. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
35. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
36. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
37. Killick, R. et al. Optimal detection of changepoints with a linear computation cost. *J. Am. Stat. Assoc.* **107**, 1590–1598 (2012).
38. Wambui, G. D. et al. The power of the pruned exact linear time (PELT) test in multiple changepoint detection. *Am. J. Theor. Appl. Stat.* **4**, 581–586 (2015).
39. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).
40. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
41. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
42. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
43. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
44. Hovestadt, V. et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).
45. Berman, B. P. et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* **44**, 40–46 (2012).

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

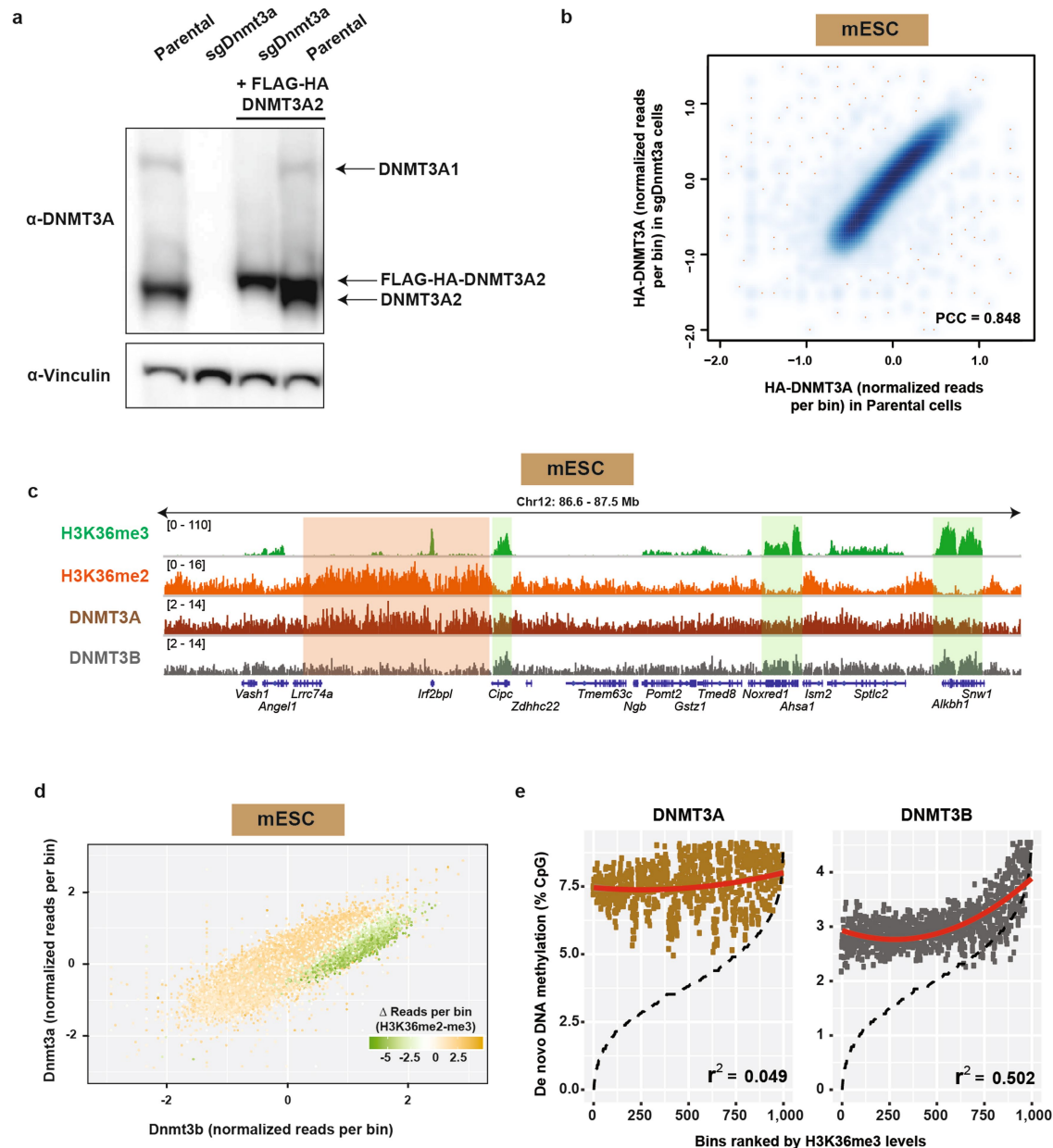**Extended Data Fig. 1 | H3K36me2 and H3K36me3 mark transcriptionally active euchromatin. a**, ChIP–seq normalized reads of H3K36me2 relative to the average percentage of CpG methylation in mouse MSCs for 100-kb non-overlapping bins ($n = 25,624$). Pearson's correlation coefficient is indicated. **b**, Quantification of ChIP–seq normalized reads for averaged CpG methylation, DNMT3A1 and DNMT3B within H3K36me2 and H3K36me3 ($n = 591$), H3K27me3 ($n = 283$) or H3K9me3 ($n = 545$) domains. P values were $<2.2 \times 10^{-16}$ for all pair-wise comparisons, as determined by Wilcoxon's rank-sum test (two-sided), except for DNMT3B ($P = 3.73 \times 10^{-14}$) between H3K27me3 and H3K9me3 domains. For box plots, the centre lines represent the median, the box limits are the 25th and 75th percentiles, the whiskers are the minimum to maximum values and discrete points represent outliers. **c**, Quantification of ChIP–seq normalized reads for gene expression, H3K4me1 and H3K27ac within H3K36me2 and H3K36me3

($n = 591$), H3K27me3 ($n = 283$) or H3K9me3 ($n = 545$) domains. P values were $<2.2 \times 10^{-16}$ for all pair-wise comparisons, as determined by Wilcoxon's rank-sum test (two-sided), except for gene expression ($P = 1.72 \times 10^{-6}$) between H3K27me3 and H3K9me3 domains. Box plots as in **b**. **d**, Three-dimensional genome browser representation of Hi-C chromatin conformation capture data from mouse myoblasts compared to ChIP–seq normalized reads for histone PTMs in mouse MSCs at Chr4: 52.2–99.1 Mb. The levels of CpG methylation (CpGme) are depicted as a heat map (blue, low; white, intermediate; red, high). For H3K36me3 and H3K36me2, data are representative of two independent ChIP–seq experiments; for H3K27me3 and H3K9me3, ChIP–seq was performed once and an independent ChIP was performed in which genomic regions of selective enrichment and depletion were confirmed by qPCR. WGBS was performed once.
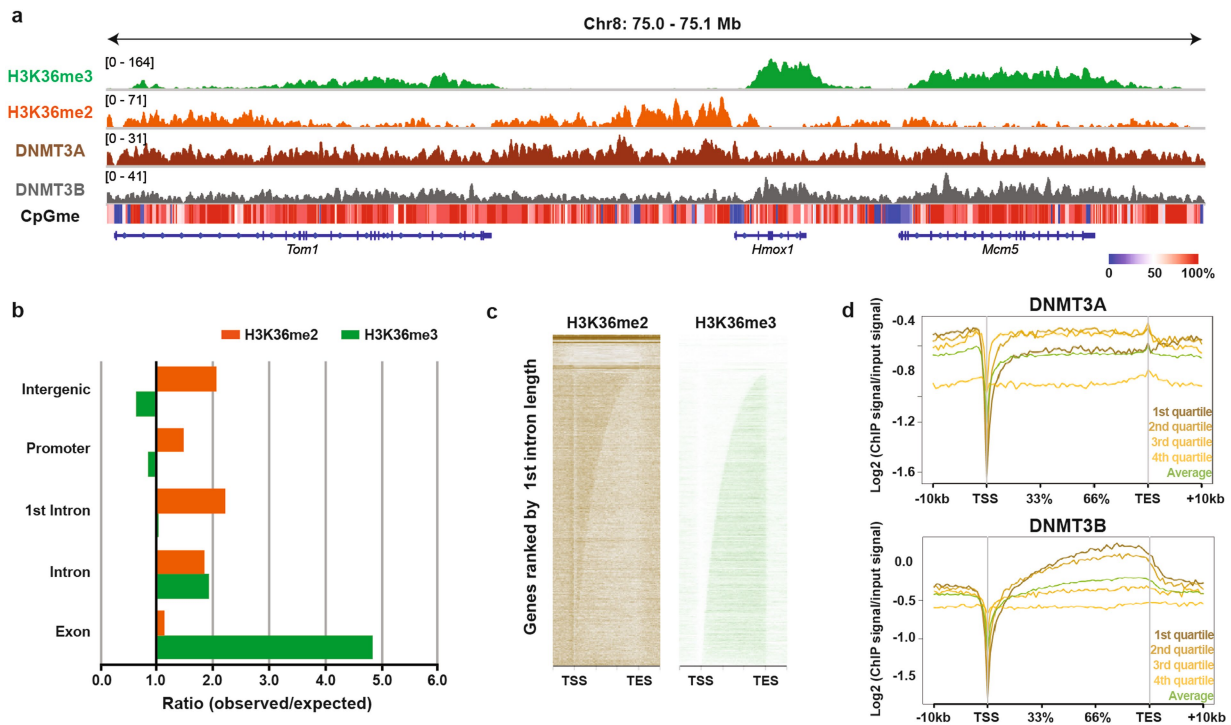
Extended Data Fig. 2 | Distinct enrichment patterns between H3K36me2 and H3K36me3 at euchromatin. a, Genome browser representation of ChIP–seq normalized reads for H3K36me3, H3K36me2, DNMT3A1 and DNMT3B in mouse MSCs at Chr8: 75.0–75.1 Mb. The levels of CpG methylation are depicted as a heat map (blue, low; white, intermediate; red, high) and RefSeq genes are annotated at the bottom. For H3K36me3, H3K36me2 and DNMT3A1, data are representative of two independent ChIP–seq experiments; for DNMT3B, ChIP–seq was performed once and an independent ChIP was performed in which genomic regions of selective enrichment and depletion were confirmed by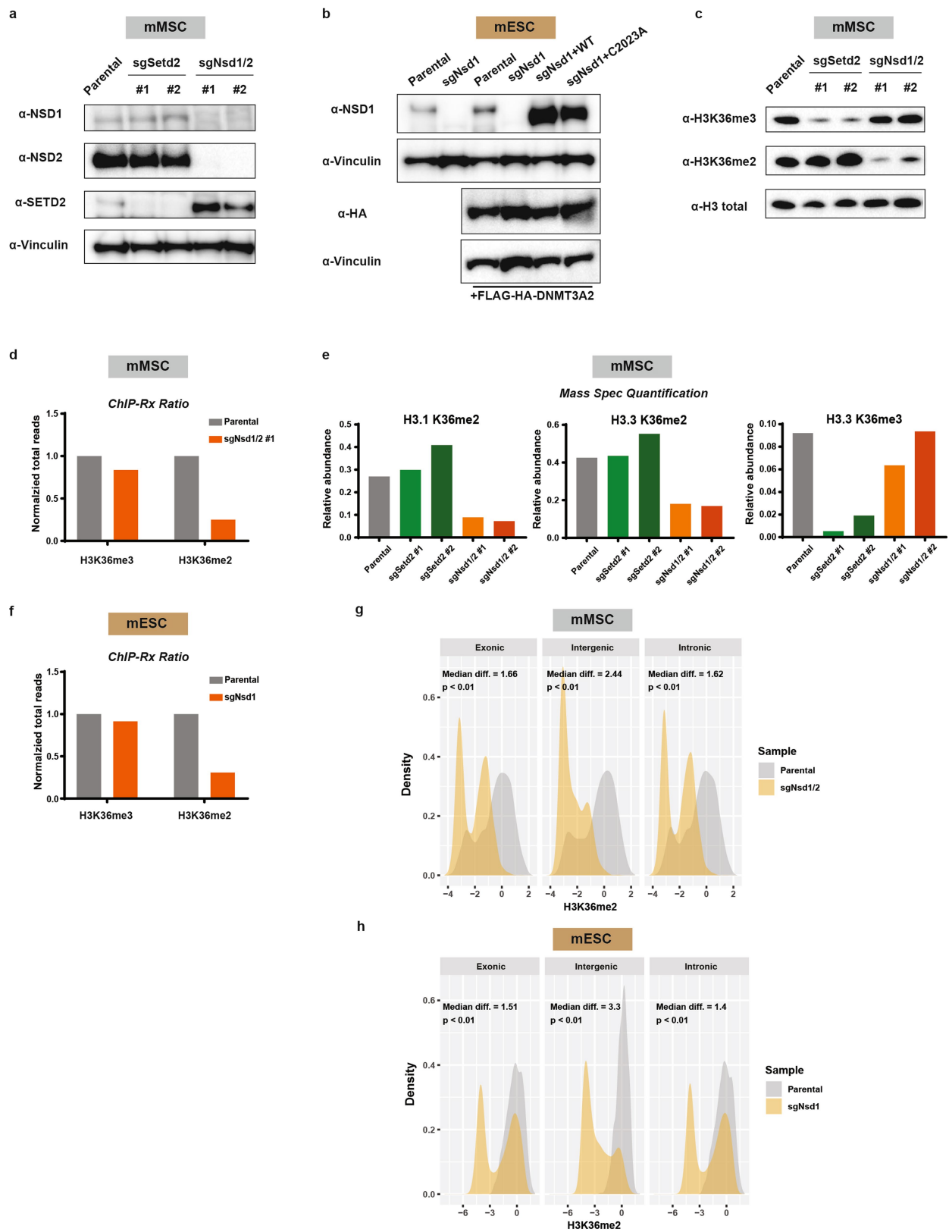 qPCR. WGBS was performed once. **b**, Ratio of observed-to-expected ChIP–seq reads for H3K36me2 and H3K36me3 in annotated genomic regions. Numbers of expected reads were generated assuming equivalent genomic distribution to input. **c**, Heat maps representing ChIP–seq signal density for H3K36me2 and H3K36me3 in mouse MSCs across all gene bodies. Genes are ranked by the length of the first intron. Each gene is displayed as a row. **d**, Averaged ChIP–seq normalized signal across gene bodies stratified by expression quartile, represented as $\log_2$(ChIP signal/input signal) over input for DNMT3A1 (top) and DNMT3B (bottom) in parental MSCs. Sample sizes (which are the same for DNMT3A1 and DNMT3B) are $n = 7,524$, $n = 7,435$, $n = 6,923$ and $n = 8,550$ for the first, second, third and fourth quartiles, respectively; and $n = 35,777$ for the average. TES, transcription end site; TSS, transcription start site.

Extended Data Fig. 3 | Genome-wide colocalization between H3K36me2 and DNMT3A2 in mouse ES cells. a, Immunoblots of lysates generated from parental and sgDnmt3a mouse ES cells (termed mESCs in the Extended Data figures) that ectopically express HA-tagged DNMT3A2. Vinculin was used as a loading control. Endogenous expression of the long isoform (DNMT3A1) and the short isoform (DNMT3A2) of DNMT3A is indicated. Data are representative of two independent experiments. **b**, ChIP–seq normalized reads of HA-tagged DNMT3A2 in sgDnmt3a relative to parental ES cells for 100-kb non-overlapping bins ($n = 26,181$). Pearson's correlation coefficient is indicated. **c**, Genome browser representation of ChIP–seq normalized reads for H3K36me3, H3K36me2, DNMT3A2 and DNMT3B in ES cells at Chr12: 86.6–87.5 Mb. RefSeq genes are annotated at the bottom. The shaded areas indicate H3K36me2-enriched intergenic regions (orange) and H3K36me3-enriched genic regions (green) in parental cells. For H3K36me2 and DNMT3A2,

ChIP–seq was performed once and an independent ChIP was performed in which genomic regions of selective enrichment and depletion were confirmed by qPCR. H3K36me3 and DNMT3B ChIP–seq were performed once. **d**, ChIP–seq normalized reads per 10-kb bin for DNMT3A2 (y axis) and DNMT3B (x axis) in ES cells ($n = 246,285$). Bins (dots) are colour-coded on the basis of differences between H3K36me3 and H3K36me2 ChIP–seq reads, to show selective enrichment for H3K36me2 (orange) or H3K36me3 (green). **e**, De novo methylation per bin by DNMT3A2 (brown) or DNMT3B (grey) after reintroduction into *Dnmt1*, *Dnmt3a* and *Dnmt3b* triple-knockout ES cells relative to H3K36me3. To generate bins, 1-kb genomic tiles ($n = 2,462,755$) were ranked by H3K36me3 enrichment in ES cells and grouped into 1,000 bins, ordered by rank (2,463 tiles per group). The dashed lines indicate H3K36me3 enrichment per bin. Goodness of fit was computed using a quadratic model (red lines). For gel source data, see Supplementary Fig. 1.

a mMSC

α-NSD1
α-NSD2
α-SETD2
α-Vinculin

b mESC

α-NSD1
α-Vinculin
α-HA
α-Vinculin

+FLAG-HA-DNMT3A2

c mMSC

α-H3K36me3
α-H3K36me2
α-H3 total

d mMSC

ChIP-Rx Ratio

e mMSC

Mass Spec Quantification

H3.1 K36me2

H3.3 K36me2

H3.3 K36me3

f mESC

ChIP-Rx Ratio

g mMSC

Exonic — Median diff. = 1.66, p < 0.01
Intergenic — Median diff. = 2.44, p < 0.01
Intronic — Median diff. = 1.62, p < 0.01

H3K36me2

Sample: Parental, sgNsd1/2

h mESC

Exonic — Median diff. = 1.51, p < 0.01
Intergenic — Median diff. = 3.3, p < 0.01
Intronic — Median diff. = 1.4, p < 0.01

H3K36me2

Sample: Parental, sgNsd1

Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Genetic ablation of *Nsd1* and *Nsd2* in mouse MSCs and *Nsd1* in mouse ES cells. a, Immunoblots of lysates from parental and H3K36-methyltransferase-knockout mouse MSC clonal lines for NSD1, NSD2 and SETD2. Vinculin was used as a loading control. **b**, Immunoblots of lysates from parental and sgNsd1 mouse ES cells that express HA-tagged DNMT3A. sgNsd1 cells were rescued with ectopic expression of wild-type NSD1 or the C2023A catalytic mutant. Vinculin was used as a loading control. **c**, Immunoblots of lysates generated from parental, sgSetd2 and sgNsd1/2 MSCs for H3K36me3 and H3K36me2. Total H3 was used as a loading control. **d**, Ratios of ChIP–seq reads for H3K36me2 and H3K36me3 in mouse MSCs between target chromatin (mouse) and reference spike-in chromatin (*Drosophila*). Data are representative of two independent experiments. **e**, Quantitative mass spectrometry measurement of the abundance of histone PTMs in acid-extracted histones derived from the indicated MSC lines. Data are from one experiment. **f**, Ratios of ChIP–seq reads for H3K36me2 and H3K36me3 in ES cells between target chromatin (mouse) and reference spike-in chromatin (*Drosophila*). Data are from one experiment. **g**, Density plots of H3K36me2 levels at intergenic ($n = 1,165$), exonic ($n = 13,601$) and intronic ($n = 12,364$) regions for parental (grey) and sgNsd1/2 (orange) MSCs. P values were determined by Wilcoxon's rank-sum test (two-sided). **h**, Density plots of H3K36me2 levels at intergenic ($n = 1,165$), exonic ($n = 13,601$) and intronic ($n = 12,364$) regions for parental (grey) and sgNsd1 (orange) ES cells. P values were determined by Wilcoxon's rank-sum test (two-sided). The immunoblot experiments in **a–c** were independently repeated twice with similar results. For gel source data, see Supplementary Fig. 1.
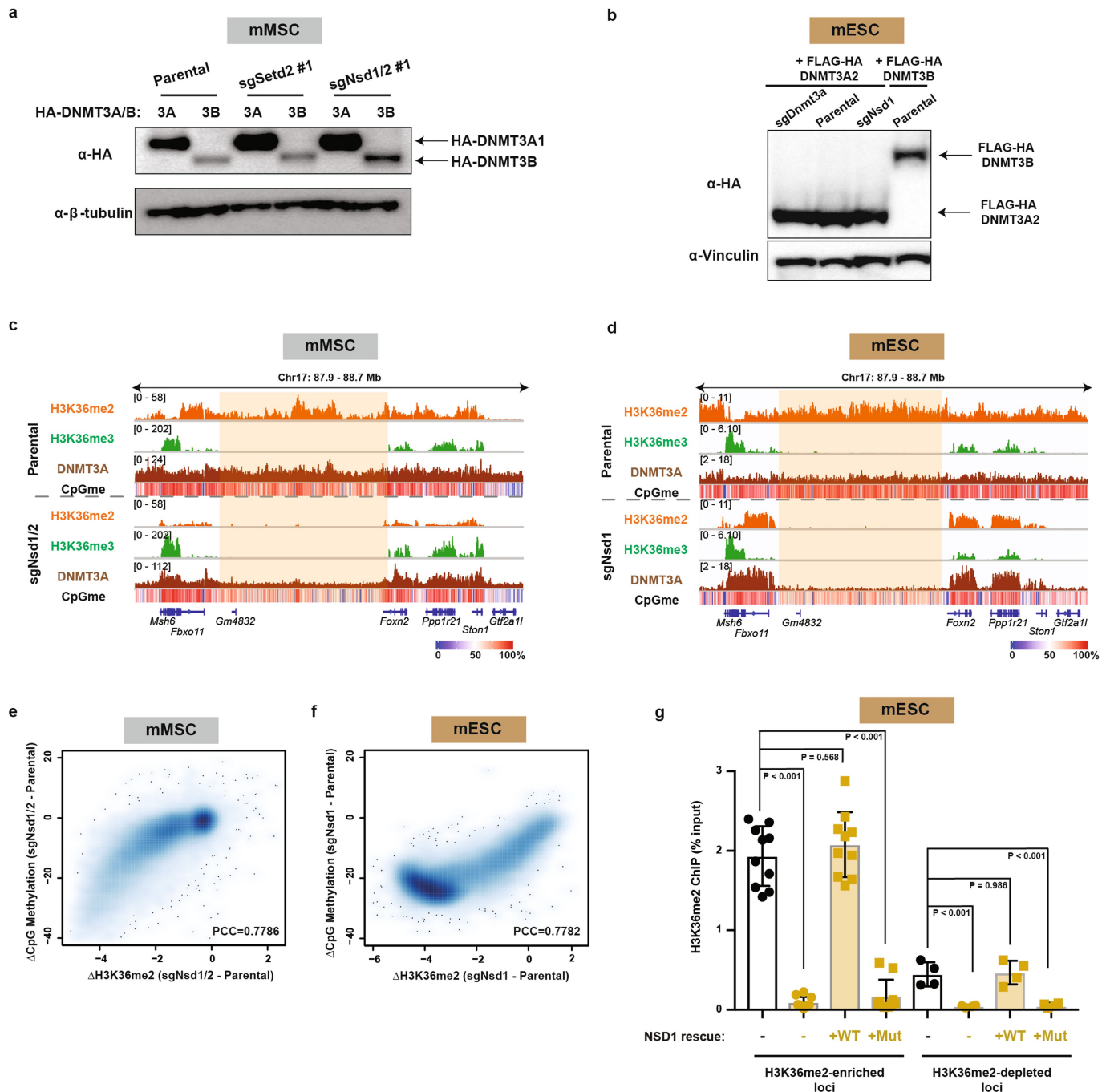
Extended Data Fig. 5 | H3K36me2 depletion impairs genomic targeting of DNMT3A and reduces intergenic CpG methylation.
**a**, Immunoblots of lysates from parental, sgSetd2 and sgNsd1/2 mouse MSCs that ectopically express HA-tagged DNMT3A1 or DNMT3B. β-tubulin was used as a loading control. **b**, Immunoblots of lysates from sgDnmt3a, parental and sgNsd1 mouse ES cells that ectopically express HA-tagged DNMT3A2 or DNMT3B. Vinculin was used as a loading control. **c**, Genome browser representation of ChIP–seq normalized reads for H3K36me3, H3K36me2 and DNMT3A1 in parental and sgNsd1/2 MSCs at Chr17: 87.9–88.7 Mb. The levels of CpG methylation (CpGme) are depicted as a heat map (blue, low; white, intermediate; red, high) and RefSeq genes are annotated at the bottom. The shaded area indicates H3K36me2-enriched intergenic region in parental cells. For H3K36me2, H3K36me3 and DNMT3A1 in parental cells, data are representative of two independent ChIP–seq experiments. DNMT3A1 ChIP–seq in sgNsd1/2 cells and WGBS in both lines were performed once. **d**, Genome browser representation of ChIP–seq normalized reads for H3K36me3, H3K36me2 and DNMT3A2 in parental and sgNsd1 ES cells at Chr17: 87.9-88.7 Mb. The levels of CpG methylation are depicted as a heat map (blue, low; white, intermediate; red, high) and RefSeq genes are annotated at the bottom.

The shaded area indicates the H3K36me2-enriched intergenic region in parental cells. For H3K36me2 and DNMT3A2, ChIP–seq was performed once and an independent ChIP was performed in which genomic regions of selective enrichment and depletion were confirmed by qPCR. WGBS and H3K36me3 ChIP–seq were performed once. **e**, Per cent change in averaged CpG methylation between parental and sgNsd1/2 MSCs relative to changes in ChIP–seq normalized reads of H3K36me2 for 100-kb non-overlapping bins ($n = 25,611$). Pearson's correlation coefficient is indicated. **f**, Per cent change in averaged CpG methylation between parental and sgNsd1 ES cells relative to changes in ChIP–seq normalized reads of H3K36me2 for 100-kb non-overlapping bins ($n = 26,044$). Pearson's correlation coefficient is indicated. **g**, The enrichment (per cent input) of H3K36me2 at various intergenic regions in parental (black) and sgNsd1 (orange) ES cells rescued with ectopic expression of wild-type NSD1 or the C2023A catalytic mutant was measured by ChIP–qPCR. Each data point represents a genomic locus ($n = 10$ for H3K36me2-enriched regions, $n = 4$ for H3K36me2-depleted regions). Data are mean ± s.d. P values were determined by one-way ANOVA. The immunoblot experiments in **a**, **b** were independently repeated twice with similar results. For gel source data, see Supplementary Fig. 1.
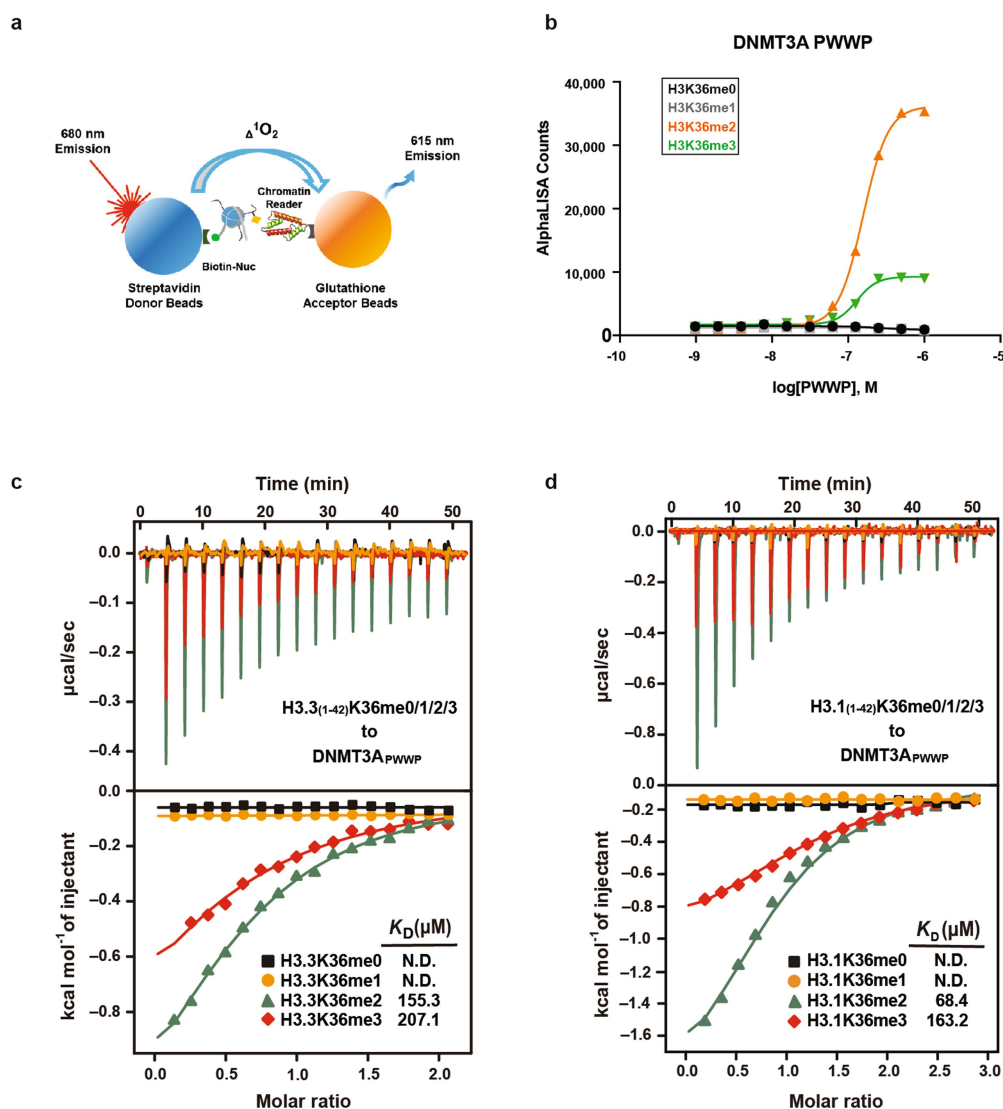
Extended Data Fig. 6 | The PWWP domain of DNMT3A preferentially binds to H3K36me2. a, The dCypher approach for interrogating chromatin readers (Methods). Biotinylated nucleosomes are immobilized on streptavidin donor beads and GST-tagged DNMT3A$_{PWWP}$ on glutathione AlphaLISA acceptor beads. Laser excitation of the donor generates singlet oxygen, which diffuses to activate emission from the acceptor. Fluorescence counts are directly proportional to the amount of donor–acceptor complex bridged by the nucleosome-reader interaction. **b,** Quantification of AlphaLISA counts for isolated GST-tagged DNMT3A$_{PWWP}$ interaction, titrated against H3K36-modified nucleosomes. Data are the mean of two replicates. **c,** ITC titration and fitting curves of human DNMT3A$_{PWWP}$ with H3.3$_{(1–42)}$ K36-modified peptides. **d,** ITC titration and fitting curves of human DNMT3A$_{PWWP}$ with H3.1$_{(1–42)}$ K36-modified peptides.
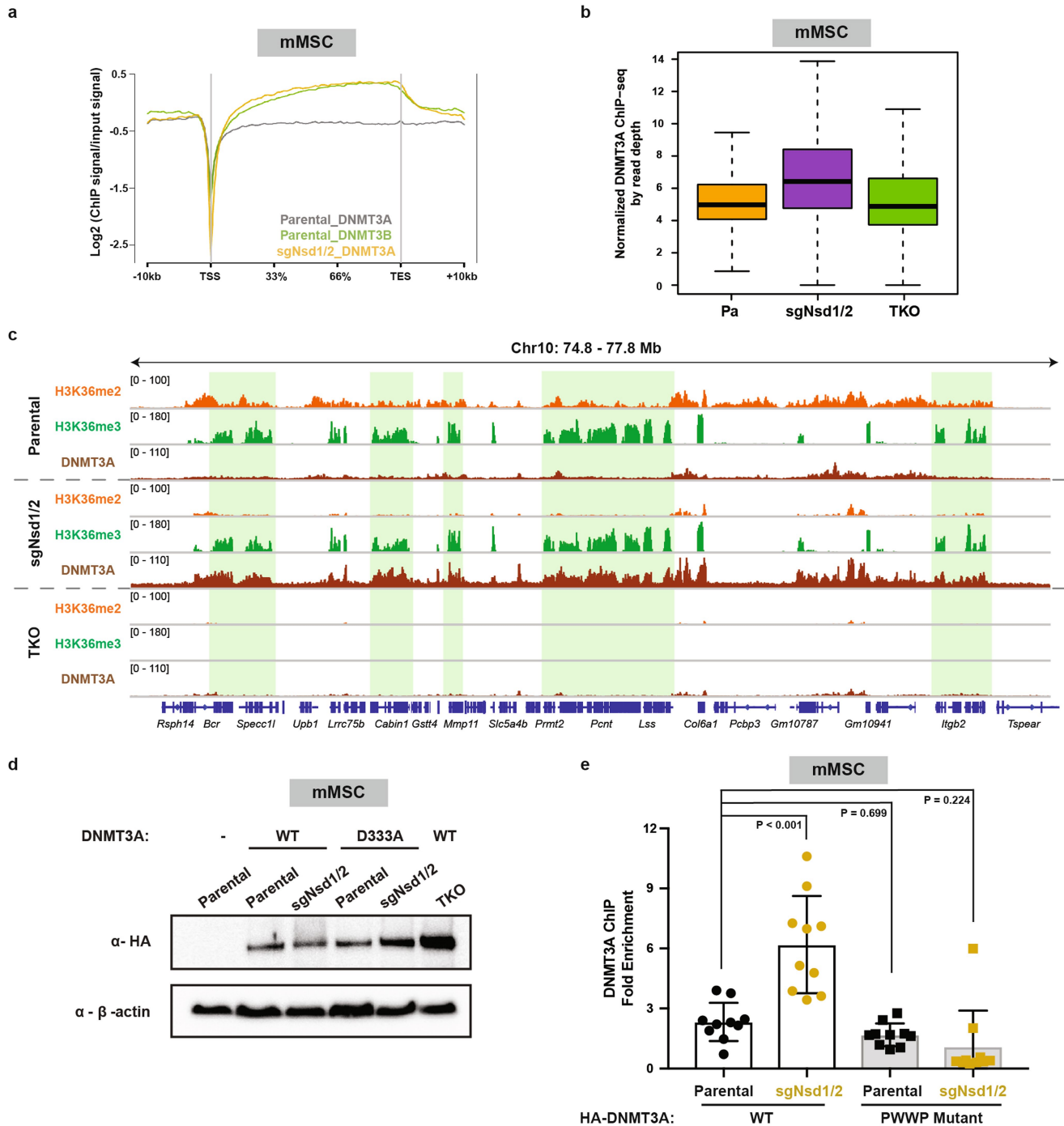
Extended Data Fig. 7 | Redistribution of DNMT3A to H3K36me3-marked gene bodies after loss of H3K36me2. a, Averaged ChIP–seq normalized signal across all gene bodies, represented as $\log_2$(ChIP signal/input signal) for DNMT3A1 (grey, $n = 14{,}959$) and DNMT3B (green, $n = 14{,}959$) in parental mouse MSCs, and for DNMT3A1 (orange, $n = 14{,}311$) in sgNsd1/2 MSCs. **b**, Quantification of ChIP–seq normalized reads of DNMT3A1 in parental, sgNsd1/2 and triple-knockout (TKO) MSCs at 10-kb non-overlapping bins enriched for H3K36me3 in parental cells (top 20% of bins, $n = 54{,}624$). Reads were normalized by read depth. P values were $<2.2 \times 10^{-16}$ for all pair-wise comparisons, as determined by Wilcoxon's rank-sum test (two-sided). For box plots, the centre lines represent the median, the box limits are the 25th and 75th percentiles, the whiskers are the minimum to maximum values and discrete points represent outliers. **c**, Genome browser representation of ChIP–seq normalized reads for H3K36me2, H3K36me3 and DNMT3A1 in parental, sgNsd1/2 and triple-knockout MSCs at Chr10: 74.8–77.8 Mb. RefSeq genes are annotated at the bottom. The shaded areas indicate H3K36me3-enriched genic regions in parental cells. For H3K36me2 and H3K36me3, data in parental and sgNsd1/2 cells are representative of two independent ChIP–seq experiments. ChIP–seq in triple-knockout cells for H3K36me2 and H3K36me3 was performed once. For DNMT3A1, data in parental cells are representative of two independent ChIP–seq experiments. ChIP–seq in sgNsd1/2 and triple-knockout cells for DNMT3A1 was performed once. **d**, Immunoblots of lysates from parental, sgNsd1/2 and triple-knockout MSCs that ectopically express HA-tagged wild-type or D333A-mutant DNMT3A1. β-actin was used as a loading control. Data are representative of two independent experiments. **e**, Fold enrichment of wild-type or PWWP-mutant (D333A) DNMT3A1 at gene-body versus intergenic regions in parental (black) and sgNsd1/2 (orange) MSCs was measured by ChIP–qPCR. Each data point represents a genomic locus ($n = 10$). Data are mean ± s.d. P values were determined by one-way ANOVA. For gel source data, see Supplementary Fig. 1.
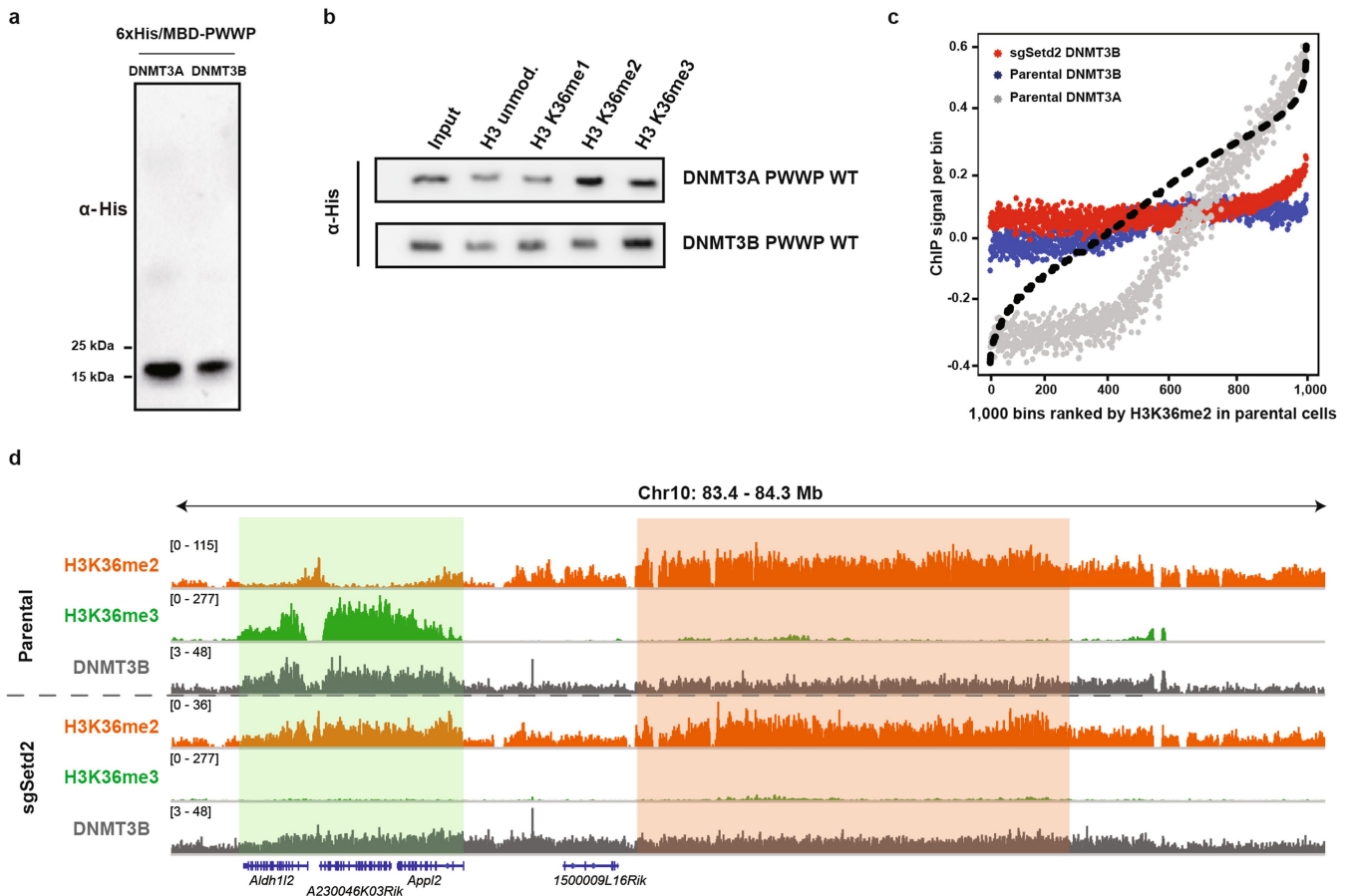
Extended Data Fig. 8 | Assessment of binding affinity between DNMT3B and H3K36me2 in vitro and in cells. a, Immunoblots of His–MBP-tagged recombinant DNMT3A and DNMT3B PWWP domains. **b**, Immunoblots of recombinant His–MBP-tagged wild-type DNMT3A and DNMT3B PWWP domains bound to H3K36-modified recombinant nucleosomes following the in vitro pull-down assay. **c**, ChIP–seq normalized reads per bin for DNMT3A1 (grey) and DNMT3B (blue) in parental mouse MSCs and DNMT3B (red) in sgSetd2 MSCs relative to H3K36me2. To generate bins, 1-kb genomic tiles were ranked by H3K36me2 enrichment in parental MSCs and grouped into 1,000 bins, ordered by rank. The dashed line indicates H3K36me2 enrichment per bin. **d**, Genome browser representation of ChIP–seq normalized reads for H3K36me2, H3K36me3 and DNMT3B in parental and sgSetd2 MSCs at Chr10: 83.4–84.3 Mb. RefSeq genes are annotated at the bottom. The shaded areas indicate H3K36me2-enriched intergenic regions (orange) and H3K36me3-enriched genic regions (green) in parental cells. For H3K36me2 and H3K36me3, data are representative of two independent ChIP–seq experiments; for DNMT3B, ChIP–seq was performed once and an independent ChIP was performed in which genomic regions of selective enrichment and depletion were confirmed by qPCR. The immunoblot experiments in **a**, **b** were independently repeated twice with similar results. For gel source data, see Supplementary Fig. 1.
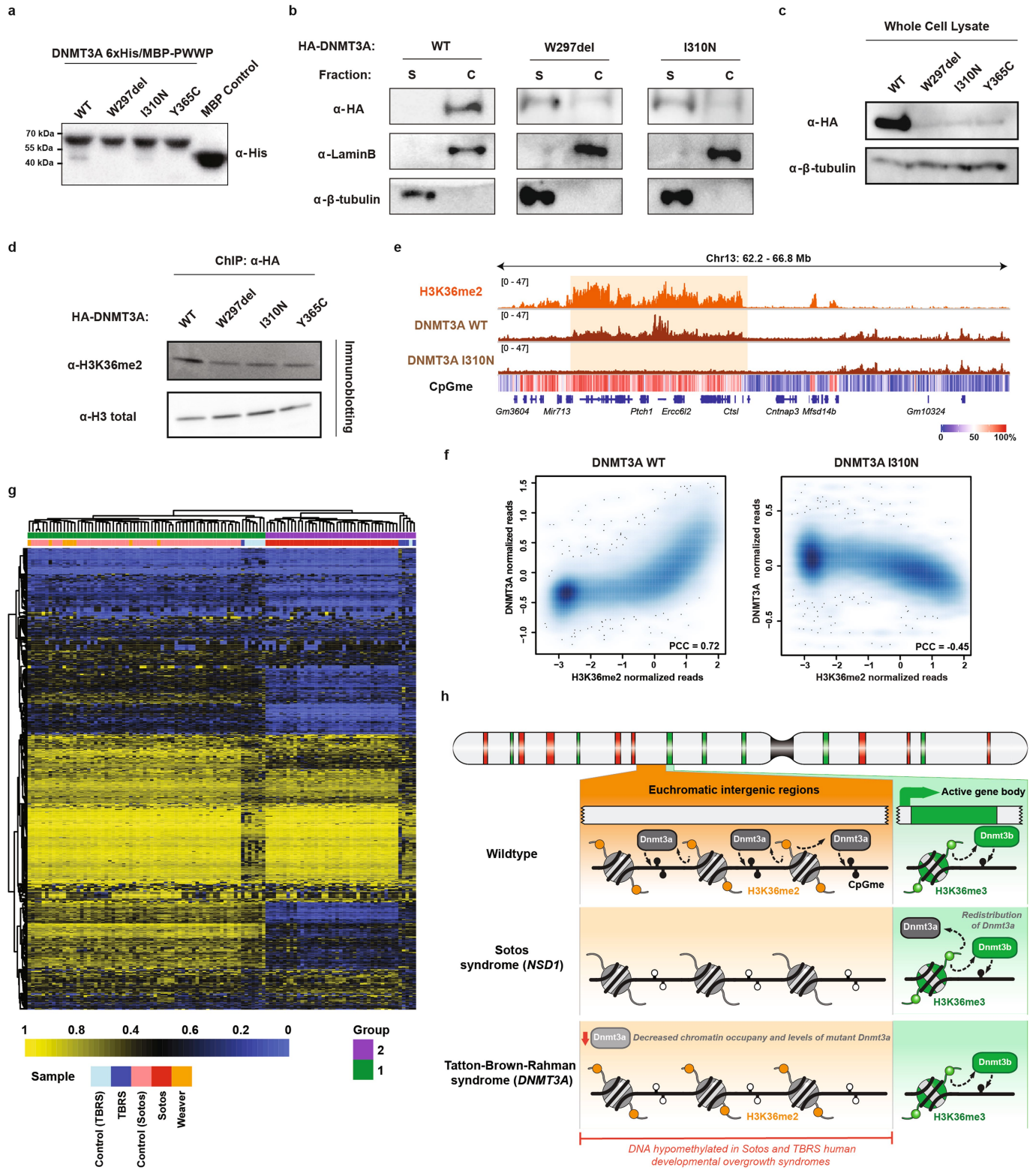
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | TBRS-associated mutations in DNMT3A are loss-of-function and result in DNA hypomethylation similar to that observed in Sotos syndrome. a, Immunoblots of the recombinant His–MBP-tagged DNMT3A PWWP domain containing TBRS-associated mutations. MBP alone was used as a control. **b**, Immunoblots of soluble or chromatin-associated lysates generated from cells that ectopically express HA-tagged wild-type or mutant (W297del or I310N) DNMT3A. β-tubulin and lamin B1 were used as loading controls for the soluble and chromatin-associated fractions, respectively. **c**, Immunoblots of lysates from parental mouse MSCs that ectopically express HA-tagged wild-type or TBRS-associated mutant (W297del, I310N or Y365C) DNMT3A. Data are from one experiment. β-tubulin was used as a loading control. **d**, Immunoblots of nucleosomes bound to HA-tagged wild-type or mutant (W297del, I310N or Y365C) DNMT3A after ChIP with anti-HA-tag antibodies. Total H3 was used as a loading control and to normalize for differences in protein expression and nucleosome pull-down efficiency between samples. **e**, Genome browser representation of ChIP–seq normalized reads for H3K36me2, wild-type DNMT3A1 and I310N-mutant DNMT3A1 in mouse MSCs at Chr13: 62.2–66.8 Mb. The levels of CpG methylation (CpGme) are depicted as a heat map (blue, low; white, intermediate; red, high) and RefSeq genes are annotated at the bottom. **f**, ChIP–seq normalized reads of wild-type or PWWP-mutant (I310N) DNMT3A1 relative to that of H3K36me2 for 100-kb non-overlapping

bins (wild type: $n = 25,694$; I310N: $n = 25,757$). Pearson's correlation coefficient is indicated. For H3K36me2 and wild-type DNMT3A1, data are representative of two independent ChIP–seq experiments; for I310N-mutant DNMT3A1, ChIP–seq was performed once and an independent ChIP was performed in which genomic regions of selective enrichment and depletion were confirmed by qPCR. WGBS was performed once. **g**, Unsupervised hierarchical clustering of publicly available Infinium Human Methylation 450K array profiles of blood samples from patients with TBRS, Sotos syndrome and Weaver syndrome and healthy control individuals, based on the top 1,000 most-variable probes. **h**, Model depicting the changes in chromatin landscape in human developmental overgrowth syndromes (Sotos syndrome and TBRS). In normal development, DNMT3A and DNMT3B act in parallel to methylate CpG dinucleotides at H3K36me2-enriched intergenic and H3K36me3-enriched genic regions, respectively. Haploinsufficiency of *NSD1* and depletion of intergenic H3K36me2 levels in Sotos syndrome abrogate PWWP-mediated intergenic recruitment of DNMT3A, which leads to hypomethylation of intergenic DNA and the redistribution of DNMT3A to H3K36me3-enriched gene bodies. Mutations within the PWWP domain of DNMT3A in TBRS impair chromatin occupancy and reduce cellular levels of the protein—thereby also causing hypomethylation of intergenic DNA. The immunoblot experiments in **a**, **c**, **d** were independently repeated twice with similar results. For gel source data, see Supplementary Fig. 1.

Corresponding author(s):   C. David Allis, Jacek Majewski, Chao Lu

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect data for this study.
Data analysis	Histone post-translational modification (PTM) profiling by nano-LC-MS was analyzed using EpiProfile 2.0. Isothermal titration calorimetry data was analyzed using Origin 8.0. ChIP-seq reads were aligned with BWA version 0.7.17 and reads were summed across bins using bedtools version 2.22.1. Calling of broad histone PTM domains was performed using PELT. Gene body plots were generated using ngs.plot.r and annotated genomic regions within H3K36me2 domains were identified using a custom PERL script (available upon request) and reads in these regions were observed by featureCounts v1.5.3. Expected reads in the same regions were simulated using bedtools version 2.22.1. Data was visualized using IGV 2.3. RNA-seq reads were aligned using STAR version 2.5.3a and gene-level read counts were obtained from featureCounts v1.5.3. WGBS reads were aligned using BWA version 0.6.1 and methylation of individual CpGs was called using Samtools version 0.1.18. DNA methylation clustering was performed using the pheatmap R function (Pretty Heatmaps v1.0.10) with parameters: clustering_method = 'complete', clustering_distance_cols = 'euclidean', and cutree_cols = 2.

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 upon request. 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

> The WGBS, ChIP-seq, and RNA-seq data have been deposited in the Gene Expression Omnibus (GEO) database under accession number GSE118785.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	For high throughput sequencing experiments, two independent biological replicates were used where possible and to strengthen the main conclusions of the study. This included performing similar sequencing experiments independently in different cell lines (e.g. H3K36me2 ChIP-seq and WGBS in both mMSCs and mESCs, with and without intact NSD1/2 activity). The sample sizes of subsequent bioinformatic analyses were based on grouping the large number of loci into non-overlapping regions to achieve comprehensive genome-wide representation of the sequencing data. Sample sizes for other assays (qPCR, dCypher assays, etc.) were not predetermined and were chosen based on our prior experience and common standards in the field for detecting statistically significant differences between conditions.
Data exclusions	No samples were excluded.
Replication	All findings were reproducible. Confirmation of reproducibility was assessed with a biological replicate and, in the case of high throughput sequencing analysis, through the use of sample sizes that provided comprehensive genome-wide coverage of chromatin landscape changes in a cell line or condition.
Randomization	Samples were allocated to groups according to genotype. No randomization was required because the sequencing of nucleic acid libraries are not affected by sample randomization.
Blinding	Studies were not performed in a blinded fashion. Group allocation was done before data collection.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
[×]	Unique biological materials
[ ]	[×] Antibodies
[ ]	[×] Eukaryotic cell lines
[×]	Palaeontology
[×]	Animals and other organisms
[×]	Human research participants

Methods

n/a	Involved in the study
[ ]	[×] ChIP-seq
[×]	Flow cytometry
[×]	MRI-based neuroimaging

Antibodies

Antibodies used	For Immunoblotting, the following primary antibodies were used: anti-H3K36me2 (Cell Signaling Tech 2901, clone C75H12), anti-H3K36me3 (Active Motif 61101, Lot #08617004), anti-NSD1 (Abbexa abx135901, Lot #L.AB1810J177), anti-NSD2 (Millipore Sigma MABE191, clone 29D1), anti-SETD2 (Abcam ab31358, Lot #GR168354-1), anti-DNMT3A (Abcam ab188470, Lot #GR224165-13), anti-Vinculin (Cell Signaling 13901, clone E1E9V), anti-His (ZSGB-Bio TA-02, clone OTI2B5), anti-Lamin B1 (Cell Signaling 12586, clone D4Q4Z), anti-β-Tubulin (Cell Signaling 2128, clone 9F3), anti-β-actin (Abcam ab8224, Lot #GR289543-12), anti-H3 (Abcam ab1791, Lot #GR292835) and anti-HA (Biolegend 901501, clone 16B12). All primary antibodies used at 1:1000 dilution for immunoblotting, except anti-H3 which was used at 1:10,000. The following secondary antibodies were used at a 1:5000 dilution: anti-Rabbit IgG HRP (Dako P0399, Lot #20028547), and anti-Mouse IgG HRP (GE NA931, Lot #9729340). For ChIP, the following primary antibodies were used: Pierce anti-HA beads (Thermo Scientific 88836, clone 2-2.2.14, 100 ul beads per sample), anti-H3K4me1 (Abcam ab8895, Lot #GR3206285-1, 1:50 dilution), anti-H3K9me3 (Abcam ab8898, Lot #GR3217826-1, 1:100 dilution), anti-H3K27ac (Active Motif 39133, Lot #10918010, 1:100 dilution), anti-H3K27me3 (Cell Signaling 9733, clone C36B11, 1:50 dilution), anti-H3K36me2 (Cell Signaling 2901, clone C75H12, 1:50 dilution) and anti-H3K36me3 (Active Motif 61101, Lot #08617004, 1:50 dilution).
Validation	For Immunoblot, the correct size of the detected bands was assessed based on the protein marker. For ChIP, antibodies were first tested by immunoblotting. The specificities of anti-H3K36me3 and anti-H3K36me2 antibodies were validated using Setd2 knockout and Nsd1/2 double knockout cell lines. anti-H3K36me2: Manufacturer indicates human and mouse reactivity by immunoblot. We have demonstrated in prior work (Lu et al. 2016) that this antibody can be used for ChIP-seq and confirmed its specificity using Setd2 knockout and Nsd1/2 double knockout lines. anti-H3K36me3: Manufacturer indicates human and mouse reactivity by immunoblot and ChIP. Specificity confirmed using Setd2 knockout and Nsd1/2 double knockout cell lines. anti-NSD1: Manufacturer indicates human and mouse reactivity by immunoblot. anti-NSD2: Manufacturer indicates human reactivity by immunoblot. We have demonstrated in prior work (Lu et al. 2016) that this antibody also recognizes mouse NSD2. anti-SETD2: Manufacturer indicates human reactivity by immunoblot. We confirmed using Setd2 knockout mMSCs that this antibody also recognizes mouse SETD2. anti-DNMT3A: Manufacturer indicates human and mouse reactivity by immunoblot. anti-Vinculin: Manufacturer indicates human and mouse reactivity by immunoblot. anti-His: Manufacturer indicates reactivity by immunoblot to detect His-tagged proteins. anti-LaminB1: Manufacturer indicates human and mouse reactivity by immunoblot. anti-β-Tubulin: Manufacturer indicates human and mouse reactivity by immunoblot. anti-β-actin: Manufacturer indicates human and mouse reactivity by immunoblot. anti-H3: Manufacturer indicates human and mouse reactivity by immunoblot. anti-HA: Manufacturer indicates reactivity by immunoblot to detect HA epitope-tagged proteins. anti-H3K4me1: Manufacturer indicates human and mouse reactivity by immunoblot and ChIP. anti-H3K9me3: Manufacturer indicates human and mouse reactivity by immunoblot and ChIP. anti-H3K27ac: Manufacturer indicates human and mouse reactivity by immunoblot and ChIP. anti-H3K27me3: Manufacturer indicates human and mouse reactivity by immunoblot and ChIP.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	293T (ATCC), C3H10T1/2 (ATCC), Fadu (ATCC), SKN-3 (JCRB cell bank), Cal27 (ATCC), SCC-4 (ATCC)
Authentication	C3H10T1/2 sgNsd1/2, sgSetd2, and TKO clones and mESC sgNsd1 and sgDnmt3a clones were individually verified through Sanger sequencing of the target loci. Immunoblot analysis was performed to confirm knockdown of target protein levels. 293T, C3H10T1/2, mESC, and HNSCC parental lines were authenticated based on the testing and monitoring of phenotypic features (morphology, differentiation potential, growth conditions, etc.) characteristic of each line that were previously reported by manufacturers and other groups. Immunoblot analysis was performed to confirm the NSD1 mutational status of each HNSCC line.
Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study.

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.*	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE118785 Enter token efmdsmwcbnsfbgn into the box
Files in database submission	H3K36me2_ChIPseq_Parental.bw H3K36me3_ChIPseq_Parental.bw H3K27me3_ChIPseq_Parental.bw H3K9me3_ChIPseq_Parental.bw H3K4me1_ChIPseq_Parental.bw H3K27ac_ChIPseq_Parental.bw H3K36me2_ChIPseq_SETD2_KO.bw H3K36me2_ChIPseq_NSD1_2_DKO.bw H3K36me3_ChIPseq_SETD2_KO.bw H3K36me3_ChIPseq_NSD1_2_DKO.bw H3K36me2_ChIPseq_mESCs.bw HA-DNMT3A_ChIPseq_Parental.bw HA-DNMT3B_ChIPseq_Parental.bw HA-DNMT3A_I310N_ChIPseq_Parental.bw HA-DNMT3A_ChIPseq_SETD2_KO.bw HA-DNMT3B_ChIPseq_SETD2_KO.bw HA-DNMT3A_ChIPseq_NSD1_2_DKO.bw HA-DNMT3B_ChIPseq_NSD1_2_DKO.bw H3K36me2_ChIPseq_Cal27.bw H3K36me2_ChIPseq_Fadu.bw H3K36me2_ChIPseq_SCC_4.bw H3K36me2_ChIPseq_SKN_3.bw RNAseq_Parental.bw WGBS_Parental.tdf WGBS_NSD_1_2_DKO.tdf WGBS_Cal27.tdf WGBS_Fadu.tdf WGBS_SCC_4.tdf WGBS_SKN_3.tdf WGBS_mESC_Parental.tdf WGBS_mESC_NSD1_KO.tdf H3K36me2_ChIPseq_TKO.bw H3K36me3_ChIPseq_TKO.bw HA-DNMT3A_ChIPseq_TKO.bw H3K36me2_ChIPseq_mESC_NSD1_KO.bw H3K36me3_ChIPseq_mESC_Parental.bw H3K36me3_ChIPseq_mESC_NSD1_KO.bw HA-DNMT3A2_ChIPseq_mESC_Parental.bw HA-DNMT3A2_ChIPseq_mESC_DNMT3A_KO.bw HA-DNMT3A2_ChIPseq_mESC_NSD1_KO.bw HA-DNMT3B_ChIPseq_mESC_Parental.bw
Genome browser session (e.g. UCSC)	Not applicable- visualized data using IGV.

Methodology

Replicates	No technical replicates.
Sequencing depth	ChIP-seq experiments: 50 bp, single-end. H3K36me2_ChIPseq_Parental, total number of reads: 57412745 H3K36me3_ChIPseq_Parental, total number of reads: 53852142 H3K27me3_ChIPseq_Parental, total number of reads: 46870731 H3K9me3_ChIPseq_Parental, total number of reads: 81667574

H3K4me1_ChIPseq_Parental, total number of reads: 51388165
H3K27ac_ChIPseq_Parental, total number of reads: 34252569
H3K36me2_ChIPseq_SETD2_KO, total number of reads: 54721985
H3K36me2_ChIPseq_NSD1_2_DKO, total number of reads: 54646530
H3K36me3_ChIPseq_SETD2_KO, total number of reads: 96166028
H3K36me3_ChIPseq_NSD1_2_DKO, total number of reads: 66334380
H3K36me2_ChIPseq_mESCs, total number of reads: 64997042
HA-DNMT3A_ChIPseq_Parental, total number of reads: 90441880
HA-DNMT3B_ChIPseq_Parental, total number of reads: 78109264
HA-DNMT3A_I310N_ChIPseq_Parental, total number of reads: 66916451
HA-DNMT3A_ChIPseq_SETD2_KO, total number of reads: 56731179
HA-DNMT3B_ChIPseq_SETD2_KO, total number of reads: 114103443
HA-DNMT3A_ChIPseq_NSD1_2_DKO, total number of reads: 211784261
HA-DNMT3B_ChIPseq_NSD1_2_DKO, total number of reads: 82201505
H3K36me2_ChIPseq_Cal27, total number of reads: 71957820
H3K36me2_ChIPseq_Fadu, total number of reads: 73141180
H3K36me2_ChIPseq_SCC_4, total number of reads: 93741149
H3K36me2_ChIPseq_SKN_3, total number of reads: 70520001
H3K36me2_ChIPseq_TKO, total number of reads: 55923287
H3K36me3_ChIPseq_TKO, total number of reads: 61710970
HA-DNMT3A_ChIPseq_TKO, total number of reads: 35949830
H3K36me2_ChIPseq_mESC_NSD1_KO, total number of reads: 58380101
H3K36me3_ChIPseq_mESC_Parental, total number of reads: 53420599
H3K36me3_ChIPseq_mESC_NSD1_KO, total number of reads: 47372438
HA-DNMT3A2_ChIPseq_mESC_Parental, total number of reads: 57826204
HA-DNMT3A2_ChIPseq_mESC_DNMT3A_KO, total number of reads: 48796383
HA-DNMT3A2_ChIPseq_mESC_NSD1_KO, total number of reads: 54961173
HA-DNMT3B_ChIPseq_mESC_Parental, total number of reads: 42085703

RNA-seq experiments: 125bp, paired-end.

C3H10T_Parental_RNAseq, total number of reads: 188681858

WGBS experiments: 150bp, paired-end.

WGBS_Parental, total number of reads: 887915518
WGBS_NSD_1_2_DKO, total number of reads: 872637038
WGBS_Cal27, total number of reads: 984952006
WGBS_Fadu, total number of reads: 955264598
WGBS_SCC_4, total number of reads: 1007047455
WGBS_SKN_3, total number of reads: 1008909116
WGBS_mESC_Parental, total number of reads: 749430350
WGBS_mESC_NSD1_KO, total number of reads: 812600584

Antibodies	For ChIP, the following primary antibodies were used: 100 μl Pierce anti-HA beads (Themo Scientific, 88836), anti-H3K4me1 (Abcam, ab8895), anti-H3K9me3 (Abcam, ab8898), anti-H3K27ac (Active Motif, 39133), anti-H3K27me3 (Cell Signaling Tech, 9733), anti-H3K36me2 (Cell Signaling, 2901) and anti-H3K36me3 (Active Motif, 61101).
Peak calling parameters	No peak calling was performed in this study.
Data quality	For ChIP-seq, Individual sequencing experiments were assessed for their percent of mapped reads (to hg19 and dm6, when applicable) to ensure proper coverage. Mapped reads with mapping quality were discarded from further analysis. Antibody pulldown efficacy was visually assessed by looking at tracks and comparing with spiked-in distribution of marks, when applicable. For DNA methylation, bi-sulfite conversion was confirmed by lambda-phage spike in and methylation levels of cytosines were considered. Cytosines with less than 5x coverage were discarded from further analysis.
Software	ChIP-seq reads were aligned with BWA version 0.7.17 and reads were summed across bins using bedtools version 2.22.1. Calling of broad histone PTM domains was performed using PELT. Gene body plots were generated using ngs.plot.r and annotated genomic regions within H3K36me2 domains were identified using a custom PERL script (available upon request) and reads in these regions were observed by featureCounts v1.5.3. Expected reads in the same regions were simulated using bedtools version 2.22.1. Data was visualized using IGV 2.3. RNA-seq reads were aligned using STAR version 2.5.3a and gene-level read counts were obtained from featureCounts v1.5.3. WGBS reads were aligned using BWA version 0.6.1 and methylation of individual CpGs was called using Samtools version 0.1.18.