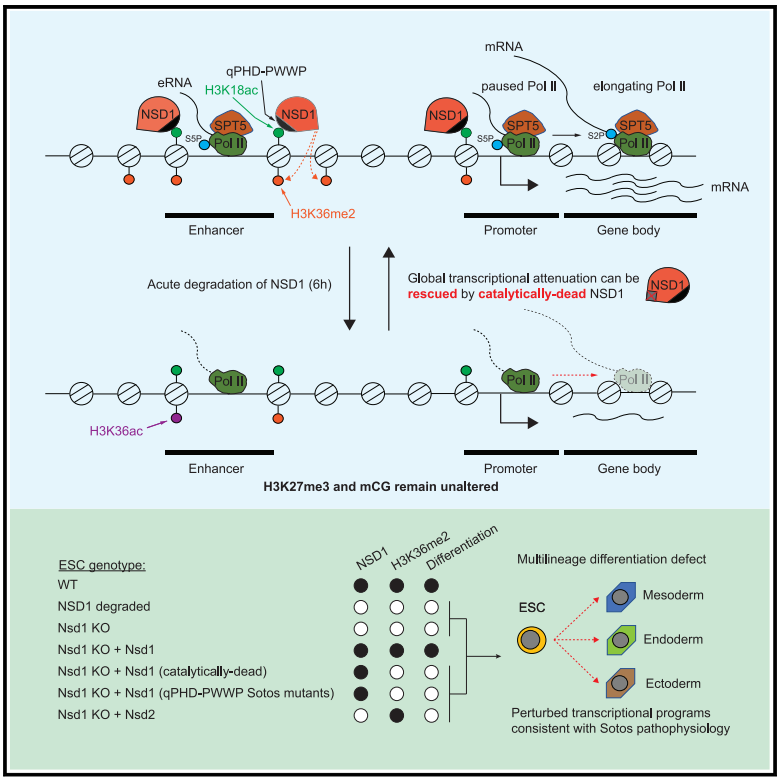
Molecular Cell

Article

Chromatin regulation of transcriptional enhancers and cell fate by the Sotos syndrome gene *NSD1*

Graphical abstract



Highlights

- NSD1 is enriched at active enhancers through a tandem qPHD-PWWP module
- NSD1 has a catalytic-independent transcriptional coactivator function
- NSD1 promotes enhancer activity and RNA Pol II promoter pause release
- NSD1 activates developmental transcriptional programs perturbed in Sotos syndrome

Authors

Zhen Sun, Yuan Lin, Mohammed T. Islam, ..., Thomas Vierbuchen, Charles L. Sawyers, Kristian Helin

Correspondence

sunz@mskcc.org (Z.S.), kristian.helin@icr.ac.uk (K.H.)

In brief

Sun et al. report an enhancer-enriched catalytic-independent coactivator function of NSD1, therefore revealing a direct mechanism of transcriptional regulation by NSD1 beyond its impact on H3K27me3 and DNA methylation through catalyzing H3K36me2. This work also sheds light on how NSD1 alterations contribute to Sotos syndrome development.

Sun et al., 2023, Molecular Cell *83*, 2398–2416
July 20, 2023
https://doi.org/10.1016/j.molcel.2023.06.007

CellPress

CellPress
OPEN ACCESS

Molecular Cell

Article

Chromatin regulation of transcriptional enhancers and cell fate by the Sotos syndrome gene *NSD1*

Zhen Sun,[1,2,3,*] Yuan Lin,[4,5] Mohammed T. Islam,[4,5,6] Richard Koche,[2] Lin Hedehus,[1,2,7] Dingyu Liu,[4,5,6] Chang Huang,[1,2,8] Thomas Vierbuchen,[4,5] Charles L. Sawyers,[3,9] and Kristian Helin[1,2,7,8,10,*]

[1]Cell Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[2]Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[3]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[4]Developmental Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[5]Center for Stem Cell Biology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[6]Louis V. Gerstner Jr. Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[7]Biotech Research and Innovation Centre, University of Copenhagen 2200 Copenhagen N, Denmark
[8]Division of Cancer Biology, Institute of Cancer Research, London SW7 3RP, UK
[9]Howard Hughes Medical Institute, Chevy Chase, MD, USA
[10]Lead contact
*Correspondence: sunz@mskcc.org (Z.S.), kristian.helin@icr.ac.uk (K.H.)
https://doi.org/10.1016/j.molcel.2023.06.007

SUMMARY

Nuclear receptor-binding SET-domain protein 1 (NSD1), a methyltransferase that catalyzes H3K36me2, is essential for mammalian development and is frequently dysregulated in diseases, including Sotos syndrome. Despite the impacts of H3K36me2 on H3K27me3 and DNA methylation, the direct role of NSD1 in transcriptional regulation remains largely unknown. Here, we show that NSD1 and H3K36me2 are enriched at *cis*-regulatory elements, particularly enhancers. NSD1 enhancer association is conferred by a tandem quadruple PHD (qPHD)-PWWP module, which recognizes p300-catalyzed H3K18ac. By combining acute NSD1 depletion with time-resolved epigenomic and nascent transcriptomic analyses, we demonstrate that NSD1 promotes enhancer-dependent gene transcription by facilitating RNA polymerase II (RNA Pol II) pause release. Notably, NSD1 can act as a transcriptional coactivator independent of its catalytic activity. Moreover, NSD1 enables the activation of developmental transcriptional programs associated with Sotos syndrome pathophysiology and controls embryonic stem cell (ESC) multilineage differentiation. Collectively, we have identified NSD1 as an enhancer-acting transcriptional coactivator that contributes to cell fate transition and Sotos syndrome development.

INTRODUCTION

The eukaryotic genome is presented in the form of chromatin, which impacts all DNA-templated events including transcription.[1,2] Transcriptional enhancers are distal *cis*-regulatory DNA elements that are critical for robust transcriptional output and precise spatiotemporal control of gene expression, which are fundamental for cell fate specification and organismal development.[3,4] Enhancer chromatin is engaged with a diverse array of transcription factors/coactivators and is dynamically modified by a plethora of histone post-translational modifications (PTMs),[5,6] e.g., H3K27ac and H3K4me1.[7–9] Furthermore, many coactivators themselves, e.g., MLL3/4, p300/CBP, and BRD4, bear "writing" and/or "reading" activities of histone PTMs conferred by specific domains.[10–15] However, despite recent advances in dissecting enhancer chromatin composition,[16–18] we still do not have a comprehensive understanding of how chromatin factors and modifications regulate enhancer activity.

Nuclear receptor-binding SET-domain protein 1 (NSD1) is the founding member of the NSD family of H3K36 methyltransferases that catalyze H3K36 mono- and di-methylation.[19] The NSD enzymes are large modular proteins that contain a catalytic module (AWS-SET-postSET) and multiple domains typically involved in chromatin association (PHD, PWWP, and C5HCH).[19,20] *In vitro* studies have shown that the catalytic module of NSD1 alone is sufficient for its ability and specificity in methylating H3K36 on nucleosome substrates.[21,22] However, the binding substrates of NSD1 chromatin reader domains and their contribution to NSD1 function remains incompletely characterized.[23,24] Although NSD1 was first identified as a nuclear receptor-interacting protein and putative transcriptional cofactor,[25,26] its direct role in gene transcription stays elusive. Moreover, the genome-wide distribution of NSD1 remains unknown, precluding the understanding of its role in gene regulation. H3K36me2 has been reported to occupy broad intergenic regions, to antagonize polycomb repressive complex 2 (PRC2)-catalyzed H3K27me3, and to promote DNA

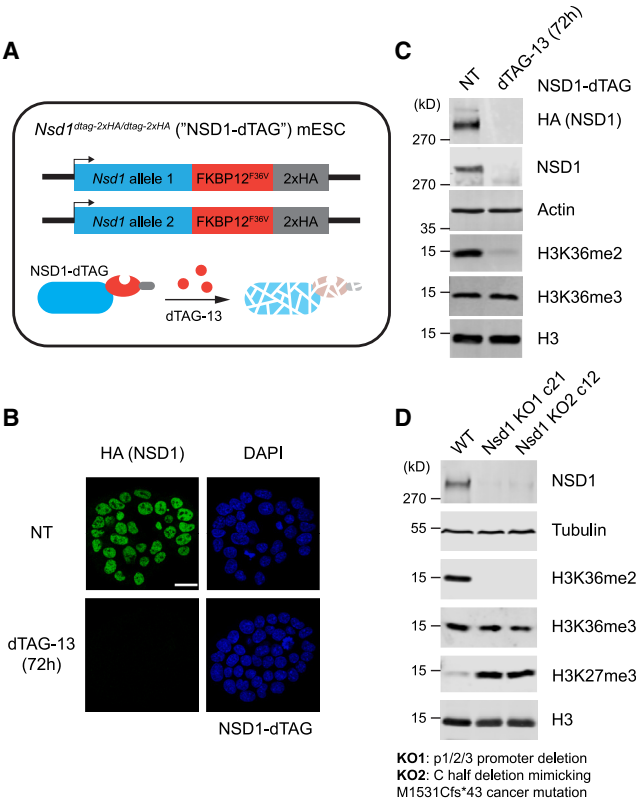


Figure 1. NSD1 catalyzes the majority of H3K36me2 in mESCs

(A) Schematic showing the NSD1-dTAG alleles and targeted proteolysis.

(B) Immunofluorescence in NSD1-dTAG mESCs. Scale bar represents 20 μm applies to all panels. In (B) and (C), NT, no treatment.

(C and D) Western blots of whole-cell extracts from indicated mESCs. Actin, tubulin, and H3 used as loading controls.

See also Figure S1.

methylation.[27–35] However, beyond the context of its interplay with other chromatin modifications, the functional impact of H3K36me2 on transcription remains poorly understood.

NSD1 is essential for mammalian development demonstrated by gastrulation defects and embryonic lethality before E10.5 in *Nsd1* knockout mice.[36] In humans, germline *NSD1* haploinsufficiency leads to Sotos syndrome (OMIM 117550), a multisystemic developmental disorder, with remarkably high penetrance.[37] Sotos syndrome is characterized by childhood overgrowth, intellectual disability, and facial dysmorphism, and patients show a spectrum of abnormalities, including cardiac anomalies, renal anomalies, advanced bone age, joint hyperlaxity, and scoliosis.[37] Additionally, Sotos syndrome is associated with increased cancer risk,[37,38] consistent with the presence of *NSD1*-inactivating mutations in a variety of cancers (Figure S1A).[39] Despite a strong link between NSD1 and early development, the transcriptional programs regulated by NSD1 during cell fate specification and the underlying mechanism remain ill-defined, hindering the understanding of how *NSD1* alterations lead to Sotos syndrome. Moreover, although Sotos syndrome pathogenic mutations are concentrated at chromatin regulatory domains of NSD1,[37,38,40,41] their impact on NSD1 molecular function and Sotos syndrome development has not been determined.

In this study, we report on the discovery of a catalytic-independent transcriptional coactivator function of NSD1 that facilitates enhancer activity and RNA polymerase II (RNA Pol II) pause release, mechanisms of its chromatin association and enrichment at enhancers, and its regulation of developmental gene expression programs associated with Sotos syndrome pathogenesis.

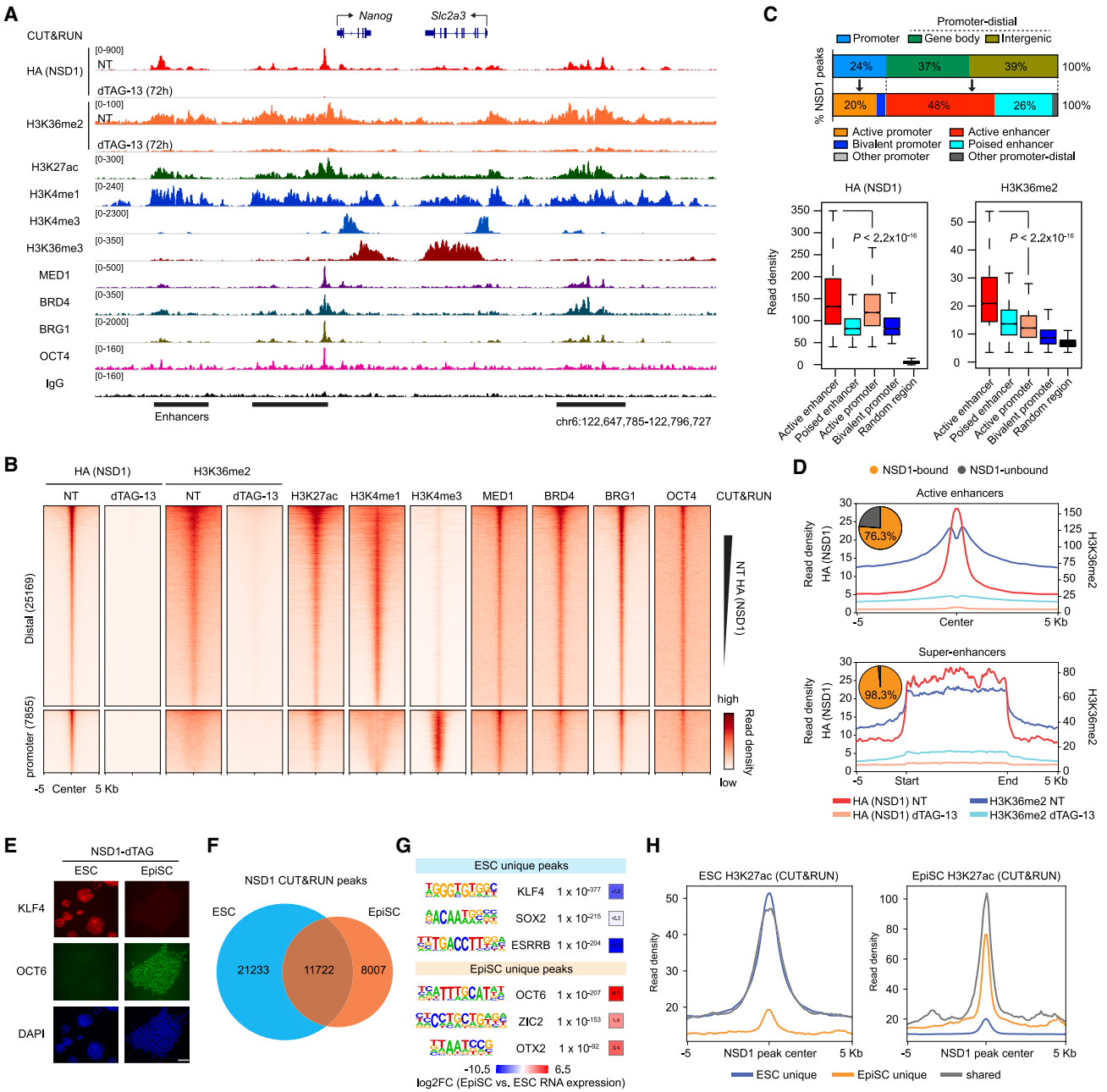
RESULTS

NSD1 catalyzes the majority of H3K36me2 in mESCs

To investigate the molecular functions of NSD1, we examined its contribution to H3K36me2/3 in mouse embryonic stem cells (mESCs). We established NSD1 loss-of-function systems through both targeted protein degradation and genetic ablation. Using CRISPR-mediated knockin, we generated "NSD1-dTAG (degradation tag)" mESCs, where both alleles of *Nsd1* are fused with sequences encoding the $FKBP12^{F36V}$ dTAG.[42] This allows for an acute and selective depletion of endogenous NSD1 protein upon treatment with the dTAG-13 molecule (Figure 1A). Simultaneously, we knocked in the 2xHA tag as a molecular handle (Figure 1A). Although dTAG knockin did not affect the expression levels of endogenous NSD1 (Figure S1B), dTAG-13 treatment induced near-complete NSD1 degradation (Figures 1B, 1C, and S1C).

We utilized two strategies to constitutively knockout *Nsd1* in mESCs. First, we abolished the transcription of endogenous *Nsd1* through promoter deletion. Interestingly, we detected four NSD1 isoforms through the C-terminal 2xHA tag, with the major one being the less-characterized NSD1.1S[43] (Figure S1C). By integrating transcription start site (TSS) annotation[44] and estimated sizes of the isoforms, we found that these four isoforms arise from three alternative promoters. The canonical promoter ("p1") produces the two longer isoforms (*Nsd1.1S* and *Nsd1.1L*), whereas two "intergenic" promoters, "p2" and "p3," give rise to the shorter isoforms *Nsd1.2* and *Nsd1.3*, respectively (Figures S1C and S1D). The most abundant ESC isoform NSD1.1S is generated through the exclusion of an alternative intron within exon 2 and largely lacks the N-terminal unstructured region compared to the canonical isoform NSD1.1L (Figures S1E and S1F). To abolish the expression of all isoforms, we performed serial CRISPR-mediated deletion of p1 to p3 (Figure S1D), which showed that p1, p2, and p3 produce approximately 60%, 20%, and 20% of Nsd1 transcripts, respectively, in mESCs (Figure S1G). Triple promoter knockout (*Nsd1* "KO1") nearly abrogated NSD1 transcripts and proteins (Figures S1G and S1H). Alternatively, we used CRISPR editing to delete a large genomic region encoding a cluster of NSD1 C-terminal domains, including the PWWP, PHD, and SET domains (*Nsd1* "KO2") (Figure S1D). Such deletion functionally mimics a recurrent cancer-associated frameshift mutation, M1531Cfs*43, and led to the expression of a truncated NSD1 (Figures S1A and S1H).

The loss of NSD1 expression, by either NSD1 degradation in NSD1-dTAG cells or deletion of the three *Nsd1* promoters in *Nsd1* KO1 cells, and the functional inactivation of NSD1 in *Nsd1* KO2 cells did not impact the self-renewal capability of mESCs, with only moderately reduced proliferation in *Nsd1* KO1/2 cells (Figures S1I and S1J). However, loss of NSD1 nearly abrogated



(legend on next page)

H3K36me2 without affecting H3K36me3 levels (Figures 1C, 1D, and S1H). These results demonstrate that NSD1 is the predominant H3K36 di-methyltransferase in mESCs.

NSD1 and H3K36me2 are enriched at active enhancers

To determine the genome-wide distribution of NSD1 and H3K36me2, we utilized the NSD1-dTAG mESCs and performed cleavage under targets and release using nuclease (CUT&RUN).[45] NSD1 and H3K36me2 levels were diminished genome-wide upon NSD1 degradation (72 h) (Figures 2A and 2B), validating that NSD1 deposits the bulk of H3K36me2 in mESCs and demonstrating high specificity in our system. To comprehensively annotate the epigenome, we performed CUT&RUN for an array of histone PTMs, transcription factors, and coactivators. Strikingly, both NSD1 and H3K36me2 were enriched at *cis*-regulatory elements, particularly active enhancers defined by H3K4me1 and H3K27ac, where they colocalized with transcription (co)factors Mediator (MED1), BRD4, BAF complex (BRG1), and OCT4[7–9,13,46] (Figures 2A–2C, S2A, and S2B). Furthermore, NSD1 and H3K36me2 showed a distribution distinct from H3K4me3 and H3K36me3, which demarcate active promoters and gene bodies, respectively (Figures 2A, 2B, and S2A), and a global anti-correlation with repressive histone modifications H3K27me3 and H3K9me3 (Figure S2B). Similar to other transcription coactivators, NSD1 showed sharp binding peaks at active enhancers including constituents of super-enhancers,[46] whereas H3K36me2 displayed broader distribution with enrichment at active enhancers (Figures 2A and 2D). Notably, we observed NSD1 peaks at 76.3% of active enhancers and nearly all super-enhancers (Figure 2D), demonstrating that NSD1 occupancy is a prevalent feature of active enhancers. Moreover, NSD1-bound enhancers showed elevated histone acetylation levels and coactivator occupancy, indicating higher activity (Figure S2C).

Next, we queried whether the distribution of NSD1 correlates with enhancer activity during cell fate transition. We converted NSD1-dTAG ESCs to epiblast stem cells (EpiSCs)[47] (Figures 2E and S2D–S2F), a process representing epiblast maturation from naive to primed pluripotency during early development when dramatic enhancer reprogramming takes place.[48] NSD1.1S remained the dominant isoform in EpiSCs (Figure S2G). We observed a profound genomic redistribution of NSD1 during ESC-to-EpiSC transitions (Figure 2F). Of note, ESC- and EpiSC-specific NSD1-binding sites show motif enrichment of key transcription factors of naïve and primed pluripotency, respectively (Figure 2G). Moreover, such redistribution strongly correlated with the reprogramming of enhancer activity indicated by H3K27ac levels (Figures 2H and 2I). These results demonstrate that NSD1 binding is dynamically regulated during the transition of cellular identity to occupy active enhancers.

The tandem qPHD-PWWP module is required for NSD1 chromatin association and enhancer enrichment

To identify domains of NSD1 that regulate its chromatin association, we first tested their ability to support global H3K36me2 levels. We generated a series of NSD1 truncation mutants (Figures 3A and S3A) and re-expressed them in *Nsd1* KO mESCs. We found that the NSD1 "C-terminal half" (C) fragment, which is structurally similar to the NSD1.3 isoform, was sufficient to rescue global H3K36me2 (Figures 3B and S3B). In contrast, the "N-terminal half" (N) fragment, which resembles the truncated NSD1 caused by the M1531Cfs*43 mutation in cancer, could not rescue H3K36me2 levels (Figure S3B). Furthermore, PHD1–4 and PWWP2 were each required and together sufficient, in the presence of the catalytic domain, to rescue global H3K36me2 (Figure 3B). In contrast, PWWP1, PHD5, C5HCH, and most unstructured regions were dispensable to deposit global H3K36me2 (Figure 3B). To understand whether PHD1–4 and PWWP2 domains contribute to NSD1 chromatin association, we then performed sequential salt extraction. NSD1 C without PHD1–4 and/or PWWP2, although nuclear localized, showed a pronounced reduction of chromatin affinity as demonstrated by elution at no/low salt conditions (Figures 3C, S3C, and S3D). We next investigated the role of PHD1–4 and PWWP2 in NSD1 genomic occupancy and performed CUT&RUN. We showed that NSD1 C recapitulated the enhancer-enriched binding pattern of endogenous NSD1 (Figures 3D, S3E, and S3F). In contrast, NSD1 C without PHD1–4 and/or PWWP2 showed a dramatic decrease of occupancy (Figures 3D, 3E, S3G, and S3H). The contribution of PHD1–4 and PWWP2 in enhancer association was further validated with full-length (FL) NSD1 lacking this tandem module (Figures S3I–S3M).

To identify the chromatin substrate of the tandem quadruple PHD (qPHD)-PWWP module, we performed nucleosome pull-down assays using nuclear extracts from *Nsd1* KO2 cells expressing a NSD1 fragment comprised the qPHD-PWWP module and the catalytic module ("PP-SET") and a panel of recombinant nucleosomes with distinct histone PTMs. Notably, we observed a strong selective interaction between PP-SET and H3K18ac, enhancer-enriched acetylation catalyzed by p300/CBP,[49,50] as well as a weak ubiquitous nucleosome affinity largely independent of PTM status (Figure 3F). The fragment also interacted

Figure 2. NSD1 and H3K36me2 are enriched at active enhancers

(A) Genome browser view of CUT&RUN profiles in NSD1-dTAG mESCs. Black bars represent active enhancers. In (A), (B), and (D), 72 h dTAG-13 treated cells were used as control; NT, no treatment.

(B) Heatmap showing enrichment of chromatin features at all NSD1-occupied promoter and promoter-distal regions.

(C) Upper: genomic distribution of NSD1. Lower: relative enrichment of NSD1/H3K36me2 at distinct genomic regions. p values calculated using Welch's t-test.

(D) Average profile of NSD1 and H3K36me2 at active enhancers (n = 15,986) and super-enhancers (n = 705). Inset showing percentage occupied by NSD1.

(E) Immunofluorescence of pluripotency markers KLF4 (naive) and OCT6 (primed). Scale bar represents 20 μm applies to all panels.

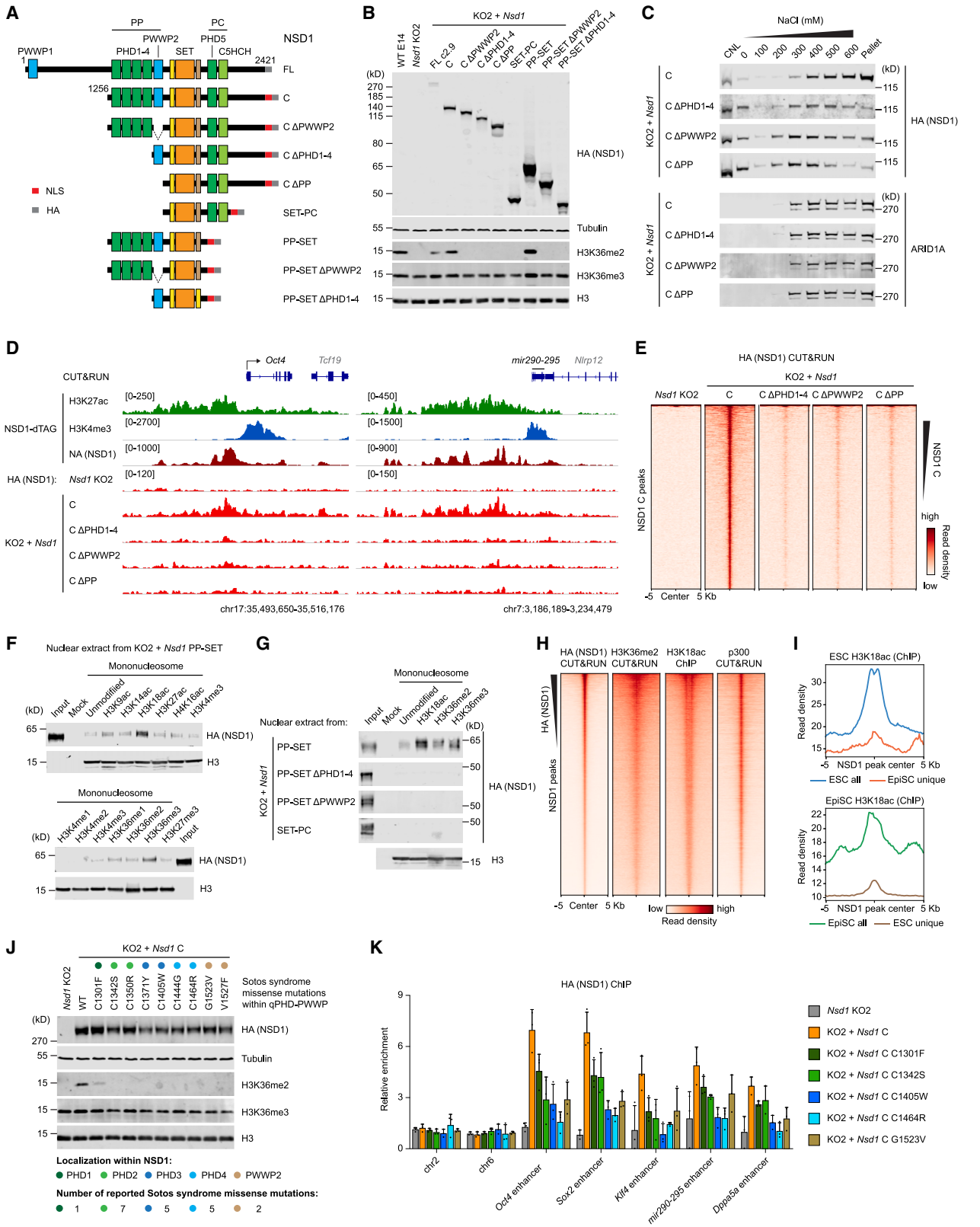
(F) Overlap of NSD1 peaks in ESCs vs. EpiSCs.

(G) Top three *de novo* motifs enriched at ESC and EpiSC unique NSD1 peaks (target coverage >10%; ranked by p values). Heatmap shows relative expression.

(H) Average profile showing H3K27ac occupancy in ESCs and EpiSCs at indicated categories of NSD1 peaks.

(I) Genome browser view of CUT&RUN profiles in NSD1-dTAG ESCs vs. EpiSCs.

See also Figure S2.



(legend on next page)

moderately with H3K36me2/3 (Figures 3F and 3G). Moreover, deletion of either PHD1–4 or PWWP2 abolished nucleosome interaction (Figure 3G), suggesting that both are required for basal nucleosome affinity, which in turn allows for PTM selectivity. Additionally, a fragment consisting of the catalytic module, PHD5, and C5HCH ("SET-PC") was devoid of nucleosome affinity. Together, these results suggest that the NSD1 qPHD-PWWP module engages in multivalent chromatin interaction with specificity toward H3K18ac. Consistent with a role in NSD1 recruitment, H3K18ac and p300 showed genome-wide co-enrichment with NSD1 and H3K36me2 in mESCs (Figure 3H), and H3K18ac showed genome-wide redistribution correlating with NSD1 localization during ESC-to-EpiSC transition (Figure 3I).

Interestingly, the qPHD-PWWP module is a hotspot for Sotos syndrome missense mutations, and at least twenty pathogenic variants across the five constituent domains have been reported (Figure 3J).[37,38,40,41,51,52] Of note, all pathogenic variants were largely unable to rescue global H3K36me2 levels when expressed in *Nsd1* KO cells (Figure 3J). Moreover, Sotos mutation at each constituent domain led to reduced affinity with bulk chromatin (Figure S3N) and decreased occupancy at enhancers (Figure 3K), suggesting that the five domains cooperate to drive NSD1 chromatin binding. Together, these results demonstrate that the tandem qPHD-PWWP module is critical for NSD1 chromatin recruitment and enrichment at active enhancers.

Impact of acute NSD1 degradation on H3K36 modifications and associated chromatin features

To investigate the kinetics of NSD1 and H3K36me2 loss using our degron system, we treated NSD1-dTAG mESCs with dTAG-13, which revealed that NSD1 reached near-complete degradation within 1 h of treatment (Figure 4A), whereas other enhancer associated proteins remained unaltered (Figure S4A), demonstrating target selectivity. Global levels of H3K36me2 steadily decreased following the rapid loss of NSD1, and we observed a marked reduction of H3K36me2 as early as 6 h after dTAG-13 treatment, with the residual methylation largely abrogated by 18–24 h (Figure 4A). Consistent with the immunoblot analysis, CUT&RUN analysis revealed a genome-wide NSD1 loss by 1 h of ligand administration (Figures 4B, 4C, and S4B), although occupancy of Mediator and BRD4 remained stable by 24 h of dTAG-13 treatment (Figures 4D and 4E). The genome-wide occupancy of H3K36me2 was only mildly decreased by 1 h, but significantly depleted at 6 h and further diminished at 24 h following NSD1 degradation (Figures 4F and S4C). In comparison, H3K27ac and H3K4me1 were not affected by NSD1 loss (Figures 4G and 4H). Next, we examined the effect of NSD1 degradation on other H3K36 modifications. H3K36me3 levels and genomic distribution, particularly at active genes, remained stable in response to NSD1 degradation and H3K36me2 depletion (Figures 4A, 4I, S4D, and S4E). Interestingly, following H3K36me2 loss, we observed a reciprocal increase of H3K36ac across regions bound by NSD1 (Figures S4F–S4H), which was abolished by the pharmacological inhibition of p300/CBP (Figure S4I).[49] Additionally, chromatin accessibility as measured by assay for transposase-accessible chromatin using sequencing (ATAC-seq) remained unaltered within 24 h of NSD1 degradation (Figure 4J), and H3K4me3 showed moderate reduction at promoters by 6 h of NSD1 loss (Figure 4K).

H3K36me2 has been reported to antagonize the deposition of H3K27me3 and promote DNA methylation,[27–35] but the kinetics of such interplays in cells has not been reported. Therefore, we examined H3K27me3 and DNA methylation levels and distribution following acute degradation or constitutive knockout of NSD1. Interestingly, although *Nsd1* knockout cells displayed a global gain of H3K27me3, H3K27me3 remained unaltered for 48–72 h after NSD1 degradation and loss of H3K36me2 (Figures 1D, 4A, 4L, 4M, S4J, and S4K). Prolonged degradation of NSD1 also led to a global increase of H3K27me3 levels (Figure S4L). Similarly, acute loss of NSD1 was largely inconsequential to global DNA methylation as determined by mass spectrometric quantification and reduced representation bisulfite sequencing (RRBS), in contrast to a hypomethylated state in *Nsd1* knockout cells (Figures 4N, 4O, and S4M). These data indicate that accumulation of H3K27me3 and reduction of DNA methylation occurs gradually across multiple cell cycles following H3K36me2 loss.

Acute depletion of NSD1 leads to reduced gene transcription independent of its catalytic activity

To study the immediate consequences of NSD1 and H3K36me2 loss on gene transcription, we quantified newly synthesized RNA following NSD1 degradation using Thiol(SH)-linked alkylation for the metabolic sequencing of RNA (SLAM-seq)[53] (Figure 5A). Notably, we observed a substantial and asymmetric effect in response to NSD1 degradation: a global reduction of newly synthesized RNA was detected after 6 h of dTAG-13 treatment, preceded by a moderate yet significant decrease after 1 h (Figures 5B, S5A,

Figure 3. The tandem qPHD-PWWP module is essential for NSD1 chromatin association and enhancer recruitment

(A) Schematic of NSD1 (NSD1.1S) series. FL, full-length; PP, qPHD-PWWP; PC, PhD-C5HCH; C, C-terminal half. NLS, nuclear localization sequence.

(B) Western blots of whole-cell extracts from indicated mESCs. In (B) and (J), tubulin and H3 used as loading controls.

(C) Western blots of sequential salt extraction fractions. ARID1A used as control. CNL, cytoplasmic and nuclear-leaked fraction.

(D) Genome browser view showing occupancy of NSD1 C series.

(E) Heatmap showing occupancy of NSD1 C series at all NSD1 C peaks (n = 2,036).

(F and G) Western blot analysis of nucleosome pull-down assay using recombinant nucleosomes and mESC nuclear extracts. H3 used as input control for bait nucleosomes.

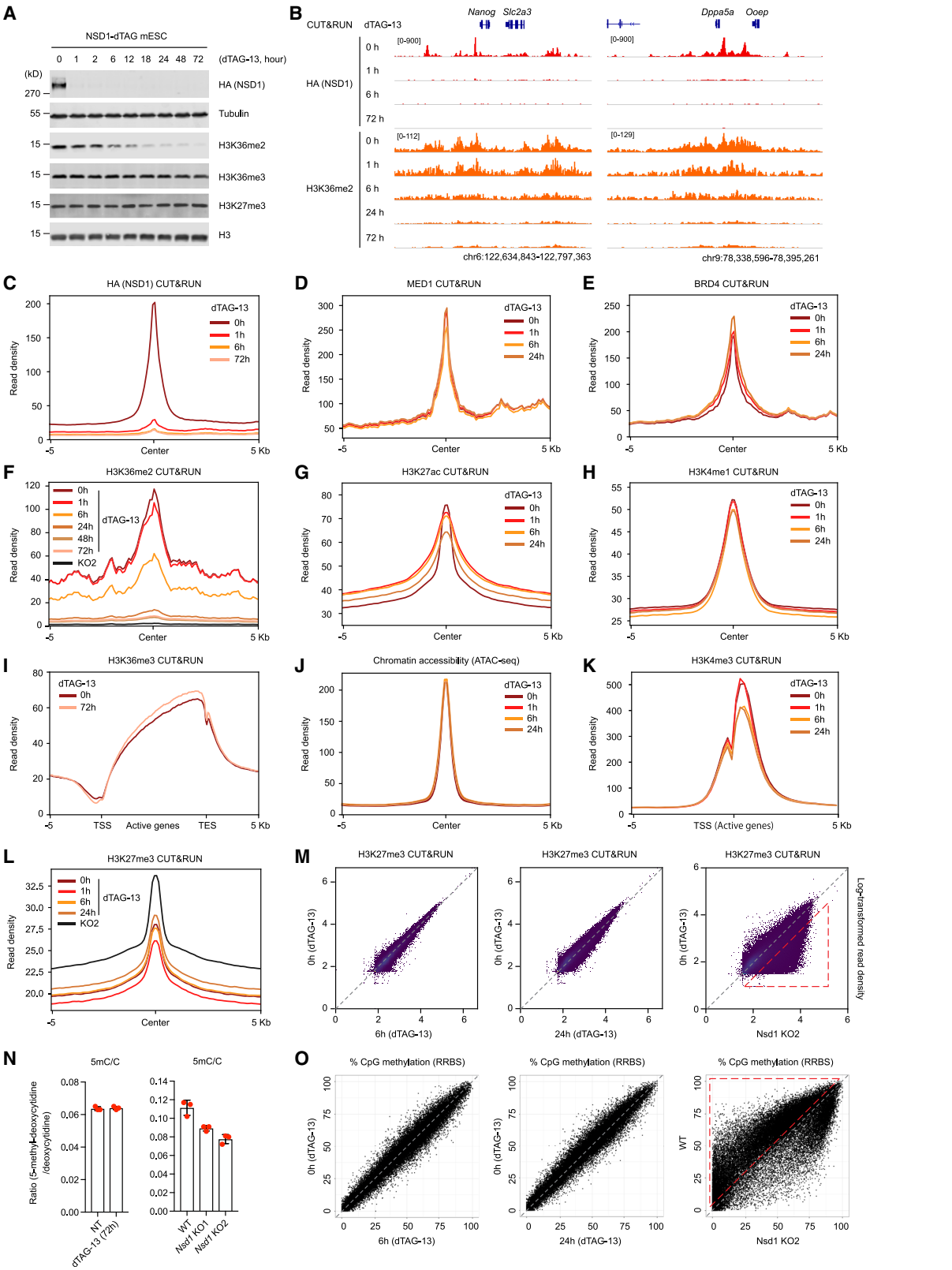
(H) Heatmap showing occupancy of indicated proteins and histone PTMs.

(I) Average profile showing H3K18ac occupancy in ESCs and EpiSCs at indicated categories of NSD1 peaks.

(J) Western blots of whole-cell extracts from indicated mESCs.

(K) Chromatin immunoprecipitation (ChIP)-qPCR analysis of relative enrichment of NSD1 C Sotos syndrome mutants at enhancers. Negative control regions (*chr2* and *chr6*) and input were used for normalization; data represent mean ± SD from n = 3 technical replicates.

See also Figure S3.



(legend on next page)

and S5B). Global nascent transcriptional changes were confirmed using transient transcriptome sequencing (TT_{chem}-seq) (Figure S5C). Moreover, NSD1 occupied the *cis*-regulatory regions of most of the downregulated genes (Figure 5C), and genes associated with super-enhancers, which are highly enriched for NSD1 occupancy, were particularly susceptible to NSD1 loss (Figure 5C). Taken together, these results indicate that NSD1 is directly involved in regulating gene transcription.

Although the reduction of transcriptional activity upon NSD1 degradation overall correlated with the loss of H3K36me2 from 0 to 6 h, we nevertheless found that genes with earlier downregulation (1 h) did not show faster loss of H3K36me2 (Figures S5D and S5E) and that prolonged NSD1 depletion (24 h), albeit leading to a further and near-complete loss of H3K36me2, did not result in additional transcriptional downregulation (Figure S5F). Therefore, we queried whether the catalytic activity is required for NSD1 to promote transcription. To this end, we ectopically expressed wild-type or the N1751Q catalytically dead mutant (also known as N1918Q[24]) in NSD1-dTAG mESCs and then induced degradation of the endogenous NSD1 (Figure 5D). Strikingly, both wild-type and catalytically dead NSD1 rescued the transcriptional attenuation following endogenous NSD1 loss (Figures 5E–5H).

To support this observation using an orthogonal assay, we employed a dCas9 activation system and utilized dCas9-$NSD1^{C}$ and/or dCas9-$p300^{core}$ fusions to target the *HS2* enhancer of the silent hemoglobin gamma A (*HBG1*) gene in HEK293T cells (Figures 5I, S5G, and S5H). Although dCas9-$NSD1^{C}$ alone only moderately turned on *HBG1* expression, it significantly enhanced the ability of dCas9-$p300^{core}$ to activate the transcription of *HBG1* (Figure 5I), suggesting cooperativity. Importantly, catalytically dead dCas9-$NSD1^{C}$ (N1751Q) maintained the transactivation activity in both contexts, in contrast to enzymatically deficient dCas9-$p300^{core}$ (D1399Y). Collectively, these results indicate that the catalytic activities of NSD1 and, therefore, H3K36me2 are dispensable for its primary coactivator function.

Next, we compared the transcriptional consequences upon NSD1 acute depletion with those following degradation of SETD2, which catalyzes H3K36me3. Although SETD2 has been suggested to regulate transcriptional elongation, direct evidence is lacking.[19] Therefore, we generated SETD2-dTAG knockin mESCs as for NSD1 (Figures 5J and S5I). We showed that SETD2 occupied active promoters and gene bodies, similar to RNA Pol II (Figures S5J and S5K), consistent with it being a RNA Pol II-interacting protein.[19] H3K36me3 was enriched at transcribed gene bodies and show negligible levels at intergenic enhancers (Figures 5K and S5L). Loss of SETD2/H3K36me3 upon dTAG-13 treatment displayed similar kinetics to that of NSD1/H3K36me2 in NSD1-dTAG cells (Figures 5J and 5K). However, in stark contrast to NSD1 loss, SETD2 degradation resulted in very minor effects on the global synthesis of new polyadenylated RNA (Figures 5L, S5M, and S5N). Together, these data demonstrate that NSD1 and SETD2 have distinct genomic distribution and roles in regulating gene expression.

NSD1 promotes enhancer activity and RNA Pol II promoter pause release

We next investigated the mechanisms underlying NSD1-regulated gene transcription. We observed that in addition to the enhancer-enriched distribution, genes with enhancer-bound NSD1 showed more profound downregulation upon NSD1 depletion than those with only promoter-bound NSD1 (Figures 6A and S6A) and that genes associated with more NSD1 binding sites were more dependent on NSD1 (Figure 6A), suggesting a distal mode of transcriptional regulation. Therefore, we next examined enhancer activity following NSD1 degradation. H3K27ac remained largely unaltered following acute NSD1 loss, similar to observations following acute depletion of the Mediator backbone MED14.[54] Since transcription is a prevalent feature of active enhancers and sensitive readout of enhancer strength, we utilized transcriptionally engaged RNA Pol II to measure enhancer activity.[5,11,54,55] Specifically, we measured RNA Pol II with serine-5 phosphorylation (S5P) at the C-terminal domain (CTD), a form associated with transcriptional initiation.[56] At 6 h following dTAG-13 exposure, we observed a pronounced reduction of RNA Pol II S5P at active enhancers including super-enhancers (Figures 6B and 6C). Consistently, enhancer RNA (eRNA) production was dampened upon NSD1 loss as measured by TT_{chem}-seq (Figure S6B). Together with our observation that dCas9-mediated NSD1 targeting can facilitate gene activation from distal enhancers, these results demonstrate that NSD1 is a coactivator that promotes enhancer activity.

We then examined RNA Pol II distribution at genes upon NSD1 loss. We observed augmented RNA Pol II at promoters and decreased occupancy downstream at the gene body, with a globally increased pausing index (Figures 6D and 6E). In accordance, occupancy of elongating RNA Pol II, defined by CTD serine-2 phosphorylation (S2P),[56] was reduced throughout transcribed regions (Figure 5H). These results suggest a defect in RNA Pol II promoter pause release and reduced productive elongation after NSD1 degradation. We therefore examined the occupancy of factors that regulate RNA Pol II pause establishment, pause release, and elongation[57] (Figure 6G). Although NELF

Figure 4. Impact of acute NSD1 degradation on H3K36 modifications and associated chromatin features

(A) Western blots of whole-cell extracts. Tubulin and H3 used as loading controls.

(B) Genome browser view showing NSD1 and H3K36me2 occupancy in NSD1-dTAG mESCs.

(C–L) Average profiles of indicated CUT&RUN and ATAC-seq signals in NSD1-dTAG mESCs at their respective peaks or all active genes (RPKM > 1) as indicated. *Nsd1* KO2 cells used as control.

(M) Scatterplots showing genome-wide correlation of H3K27me3 occupancy in NSD1-dTAG and *Nsd1* KO2 mESCs. Gain of H3K27me3 was highlighted.

(N) Mass spectrometry analysis of global DNA methylation. Data represent mean ± SD from n = 3 independent culture.

(O) Scatterplots showing genome-wide correlation of CpG methylation levels in NSD1-dTAG, WT and *Nsd1* KO2 mESCs. Loss of CpG methylation was highlighted.

See also Figure S4.

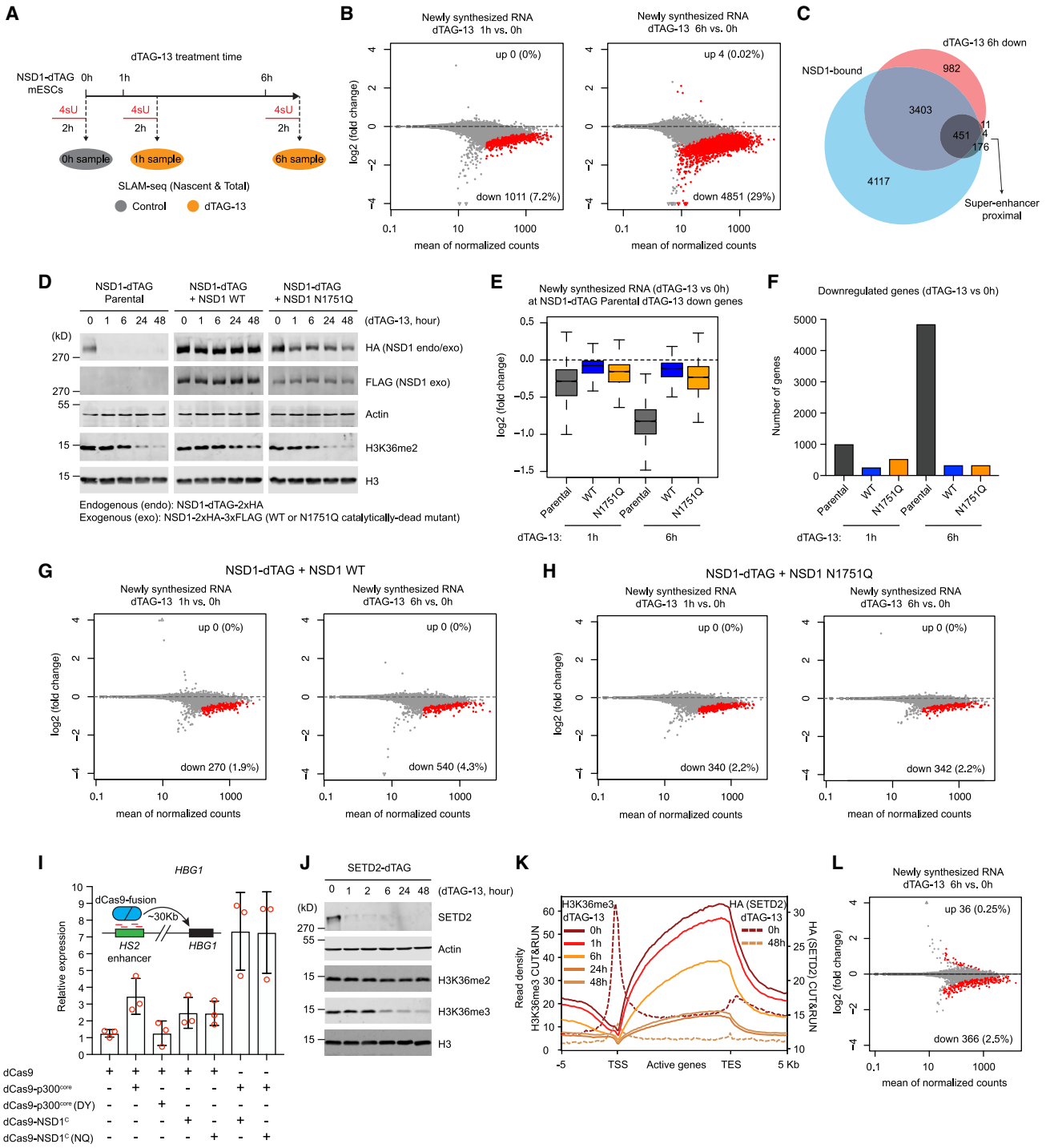


Figure 5. Acute depletion of NSD1 leads to reduced gene transcription independent of its catalytic activity

(A) Schematic of SLAM-seq analysis of newly synthesized RNA following NSD1 degradation.

(B) MA plots showing differential gene expression analysis in NSD1-dTAG mESCs. In (B), (G), (H), and (L), red dots represent genes significantly deregulated ($q < 0.05$); number of genes up- or downregulated and percentage among all genes with detectable expression are shown; n = 3 independent treatments.

(C) Overlap of genes downregulated by 6 h of dTAG-13 treatment as in (B), genes bound by NSD1 at enhancers and/or TSS, and all super-enhancer proximal genes.

(D) Western blots of whole-cell extracts from indicated mESCs. In (D) and (J), actin and H3 used as loading controls.

(E) Magnitude of transcriptional downregulation upon dTAG-13 treatment.

(F) Number of genes significantly downregulated ($q < 0.05$) upon dTAG-13 treatment.

(legend continued on next page)

binding remained stable, we found a striking decrease in SPT5 association at enhancers/promoters and throughout the gene body (Figures 6H, 6I, and S6C). PAF1 and SPT6 occupancies were also broadly reduced, although to a milder extent (Figures 6J, 6K, and S6C). Since SPT5, PAF1, and SPT6 are all components of an activated transcription elongation complex and critical for RNA Pol II pause release,[57–64] these results suggest that NSD1 promotes productive gene transcription through facilitating recruitment of pause-release/elongation factors, upholding enhancer activity, and stimulating enhancer-regulated RNA Pol II pause release.

NSD1 facilitates the activation of developmental transcriptional programs

Since NSD1 plays a critical role in gene transcription in mESCs and its loss is embryonic lethal in mice, we were interested in extending our analysis of NSD1 function to mESCs undergoing differentiation. NSD1 is largely dispensable for naive to primed pluripotency transition during ESC to epiblast-like cell (EpiLC) conversion (Figure S7A), consistent with the observation that *Nsd1*$^{-/-}$ embryos can form post-implantation epiblast but fail to properly undergo gastrulation.[36] To further examine the function of NSD1 in post-implantation development, we induced multilineage differentiation using the embryoid body (EB) formation assay with NSD1-dTAG mESCs, which can activate lineage markers to a similar extent as WT mESCs (Figure S7B). NSD1 degradation during EB differentiation led to a severe decrease in the induction of lineage markers, without affecting pluripotency exit (Figures 7A and S7C). Furthermore, global gene expression analyses revealed that NSD1 degradation during differentiation resulted in significant transcriptional changes, with the majority being strongly downregulated (Figures 7B, S7D, and S7E). More than 80% of the downregulated genes were induced during normal EB differentiation (Figures 7C and S7F), indicating compromised activation of cell identity genes without NSD1. In accordance, gene ontology (GO) analysis showed that NSD1-dependent genes during differentiation were enriched for regulators of developmental processes (Figure 7D). Notably, many of the altered processes are related to the development of organ systems implicated with major Sotos syndrome anomalies, such as the nervous, cardiovascular, genitourinary, and skeletal systems (Figure 7D). We further validated the cardiac development defect indicated by gene expression signatures and showed that mESCs failed to differentiate into cardiomyocytes (Figure 7E) and to form contracting EBs (Figure 7F) upon NSD1 degradation. We also validated the neural specification defect by directing NSD1-dTAG EpiSCs to differentiate into dorsal forebrain-patterned organoids.[47] NSD1 degradation led to altered organoid morphology and downregulation of markers of neurons across multiple brain regions as well as markers of oligodendrocyte progenitors (Figures 7G, S7G, and S7H).

We further queried whether compromised induction of developmental genes upon NSD1 loss is associated with impaired enhancer activation. In agreement with our previous results, NSD1 and H3K36me2 occupancy decreased significantly in EBs formed in the presence of dTAG-13 at *de novo* enhancers associated with NSD1-dependent genes (Figure 7H). Moreover, NSD1 degradation during EB differentiation led to a significant reduction of H3K27ac and chromatin accessibility at these developmental enhancers (Figure 7H), indicating compromised enhancer activation. Of note, we did not observe significant changes in H3K27me3 levels at these enhancers (Figure 7H) despite a global increase (Figure S7C), suggesting that H3K36me2-mediated H3K27me3 antagonism did not directly contribute to enhancer activation in this context. Taken together, these results suggest that NSD1 facilitates the induction of developmental genes by promoting enhancer activation during ESC differentiation.

Last, we investigated how functional domains of NSD1 contribute to its regulation of ESC differentiation. Consistent with our findings in the NSD1-dTAG system, *Nsd1* KO mESCs could not turn on key differentiation genes or to form contracting EBs (Figures S7J and S7K). These defects could be rescued by NSD1 FL or C fragment but not by the NSD1 N fragment (Figure S7L). Moreover, neither the qPHD-PWWP truncation nor the N1751Q catalytically dead mutant rescued such defects, in the context of either the FL protein (Figures 7I, 7J, S3A, and S3J) or a C-terminal fragment (Figures S3A, S3B, and S7M). Strikingly, a Sotos syndrome allelic series with missense mutations across the qPHD-PWWP module was largely incapable of rescuing the differentiation defects in *Nsd1* KO cells (Figures 7K and 7L), demonstrating that the qPHD-PWWP module pathological variants are loss-of-function mutants in both biochemical and physiological activities. Notably, although NSD2 overexpression was able to rescue the balance of global H3K36me2 and H3K27me3 levels in *Nsd1* KO mESCs (Figures 7M, 7N, S3A, S3B, and S7N), it failed to rescue the differentiation defects (Figure 7O), indicating a contribution of the catalytic-independent coactivator function of NSD1 in ESC differentiation. Together, these results reveal a critical role of NSD1 in ESC differentiation and suggest that its loss contributes to Sotos syndrome development by impaired activation of cell identity genes.

DISCUSSION

More than two decades after the initial discovery of NSD1 as a nuclear receptor-interacting protein,[25,26] its role in transcriptional regulation remains poorly understood. Our results demonstrate that NSD1 acts as an enhancer-enriched coactivator,

(G and H) MA plots showing differential gene expression analysis in indicated mESCs.
(I) Reverse transcription (RT)-qPCR of *HBG1* expression in HEK293T cells expressing different combinations of dCas9 fusions. *GAPDH* used as internal reference. Data represent mean ± SD from n = 3 technical replicates. DY, D1399Y; NQ, N1751Q.
(J) Western blots of whole-cell extracts from SETD2-dTAG mESCs.
(K) Average profiles of SETD2 and H3K36me3 CUT&RUN signal at active genes (RPKM > 1) in SETD2-dTAG mESCs.
(L) MA plots showing differential gene expression analysis in SETD2-dTAG mESCs.
See also Figure S5.

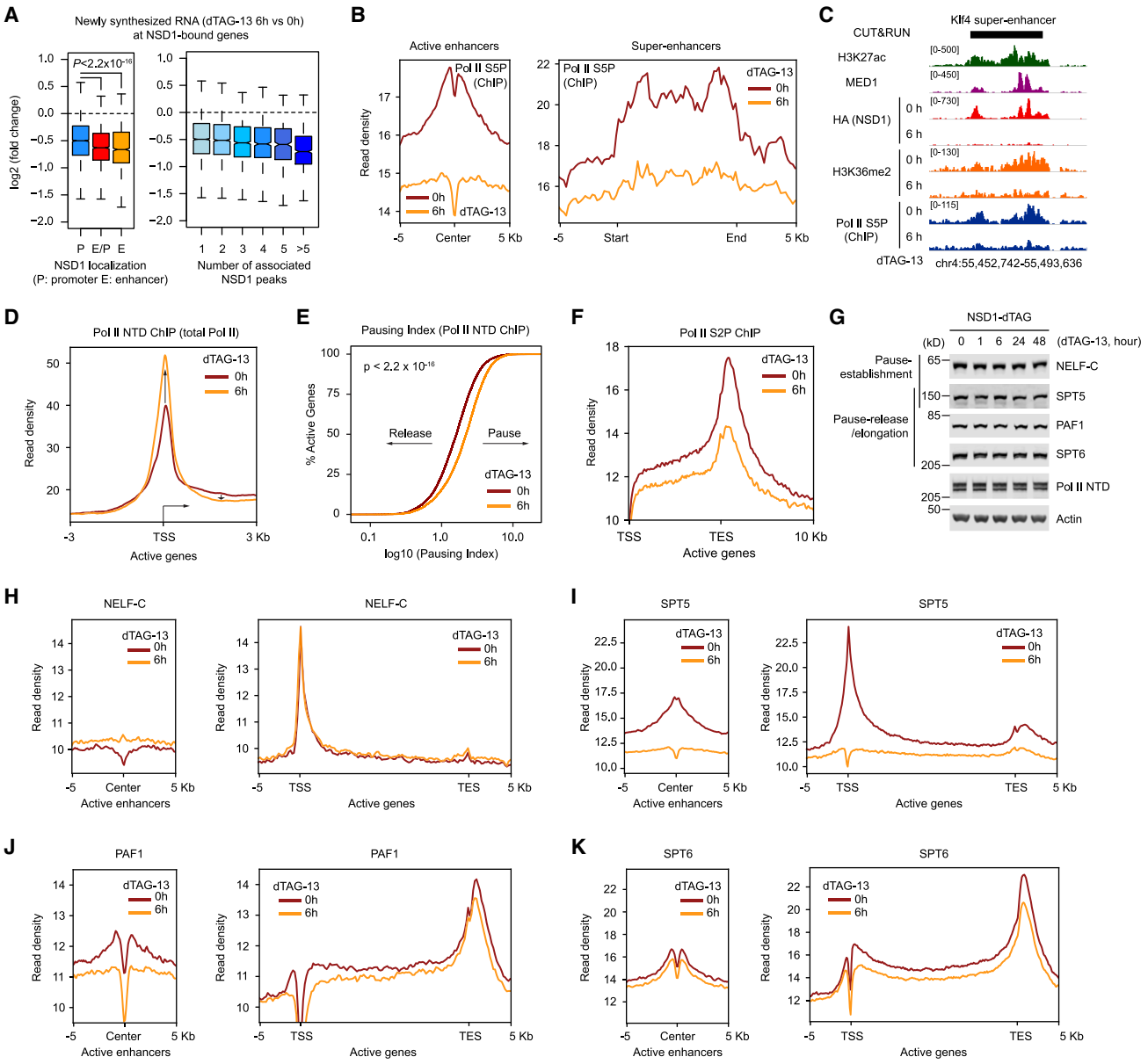


Figure 6. NSD1 promotes enhancer activity and RNA Pol II promoter pause release

(A) Magnitude of transcriptional downregulation (SLAM-seq) upon dTAG-13 treatment. p values calculated using Welch's t-test.

(B) Average profile of RNA Pol II S5P occupancy at active enhancers (n = 15,986) and super-enhancers (n = 705) in NSD1-dTAG mESCs.

(C) Genome browser view showing reduction of RNA Pol II S5P occupancy at the *Klf4* super-enhancer upon NSD1 degradation.

(D) Average profile of total RNA Pol II occupancy at TSS of active genes (RPKM > 1) in NSD1-dTAG mESCs. In (D), (F), and (H)–(K), TSS, transcription start site.

(E) Empirical cumulative distribution function (ECDF) plot of RNA Pol II pausing index following NSD1 loss. p values calculated using Welch's t-test.

(F) Average profile of RNA Pol II S2P occupancy across active gene body and after TES. In (F)–(K), TES, transcription end site.

(G) Western blots of whole-cell extracts from NSD1-dTAG mESCs.

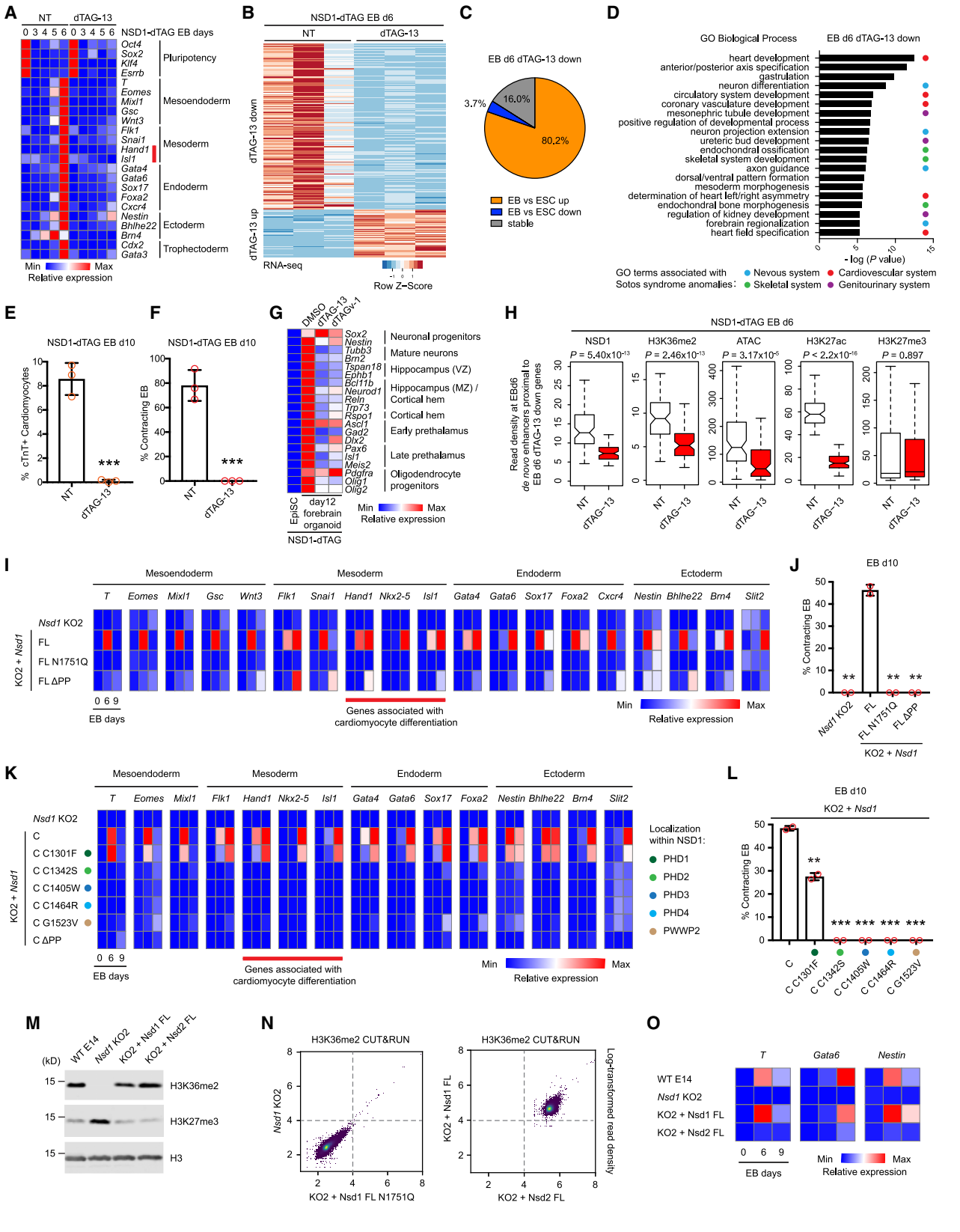
(H–K) Average profile showing occupancy of indicated factors (ChIP-seq) in NSD1-dTAG mESCs.

See also Figure S6.

which can promote gene transcription independent of its methyltransferase activity and regulate developmental transcriptional programs.

We show that NSD1 engages cell-type-specific *cis*-regulatory elements and displays particularly prevalent and prominent enrichment at active enhancers, where H3K36me2 also shows elevated occupancy. Consistently, NSD1 has been reported to associate with H3K27ac-decorated regions using chromatin proteomic approaches.[17,18] Our work has further identified a qPHD-PWWP module that promotes NSD1 chromatin

CellPress
OPEN ACCESS



(legend on next page)

association and enhancer enrichment, where all constituent domains are required. Importantly, we demonstrate that this reader module displays both selectivity toward p300/CBP-catalyzed H3K18ac and basal nucleosome affinity. These findings suggest that PHD1–4 and PWWP2 act synergistically to bind combinatorial chromatin features, conferring high avidity interaction and therefore robust enrichment at enhancers, whereas they allow transient nucleosome contacts in low-affinity chromatin contexts. Of note, the PHD fingers in this module lack essential aromatic residues for H3K4me3 recognition but resemble those binding H3K4me0.[65] Moreover, PHD3/4 shows conservation to the double PHD finger (DPF) domain present in the histone acetyltransferases MORF/MOZ and BAF subunits DPF2/3, which bind H3K14 acetylation and crotonylation.[65,66] Similarly, acetylation-engaging PHD domains are also present in MLL3/4, facilitating their enhancer association.[67] Additionally, the PWWP domain is able to bind DNA in a nonspecific manner and to interact with H3K36me2/3.[68] Therefore, the interaction of PHD1–4 and PWWP2 with the histone H3 tail and nucleosomal DNA, respectively, may underlie the basal nucleosome affinity, whereas recognition of the enhancer PTM H3K18ac contributes to the specificity of NSD1 localization. The requirement of this module in ESC differentiation further highlights its essential role in NSD1. Our structure–function study further shows that PWWP1, PHD5, and C5HCH domains are not required for NSD1 chromatin binding or ability to support differentiation in mESCs. However, since PHD5-C5HCH harbors multiple pathological missense mutations[37] and is required for the oncogenic NUP98-NSD1 fusion in acute myeloid leukemia,[24] it may have context-dependent functions.[69]

Unlike NSD1 and in addition to enhancer enrichment, we also observed broad intergenic occupancy of H3K36me2 (Figures S4H and S4K), which is in agreement with previous reports[29–32,70,71] and similar to other low methylation states, e.g., H3K4me1 and H3K27me1/2.[72] Recent studies have shown that chromatin-modifying enzymes are highly dynamic in the nucleus with only a small stably bound fraction.[73,74] The levels of H3K36me2 at distinct genomic regions may result from differential residence time of NSD1. Similarly, PRC2 deposits focal H3K27me3 at target sites of high residence time and dispersed H3K27me1/2 at broad genomic regions of dynamic contact.[72] Interestingly, in more differentiated cells, e.g., EpiSCs and EBs, or cancer cells, where NSD2 and/or NSD3 become abundant or dominant (Figures S2F and S7O), H3K36me2 seems to show a more pervasive distribution.[70,75–78] Therefore, it is of interest to compare the genomic distribution and nuclear dynamics of all NSD enzymes to understand their contribution to the H3K36me2 landscape.

Our study provides a comprehensive map of chromatin changes upon induced NSD1 loss. More than 50% of H3K36me2 was lost within 6 h of NSD1 degradation, which is much shorter than the cell division cycle of mESCs, suggesting that H3K36me2 turnover is mediated by active demethylation in addition to replication-coupled dilution. Notably, H3K36me2 loss results in a reciprocal increase of p300/CBP-mediated H3K36ac, which may underlie a compensatory mechanism in cells to buffer transcriptional perturbations after prolonged loss of NSD1. H3K36me3 distribution and levels remain largely unaltered upon NSD1 loss, consistent with previous reports,[29,30] demonstrating that NSD1-catalyzed H3K36me2 is not required for SETD2-mediated H3K36me3 catalysis. Together with genomic mapping of SETD2, our data indicate that NSD enzymes and SETD2 independently deposit H3K36me2 and H3K36me3 at distinct genomic regions. Notably, global accumulation of H3K27me3 and reduction of DNA methylation in response to H3K36me2 loss requires multiple cell cycles. These observations suggest an indirect nature of the H3K27me3-antagonism by H3K36me2 in cells or the presence of more dominant rate-limiting factors for H3K27me3 deposition at genomic sites without nearby PRC2 target sequences, i.e., transcriptionally silent CpG islands.[72] Importantly, the decoupling of transcriptional downregulation upon NSD1/H3K36me2 loss and H3K27me3 accumulation strongly suggests a H3K27me3-antagonism-independent mechanism for NSD1 in supporting

Figure 7. NSD1 facilitates activation of developmental transcriptional programs associated with Sotos syndrome pathogenesis

(A) Heatmap of RT-qPCR analysis of EB differentiation. In (A), (G), (I), (K), and (O), *Rpl7* used as internal reference; data represent mean from n = 3 technical replicates. Color represents relative expression normalized to day 0 in nontreated (NT) cells. Red bar indicates genes associated with cardiomyocyte differentiation.

(B) Heatmap of differential gene expression by RNA-seq ($|\log_2 FC| > 2$, $q < 0.05$). n = 3 independent differentiation.

(C) Percentage of dTAG-13 downregulated genes at EB day 6 (n = 164) that are up-regulated, downregulated, or stable during EB differentiation.

(D) Gene ontology (GO) analysis of dTAG-13 downregulated genes at EB day 6.

(E) Percentage of cardiomyocytes in EBs by flow cytometric analysis. In (E) and (F), data represent mean ± SD from n = 3 independent differentiation. In (E), (F), (H), (J), and (L), p values calculated using Student's t-test, $^{**}p < 0.01$, $^{***}p < 0.001$.

(F) Percentage of contracting EBs.

(G) Heatmap of RT-qPCR analysis of forebrain organoid differentiation using NSD1-dTAG EpiSCs. Color represents relative expression normalized to EpiSCs.

(H) Boxplots showing read densities of indicated chromatin features at *de novo* enhancers (EBd6 vs. ESC) associated with EB d6 dTAG-13 down genes (n = 81). p values calculated using Welch's t-test.

(I) Heatmap of RT-qPCR analysis of EB differentiation using indicated mESCs. In (I) and (K), color represents relative expression normalized to day 0 of *Nsd1* KO2 cells. PP, qPHD-PWWP.

(J) Percentage of contracting EBs. In (J) and (L), data represent mean ± SD from n = 2 independent differentiation.

(K) Heatmap of RT-qPCR analysis of EB differentiation using *Nsd1* KO2 mESCs expressing NSD1 C with Sotos syndrome missense mutations.

(L) Percentage of contracting EBs.

(M) Western blots of whole-cell extracts from mESCs. H3 used as loading control.

(N) Scatterplots showing genome-wide correlation of H3K36me2 occupancy.

(O) Heatmap of RT-qPCR analysis of EB differentiation using indicated mESCs. Color represents relative expression normalized to day 0 of WT E14 cells.

See also Figure S6.

enhancer activity and gene transcription. Additionally, the temporal decoupling and repressive nature of DNA methylation indicate that it is also unlikely to underlie the observed coactivator function of NSD1.

One of the most striking observations in our study was that acute NSD1 depletion leads to an imbalanced transcriptional consequence, with genes predominantly showing downregulation, and, surprisingly, that the methyltransferase activity of NSD1 is not required for facilitating transcription. Our data further indicate that NSD1 promotes enhancer activity to stimulate the release of paused RNA Pol II for productive elongation. These findings reveal a remarkable similarity between NSD1 and H3K4 mono-methyltransferases MLL3/4 in regulating enhancer activity, RNA Pol II pausing, and gene transcription, where H3K4me1 is also largely dispensable.[10,79] Similar to MLL3/4,[10,12,80] we show that NSD1 also functionally interacts with p300/CBP, in that NSD1 binds p300/CBP-catalyzed H3K18ac and enhances p300-mediated enhancer activation. Moreover, both NSD1 and MLL3/4 are dispensable for ESC self-renewal but required for differentiation.[80] Interestingly, H3K36me2 and H3K4me1 are also required for ESC differentiation, albeit H3K36me2 shows a broader role across lineages, likely due to its impact on global H3K27me3 and DNA methylation, whereas H3K4me1 only affects the specification of extra-embryonic fate.[81] Additionally, our results do not support a direct role for SETD2/H3K36me3 in transcriptional elongation. Taken together, our study revealed interesting similarities between the Set2 and Set1 families of histone H3K36 and H3K4 methyltransferases in gene regulation. NSD1 and MLL3/4, members that catalyze lower methylation levels, show distal enhancer enrichment and promote enhancer activity, whereas SETD2 and SET1A/B/MLL1/2, enzymes that catalyze trimethylation, are more closely associated with genes and deposit co-transcriptional histone modifications.

Our data show that NSD1 facilitates the engagement of RNA Pol II pause-release/elongation factors, particularly SPT5, to both enhancers and promoters. SPT5 is a conserved elongation factor, which heterodimerizes with SPT4 to form the DSIF complex.[57] SPT5 has pause-enhancing activity during early elongation but is converted into a positive elongation factor by P-TEFb-mediated phosphorylation and promotes RNA Pol II pause release and processive elongation.[57,58,60,61,64] SPT5 has also been reported to regulate enhancer transcription and activity.[55,82,83] Although a direct interaction has not been reported between SPT5 and NSD1, it will be interesting to further explore whether NSD1, a large protein with long intrinsically disordered regions, can facilitate SPT5 incorporation into transcriptional condensates formed at enhancers and cognate promoters, therefore facilitating its interaction with RNA Pol II. Additionally, SPT5, PAF1, and SPT6 interact with each other in the activated elongation complex,[58] and SPT5 has been shown to promote the recruitment of the PAF complex,[84] suggesting the moderate decrease of PAF1 and SPT6 occupancy upon NSD1 degradation may be a consequence of the reduction of SPT5. Both PAF1 and SPT6 have also been implicated in enhancer regulation.[85–87] Therefore, the reduced enhancer and gene transcription upon acute NSD1 depletion may result from the collectively reduced occupancy of RNA Pol II pause-release/elongation factors. Further studies will be required to better understand the mechanisms underlying the catalytic-independent transcriptional regulation by NSD1.

In summary, our work reveals chromatin-dependent genomic targeting and methyltransferase activity-independent coactivator function of NSD1 in enhancer and gene regulation. Additionally, our study both reveals a critical role of NSD1 in cell fate transition and demonstrates the potential to model Sotos syndrome and unravel its molecular pathogenesis using *in vitro* differentiation systems.

Limitations of the study

We identified H3K18ac as a substrate of the NSD1 qPHD-PWWP module by screening a set of recombinant nucleosomes with well-known histone PTMs using nuclear extracts, since we were not able to purify the qPHD-PWWP module due to solubility issues. However, an interaction study using a larger array of modified nucleosomes and a purified qPHD-PWWP module is required to further reveal the substrates of the qPHD-PWWP module and quantitatively dissect their combinatorial contribution to NSD1 enhancer recruitment.

We propose that during ESC differentiation, the catalytic-independent coactivator function of NSD1 contributes to the activation of NSD1-dependent enhancers, where H3K27me3 levels are not affected by the loss of NSD1/H3K36me2. However, since H3K36me2 is required for ESC differentiation, we cannot exclude that H3K36me2 to some degree facilitates enhancer activation in this context.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Plasmids and sequences
  - Generation of degron knockin cell lines
  - Generation of knockout cell lines
  - Transgenic expression in mESCs
  - dCas9-mediated gene activation in HEK293T cells
  - Proliferation assay
  - ESC to EpiLC/EpiSC conversion
  - Multilineage EB differentiation
  - Forebrain organoid differentiation
  - Immunofluoresence
  - Intracellular staining and flow cytometry
  - Whole-cell extract preparation and western blotting
  - Sequential salt extraction
  - Nucleosome pull-down with nuclear extract
  - Quantification of global DNA methylation by mass spectrometry

- RRBS
- ChIP
- CUT&RUN
- Purification of pAG-MNase
- Preparation of spike-in chromatin and DNA
- ATAC-seq
- RNA extraction, cDNA synthesis, and RT-qPCR
- RNA-seq
- SLAM-seq
- TT_{chem}-seq
- CUT&RUN, ChIP-seq, and ATAC-seq analysis
- RRBS analysis
- SLAM-seq analysis
- TT_{chem}-seq analysis
- RNA-seq analysis

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2023.06.007.

ACKNOWLEDGMENTS

The authors thank members of the Helin and Sawyers laboratories for advice and discussions; Zhifan Yang, Nikita Albanese, Sarah Teed, Chenyang Jiang, Sharon Nirmalakumar, and Serina Young for technical assistance; the Rockefeller University Proteomics Resource Center, particularly Dr. Hanan Alwaseem, for assistance with mass spectrometry analysis of DNA methylation; Dr. Elizabeth Wasmuth and members of the Wasmuth laboratory at UT Health San Antonio for pilot experiments in protein purification; Dr. Ruifang Li and Dr. P.J. Hamard from the CER Epigenetics Innovation Lab at MSKCC for discussions; Dr. Yuki Aoi from Northwestern University for technical advices; the High Performance Computing Center at MSKCC; and the Developmental Biology Imaging Core at MSKCC. This work was supported by an Edith C. Blum Foundation postdoctoral training grant (Z.S.), the MSKCC Functional Genomics Initiative Research Award (K.H.), MSKCC support grant P30 CA008748, and a center grant from the Novo Nordisk Foundation (NNF17CC0027852).

AUTHOR CONTRIBUTIONS

Z.S. and K.H. conceived this study, designed experiments, and interpreted results. Z.S. performed most experiments and bioinformatics analyses. Y.L. performed ATAC-seq with the help of Z.S. and collected microscopic images. M.T.I. performed forebrain organoid differentiation under the supervision of T.V. R.K. performed RRBS analysis. L.H. and D.L. assisted in generating a subset of NSD1 rescue cell lines under the supervision of Z.S. C.H. purified pAG-MNase. C.L.S. provided input into the project. Z.S. and K.H. acquired funding. Z.S. and K.H. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

C.L.S. serves on the Board of Directors of Novartis, is a co-founder of ORIC Pharmaceuticals, and is a co-inventor of enzalutamide and apalutamide. He is a science advisor to Arsenal, Beigene, Blueprint, Column Group, Foghorn, Housey Pharma, Nextech, KSQ, and PMV. K.H. is a co-founder of Dania Therapeutics and a scientific advisor for Hannibal Innovation. He was recently a scientific advisor for Inthera Bioscience AG and MetaboMed Inc.

Received: August 17, 2022
Revised: April 27, 2023
Accepted: June 5, 2023
Published: July 3, 2023

REFERENCES

1. Allis, C.D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. Nat. Rev. Genet. *17*, 487–500. https://doi.org/10.1038/nrg.2016.59.
2. Talbert, P.B., Meers, M.P., and Henikoff, S. (2019). Old cogs, new tricks: the evolution of gene expression in a chromatin context. Nat. Rev. Genet. *20*, 283–297. https://doi.org/10.1038/s41576-019-0105-7.
3. Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272–286. https://doi.org/10.1038/nrg3682.
4. Heinz, S., Romanoski, C.E., Benner, C., and Glass, C.K. (2015). The selection and function of cell type-specific enhancers. Nat. Rev. Mol. Cell Biol. *16*, 144–154. https://doi.org/10.1038/nrm3949.
5. Rickels, R., and Shilatifard, A. (2018). Enhancer logic and mechanics in development and disease. Trends Cell Biol. *28*, 608–630. https://doi.org/10.1016/j.tcb.2018.04.003.
6. Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell *167*, 1170–1187. https://doi.org/10.1016/j.cell.2016.09.018.
7. Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318. https://doi.org/10.1038/ng1966.
8. Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283. https://doi.org/10.1038/nature09692.
9. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936. https://doi.org/10.1073/pnas.1016071107.
10. Dorighi, K.M., Swigut, T., Henriques, T., Bhanu, N.V., Scruggs, B.S., Nady, N., Still, C.D., Garcia, B.A., Adelman, K., and Wysocka, J. (2017). Mll3 and Mll4 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4 monomethylation. Mol. Cell *66*, 568–576.e4. https://doi.org/10.1016/j.molcel.2017.04.018.
11. Narita, T., Ito, S., Higashijima, Y., Chu, W.K., Neumann, K., Walter, J., Satpathy, S., Liebner, T., Hamilton, W.B., Maskey, E., et al. (2021). Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release. Mol. Cell *81*, 2166–2182.e6. https://doi.org/10.1016/j.molcel.2021.03.008.
12. Wang, S.P., Tang, Z., Chen, C.W., Shimada, M., Koche, R.P., Wang, L.H., Nakadai, T., Chramiec, A., Krivtsov, A.V., Armstrong, S.A., et al. (2017). A UTX-MLL4-p300 transcriptional regulatory network coordinately shapes active enhancer landscapes for eliciting transcription. Mol. Cell *67*, 308–321.e6. https://doi.org/10.1016/j.molcel.2017.06.028.
13. Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320–334. https://doi.org/10.1016/j.cell.2013.03.036.
14. Hu, D., Gao, X., Morgan, M.A., Herz, H.-M., Smith, E.R., and Shilatifard, A. (2013). The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol. Cell. Biol. *33*, 4745–4754. https://doi.org/10.1128/MCB.01181-13.
15. Liu, W., Ma, Q., Wong, K., Li, W., Ohgi, K., Zhang, J., Aggarwal, A.K., and Rosenfeld, M.G. (2013). Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. Cell *155*, 1581–1595. https://doi.org/10.1016/j.cell.2013.10.056.
16. Myers, S.A., Wright, J., Peckner, R., Kalish, B.T., Zhang, F., and Carr, S.A. (2018). Discovery of proteins associated with a predefined genomic

locus via dCas9-APEX-mediated proximity labeling. Nat. Methods *15*, 437–439. https://doi.org/10.1038/s41592-018-0007-1.

17. Ji, X., Dadon, D.B., Abraham, B.J., Lee, T.I., Jaenisch, R., Bradner, J.E., and Young, R.A. (2015). Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. Proc. Natl. Acad. Sci. USA *112*, 3841–3846. https://doi.org/10.1073/pnas.1502971112.
18. Engelen, E., Brandsma, J.H., Moen, M.J., Signorile, L., Dekkers, D.H.W., Demmers, J., Kockx, C.E.M., Ozgür, Z., Van Ijcken, W.F.J., Van Den Berg, D.L.C., et al. (2015). Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. Nat. Commun. *6*, 7155. https://doi.org/10.1038/ncomms8155.
19. Wagner, E.J., and Carpenter, P.B. (2012). Understanding the language of Lys36 methylation at histone H3. Nat. Rev. Mol. Cell Biol. *13*, 115–126. https://doi.org/10.1038/nrm3274.
20. Musselman, C.A., Lalonde, M.E., Côté, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. Nat. Struct. Mol. Biol. *19*, 1218–1227. https://doi.org/10.1038/nsmb.2436.
21. Li, Y., Trojer, P., Xu, C.-F., Cheung, P., Kuo, A., Drury, W.J., Qiao, Q., Neubert, T.A., Xu, R.-M., Gozani, O., et al. (2009). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. J. Biol. Chem. *284*, 34283–34295. https://doi.org/10.1074/jbc.M109.034462.
22. Qiao, Q., Li, Y., Chen, Z., Wang, M., Reinberg, D., and Xu, R.M. (2011). The structure of NSD1 reveals an autoregulatory mechanism underlying histone H3K36 methylation. J. Biol. Chem. *286*, 8361–8368. https://doi.org/10.1074/jbc.M110.204115.
23. Pasillas, M.P., Shah, M., and Kamps, M.P. (2011). NSD1 PhD domains bind methylated H3K4 and H3K9 using interactions disrupted by point mutations in human Sotos syndrome. Hum. Mutat. *32*, 292–298. https://doi.org/10.1002/humu.21424.
24. Wang, G.G., Cai, L., Pasillas, M.P., and Kamps, M.P. (2007). NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. Nat. Cell Biol. *9*, 804–812. https://doi.org/10.1038/ncb1608.
25. Huang, N., Vom Baur, E., Garnier, J.M., Lerouge, T., Vonesch, J.L., Lutz, Y., Chambon, P., and Losson, R. (1998). Two distinct nuclear receptor interaction domains in NSD1, a novel SET protein that exhibits characteristics of both corepressors and coactivators. EMBO J. *17*, 3398–3412. https://doi.org/10.1093/emboj/17.12.3398.
26. Wang, X., Yeh, S., Wu, G., Hsu, C.L., Wang, L., Chiang, T., Yang, Y., Guo, Y., and Chang, C. (2001). Identification and characterization of a novel androgen receptor coregulator ARA267-α in prostate cancer cells. J. Biol. Chem. *276*, 40417–40423. https://doi.org/10.1074/jbc.M104765200.
27. Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell *42*, 330–341. https://doi.org/10.1016/j.molcel.2011.03.025.
28. Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J. Biol. Chem. *286*, 7983–7989. https://doi.org/10.1074/jbc.M110.194027.
29. Weinberg, D.N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K.N., Horth, C., McGuire, J.T., Xu, X., Nikbakht, H., et al. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature *573*, 281–286. https://doi.org/10.1038/s41586-019-1534-3.
30. Chen, H., Hu, B., Horth, C., Bareke, E., Rosenbaum, P., Kwon, S.Y., Sirois, J., Weinberg, D.N., Robison, F.M., Garcia, B.A., et al. (2022). H3K36 dimethylation shapes the epigenetic interaction landscape by directing repressive chromatin modifications in embryonic stem cells. Genome Res. *32*, 825–837. https://doi.org/10.1101/gr.276383.121.
31. Shirane, K., Miura, F., Ito, T., and Lorincz, M.C. (2020). NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. Nat. Genet. *52*, 1088–1098. https://doi.org/10.1038/s41588-020-0689-z.
32. Streubel, G., Watson, A., Jammula, S.G., Scelfo, A., Fitzpatrick, D.J., Oliviero, G., McCole, R., Conway, E., Glancy, E., Negri, G.L., et al. (2018). The H3K36me2 methyltransferase Nsd1 demarcates PRC2-mediated H3K27me2 and H3K27me3 domains in embryonic stem cells. Mol. Cell *70*, 371–379.e5. https://doi.org/10.1016/j.molcel.2018.02.027.
33. Drosos, Y., Myers, J.A., Xu, B., Mathias, K.M., Beane, E.C., Radko-Juettner, S., Mobley, R.J., Larsen, M.E., Piccioni, F., Ma, X., et al. (2022). NSD1 mediates antagonism between SWI/SNF and polycomb complexes and is required for transcriptional activation upon EZH2 inhibition. Mol. Cell *82*, 2472–2489.e8. https://doi.org/10.1016/j.molcel.2022.04.015.
34. Finogenova, K., Bonnet, J., Poepsel, S., Schäfer, I.B., Finkl, K., Schmid, K., Litz, C., Strauss, M., Benda, C., and Müller, J. (2020). Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3. eLife *9*, 1–30. https://doi.org/10.7554/eLife.61964.
35. Jani, K.S., Jain, S.U., Ge, E.J., Diehl, K.L., Lundgren, S.M., Müller, M.M., Lewis, P.W., and Muir, T.W. (2019). Histone H3 tail binds a unique sensing pocket in EZH2 to activate the PRC2 methyltransferase. Proc. Natl. Acad. Sci. USA *116*, 8295–8300. https://doi.org/10.1073/pnas.1819029116.
36. Rayasam, G.V., Wendling, O., Angrand, P.O., Mark, M., Niederreither, K., Song, L., Lerouge, T., Hager, G.L., Chambon, P., and Losson, R. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. EMBO J. *22*, 3153–3163. https://doi.org/10.1093/emboj/cdg288.
37. Tatton-Brown, K., Douglas, J., Coleman, K., Baujat, G., Cole, T.R.P., Das, S., Horn, D., Hughes, H.E., Temple, I.K., Faravelli, F., et al. (2005). Genotype-phenotype associations in Sotos syndrome: an analysis of 266 individuals with NSD1 aberrations. Am. J. Hum. Genet. *77*, 193–204. https://doi.org/10.1086/432082.
38. Saugier-Veber, P., Bonnet, C., Afenjar, A., Drouin-Garraud, V., Coubes, C., Fehrenbach, S., Holder-Espinasse, M., Roume, J., Malan, V., Portnoi, M.-F., et al. (2007). Heterogeneity of NSD1 alterations in 116 patients with Sotos syndrome. Hum. Mutat. *28*, 1098–1107. https://doi.org/10.1002/humu.20568.
39. Papillon-Cavanagh, S., Lu, C., Gayden, T., Mikael, L.G., Bechet, D., Karamboulas, C., Ailles, L., Karamchandani, J., Marchione, D.M., Garcia, B.A., et al. (2017). Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. Nat. Genet. *49*, 180–185. https://doi.org/10.1038/ng.3757.
40. Douglas, J., Hanks, S., Temple, I.K., Davies, S., Murray, A., Upadhyaya, M., Tomkins, S., Hughes, H.E., Cole, T.R.P., and Rahman, N. (2003). NSD1 mutations are the major cause of Sotos syndrome and occur in some cases of weaver syndrome but are rare in other overgrowth phenotypes. Am. J. Hum. Genet. *72*, 132–143. https://doi.org/10.1086/345647.
41. Türkmen, S., Gillessen-Kaesbach, G., Meinecke, P., Albrecht, B., Neumann, L.M., Hesse, V., Palanduz, S., Balg, S., Majewski, F., Fuchs, S., et al. (2003). Mutations in NSD1 are responsible for Sotos syndrome, but are not a frequent finding in other overgrowth phenotypes. Eur. J. Hum. Genet. *11*, 858–865. https://doi.org/10.1038/sj.ejhg.5201050.
42. Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for immediate and target-specific protein degradation. Nat. Chem. Biol. *14*, 431–441. https://doi.org/10.1038/s41589-018-0021-8.
43. Lucio-Eterovic, A.K., Singh, M.M., Gardner, J.E., Veerappan, C.S., Rice, J.C., and Carpenter, P.B. (2010). Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNApolymerase II function. Proc. Natl. Acad. Sci. USA *107*, 16952–16957. https://doi.org/10.1073/pnas.1002653107.
44. Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72–77. https://doi.org/10.1038/nature21373.

45. Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. eLife *6*, 1–35. https://doi.org/10.7554/eLife.21856.

46. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319. https://doi.org/10.1016/j.cell.2013.03.035.

47. Medina-Cano, D., Corrigan, E.K., Glenn, R.A., Islam, M.T., Lin, Y., Kim, J., Cho, H., and Vierbuchen, T. (2022). Rapid and robust directed differentiation of mouse epiblast stem cells into definitive endoderm and forebrain organoids. Development *149*, dev200561. https://doi.org/10.1242/dev.200561.

48. Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853. https://doi.org/10.1016/j.stem.2014.04.003.

49. Weinert, B.T., Narita, T., Satpathy, S., Srinivasan, B., Hansen, B.K., Schölz, C., Hamilton, W.B., Zucconi, B.E., Wang, W.W., Liu, W.R., et al. (2018). Time-resolved analysis reveals rapid dynamics and broad scope of the CBP/p300 acetylome. Cell *174*, 231–244.e12. https://doi.org/10.1016/j.cell.2018.04.033.

50. Jin, Q., Yu, L.R., Wang, L., Zhang, Z., Kasper, L.H., Lee, J.E., Wang, C., Brindle, P.K., Dent, S.Y.R., and Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. EMBO J. *30*, 249–262. https://doi.org/10.1038/emboj.2010.318.

51. Vieira, G.H., Cook, M.M., Ferreira De Lima, R.L., Frigério Domingues, C.E., De Carvalho, D.R., Soares De Paiva, I., Moretti-Ferreira, D., and Srivastava, A.K. (2015). Clinical and molecular heterogeneity in Brazilian patients with Sotos syndrome. Mol. Syndromol. *6*, 32–38. https://doi.org/10.1159/000370169.

52. Choufani, S., Cytrynbaum, C., Chung, B.H.Y., Turinsky, A.L., Grafodatskaya, D., Chen, Y.A., Cohen, A.S.A., Dupuis, L., Butcher, D.T., Siu, M.T., et al. (2015). NSD1 mutations generate a genome-wide DNA methylation signature. Nat. Commun. *6*, 10207. https://doi.org/10.1038/ncomms10207.

53. Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. Nat. Methods *14*, 1198–1204. https://doi.org/10.1038/nmeth.4435.

54. El Khattabi, L., Zhao, H., Kalchschmidt, J., Young, N., Jung, S., Van Blerkom, P., Kieffer-Kwon, P., Kieffer-Kwon, K.R., Park, S., Wang, X., et al. (2019). A pliable mediator acts as a functional rather than an architectural bridge between promoters and enhancers. Cell *178*, 1145–1158.e20. https://doi.org/10.1016/j.cell.2019.07.011.

55. Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev. *32*, 26–41. https://doi.org/10.1101/gad.309351.117.

56. Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat. Rev. Mol. Cell Biol. *18*, 263–273. https://doi.org/10.1038/nrm.2017.10.

57. Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: A nexus of gene regulation. Genes Dev. *33*, 960–982. https://doi.org/10.1101/gad.325142.119.

58. Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018). Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. Nature *560*, 607–612. https://doi.org/10.1038/s41586-018-0440-4.

59. Yu, M., Yang, W., Ni, T., Tang, Z., Nakadai, T., Zhu, J., and Roeder, R.G. (2015). RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. Science *350*, 1383–1386. https://doi.org/10.1126/science.aad2338.

60. Shetty, A., Kallgren, S.P., Demel, C., Maier, K.C., Spatt, D., Alver, B.H., Cramer, P., Park, P.J., and Winston, F. (2017). Spt5 plays vital roles in the control of sense and antisense transcription elongation. Mol. Cell *66*, 77–88.e5. https://doi.org/10.1016/j.molcel.2017.02.023.

61. Baluapuri, A., Hofstetter, J., Dudvarski Stankovic, N., Endres, T., Bhandare, P., Vos, S.M., Adhikari, B., Schwarz, J.D., Narain, A., Vogt, M., et al. (2019). MYC recruits SPT5 to RNA polymerase II to promote processive transcription elongation. Mol. Cell *74*, 674–687.e11. https://doi.org/10.1016/j.molcel.2019.02.031.

62. Ding, L., Paszkowski-Rogacz, M., Mircetic, J., Chakraborty, D., and Buchholz, F. (2021). The Paf1 complex positively regulates enhancer activity in mouse embryonic stem cells. Life Sci. Alliance *4*, 1–14. https://doi.org/10.26508/LSA.202000792.

63. Žumer, K., Maier, K.C., Farnung, L., Jaeger, M.G., Rus, P., Winter, G., and Cramer, P. (2021). Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo. Mol. Cell *81*, 3096–3109.e8. https://doi.org/10.1016/j.molcel.2021.05.028.

64. Fong, N., Sheridan, R.M., Ramachandran, S., and Bentley, D.L. (2022). The pausing zone and control of RNA polymerase II elongation by Spt5: implications for the pause-release model. Mol. Cell *82*, 3632–3645.e4. https://doi.org/10.1016/j.molcel.2022.09.001.

65. Sanchez, R., and Zhou, M.M. (2011). The PhD finger: A versatile epigenome reader. Trends Biochem. Sci. *36*, 364–372. https://doi.org/10.1016/j.tibs.2011.03.005.

66. Xiong, X., Panchenko, T., Yang, S., Zhao, S., Yan, P., Zhang, W., Xie, W., Li, Y., Zhao, Y., Allis, C.D., et al. (2016). Selective recognition of histone crotonylation by double PhD fingers of MOZ and DPF2. Nat. Chem. Biol. *12*, 1111–1118. https://doi.org/10.1038/nchembio.2218.

67. Zhang, Y., Jang, Y., Lee, J.E., Ahn, J.W., Xu, L., Holden, M.R., Cornett, E.M., Krajewski, K., Klein, B.J., Wang, S.P., et al. (2019). Selective binding of the PHD6 finger of MLL4 to histone H4K16ac links MLL4 and MOF. Nat. Commun. *10*, 2314. https://doi.org/10.1038/s41467-019-10324-8.

68. Qin, S., and Min, J. (2014). Structure and function of the nucleosome-binding PWWP domain. Trends Biochem. Sci. *39*, 536–547. https://doi.org/10.1016/j.tibs.2014.09.001.

69. Berardi, A., Quilici, G., Spiliotopoulos, D., Corral-Rodriguez, M.A., Martin-Garcia, F., Degano, M., Tonon, G., Ghitti, M., and Musco, G. (2016). Structural basis for PHDVC5HCHNSD1-C2HRNizp1 interaction: implications for Sotos syndrome. Nucleic Acids Res. *44*, 3448–3463. https://doi.org/10.1093/nar/gkw103.

70. Lu, C., Jain, S.U., Hoelper, D., Bechet, D., Molden, R.C., Ran, L., Murphy, D., Venneti, S., Hameed, M., Pawel, B.R., et al. (2016). Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. Science *352*, 844–849. https://doi.org/10.1126/science.aac7272.

71. Fang, D., Gan, H., Lee, J.H., Han, J., Wang, Z., Riester, S.M., Jin, L., Chen, J., Zhou, H., Wang, J., et al. (2016). The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. Science *352*, 1344–1348. https://doi.org/10.1126/science.aae0065.

72. Laugesen, A., Højfeldt, J.W., and Helin, K. (2019). Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. Mol. Cell *74*, 8–18. https://doi.org/10.1016/j.molcel.2019.03.011.

73. Huseyin, M.K., and Klose, R.J. (2021). Live-cell single particle tracking of PRC1 reveals a highly dynamic system with low target site occupancy. Nat. Commun. *12*, 887. https://doi.org/10.1038/s41467-021-21130-6.

74. Youmans, D.T., Schmidt, J.C., and Cech, T.R. (2018). Live-cell imaging reveals the dynamics of PRC2 and recruitment to chromatin by SUZ12-associated subunits. Genes Dev. *32*, 794–805. https://doi.org/10.1101/gad.311936.118.

75. Popovic, R., Martinez-Garcia, E., Giannopoulou, E.G., Zhang, Q., Zhang, Q., Ezponda, T., Shah, M.Y., Zheng, Y., Will, C.M., Small, E.C., et al. (2014). Histone methyltransferase MMSET/NSD2 alters EZH2 binding

and reprograms the myeloma epigenome through global and focal changes in H3K36 and H3K27 methylation. PLoS Genet. *10*, e1004566. https://doi.org/10.1371/journal.pgen.1004566.

76. Kuo, A.J., Cheung, P., Chen, K., Zee, B.M., Kioi, M., Lauring, J., Xi, Y., Park, B.H., Shi, X., Garcia, B.A., et al. (2011). NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. Mol. Cell *44*, 609–620. https://doi.org/10.1016/j.molcel.2011.08.042.

77. SenGupta, D., Zeng, L., Li, Y., Hausmann, S., Ghosh, D., Yuan, G., Nguyen, T.N., Lyu, R., Caporicci, M., Morales Benitez, A., et al. (2021). NSD2 dimethylation at H3K36 promotes lung adenocarcinoma pathogenesis. Mol. Cell *81*, 4481–4492.e9. https://doi.org/10.1016/j.molcel.2021.08.034.

78. Yuan, G., Flores, N.M., Hausmann, S., Lofgren, S.M., Kharchenko, V., Angulo-Ibanez, M., SenGupta, D., Lu, X., Czaban, I., Azhibek, D., et al. (2021). Elevated NSD3 histone methylation activity drives squamous cell lung cancer. Nature *590*, 504–508. https://doi.org/10.1038/s41586-020-03170-y.

79. Rickels, R., Herz, H.M., Sze, C.C., Cao, K., Morgan, M.A., Collings, C.K., Gause, M., Takahashi, Y.H., Wang, L., Rendleman, E.J., et al. (2017). Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. Nat. Genet. *49*, 1647–1653. https://doi.org/10.1038/ng.3965.

80. Wang, C., Lee, J.E., Lai, B., Macfarlan, T.S., Xu, S., Zhuang, L., Liu, C., Peng, W., and Ge, K. (2016). Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. Proc. Natl. Acad. Sci. USA *113*, 11871–11876. https://doi.org/10.1073/pnas.1606857113.

81. Xie, G., Lee, J.-E., Senft, A.D., Park, Y.-K., Jang, Y., Chakraborty, S., Thompson, J.J., McKernan, K., Liu, C., Macfarlan, T.S., et al. (2023). MLL3/MLL4 methyltransferase activities control early embryonic development and embryonic stem cell differentiation in a lineage-selective manner. Nat. Genet. *55*, 693–705. https://doi.org/10.1038/s41588-023-01356-4.

82. Hu, S., Peng, L., Xu, C., Wang, Z., Song, A., and Chen, F.X. (2021). SPT5 stabilizes RNA polymerase II, orchestrates transcription cycles, and maintains the enhancer landscape. Mol. Cell *81*, 4425–4439.e6. https://doi.org/10.1016/j.molcel.2021.08.029.

83. Fitz, J., Neumann, T., Steininger, M., Wiedemann, E.M., Garcia, A.C., Athanasiadis, A., Schoeberl, U.E., and Pavri, R. (2020). Spt5-mediated enhancer transcription directly couples enhancer activation with physical promoter interaction. Nat. Genet. *52*, 505–515. https://doi.org/10.1038/s41588-020-0605-6.

84. Wier, A.D., Mayekar, M.K., Héroux, A., Arndt, K.M., and VanDemark, A.P. (2013). Structural basis for Spt5-mediated recruitment of the Paf1 complex to chromatin. Proc. Natl. Acad. Sci. USA *110*, 17290–17295. https://doi.org/10.1073/pnas.1314754110.

85. Chen, F.X., Xie, P., Collings, C.K., Cao, K., Aoi, Y., Marshall, S.A., Rendleman, E.J., Ugarenko, M., Ozark, P.A., Zhang, A., et al. (2017). PAF1 regulation of promoter-proximal pause release via enhancer activation. Science *357*, 1294–1298. https://doi.org/10.1126/science.aan3269.

86. Wang, A.H., Juan, A.H., Ko, K.D., Tsai, P.F., Zare, H., Dell'orso, S., and Sartorelli, V. (2017). The elongation factor Spt6 maintains ESC pluripotency by controlling super-enhancers and counteracting polycomb proteins. Mol. Cell *68*, 398–413.e6. https://doi.org/10.1016/j.molcel.2017.09.016.

87. Crump, N.T., Smith, A., Godfrey, L., Jackson, N., Rice, S., Kim, J., Basrur, V., Fermin, D., Elenitoba-Johnson, K., Roeder, R.G., et al. (2022). PAF1 and FACT cooperate with MLL-AF4 to drive enhancer activity in leukemia. https://doi.org/10.1101/2022.05.27.493676.

88. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

89. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, 1–4. https://doi.org/10.1093/gigascience/giab008.

90. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

91. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

92. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

93. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/NAR/GKW257.

94. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

95. Neumann, T., Herzog, V.A., Muhar, M., Von Haeseler, A., Zuber, J., Ameres, S.L., and Rescheneder, P. (2019). Quantification of experimentally induced nucleotide conversions in high-throughput sequencing datasets. BMC Bioinformatics *20*, 258. https://doi.org/10.1186/s12859-019-2849-7.

96. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 1–21. https://doi.org/10.1186/s13059-014-0550-8.

97. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

98. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128. https://doi.org/10.1186/1471-2105-14-128.

99. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515. https://doi.org/10.1038/nbt.1621.

100. Hulsen, T., de Vlieg, J., and Alkema, W. (2008). BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics *9*, 488. https://doi.org/10.1186/1471-2164-9-488.

101. Brinkman, E.K., Chen, T., Amendola, M., and Van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. *42*, e168. https://doi.org/10.1093/nar/gku936.

102. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

103. Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG Islands genome wide. Mol. Cell *55*, 347–360. https://doi.org/10.1016/j.molcel.2014.06.005.

104. Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88. https://doi.org/10.1126/science.aad5227.

105. Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters

and enhancers. Nat. Biotechnol. *33*, 510–517. https://doi.org/10.1038/nbt.3199.

106. Porter, E.G., Connelly, K.E., and Dykhuizen, E.C. (2017). Sequential salt extractions for the analysis of bulk chromatin binding properties of chromatin modifying complexes. J. Vis. Exp. 1–7. https://doi.org/10.3791/55369.

107. Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *11*, 1475–1489. https://doi.org/10.1093/nar/11.5.1475.

108. Kenny, T.C., Khan, A., Son, Y., Yue, L., Heissel, S., Sharma, A., Pasolli, H.A., Liu, Y., Gamazon, E.R., Alwaseem, H., et al. (2023). Integrative genetic analysis identifies FLVCR1 as a plasma-membrane choline transporter in mammals. Cell Metab. *35*, 1057–1071.e12. https://doi.org/10.1016/j.cmet.2023.04.003.

109. Adams, K.J., Pratt, B., Bose, N., Dubois, L.G., St John-Williams, L., Perrott, K.M., Ky, K., Kapahi, P., Sharma, V., MacCoss, M.J., et al. (2020). Skyline for small molecules: A unifying software package for quantitative metabolomics. J. Proteome Res. *19*, 1447–1458. https://doi.org/10.1021/acs.jproteome.9b00640.

110. Meers, M.P., Bryson, T.D., Henikoff, J.G., and Henikoff, S. (2019). Improved cut&run chromatin profiling tools. eLife *8*, 1–16. https://doi.org/10.7554/eLife.46314.

111. Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962. https://doi.org/10.1038/nmeth.4396.

112. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218. https://doi.org/10.1038/nmeth.2688.

113. Gregersen, L.H., Mitter, R., and Svejstrup, J.Q. (2020). Using TTchem-seq for profiling nascent transcription and measuring transcript elongation. Nat. Protoc. *15*, 604–627. https://doi.org/10.1038/s41596-019-0262-3.

114. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE blacklist: identification of problematic regions of the genome. Sci. Rep. *9*, 9354. https://doi.org/10.1038/s41598-019-45839-z.

115. Javierre, B.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., Freire-Pritchett, P., Spivakov, M., Fraser, P., Burren, O.S., et al. (2016). Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell *167*, 1369–1384.e19. https://doi.org/10.1016/j.cell.2016.09.037.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
HA	Cell Signaling	Cat# 3724; RRID:AB_1549585
FLAG	Sigma	Cat# F1804; RRID:AB_262044
NSD1	Abbexa	Cat# 135901; RRID:AB_2936813
SETD2	Abcam	Cat# ab239350; RRID:AB_2936814
MED1	Abcam	Cat# ab64965; RRID:AB_1142301
p300	Cell Signaling	Cat# 86377; RRID:AB_2800077
BRD4	Abcam	Cat# ab128874; RRID:AB_11145462
BRG1	Santa Cruz	Cat# sc-17796; RRID:AB_626762
ARID1A	Santa Cruz	Cat# sc-32761; RRID:AB_673396
OCT4	Santa Cruz	Cat# sc-5279; RRID:AB_628051
RNA Pol II NTD	Cell Signaling	Cat# 14958; RRID:AB_2687876
RNA Pol II S2P	Cell Signaling	Cat# 13499; RRID:AB_2798238
RNA Pol II S5P	Cell Signaling	Cat# 13523; RRID:AB_2798246
NELF-C	Cell Signaling	Cat# 12265; RRID:AB_2797862
SPT5	BD Biosciences	Cat# 611107; RRID:AB_398420
PAF1	Cell Signaling	Cat# 12883; RRID:AB_2798052
SPT6	Cell Signaling	Cat# 15616; RRID:AB_2798748
SUZ12	Cell Signaling	Cat# 3737; RRID:AB_2196850
MEK1/2	Cell Signaling	Cat# 9122; RRID:AB_823567
SSRP1	Santa Cruz	Cat# sc-56782; RRID:AB_630253
Tubulin (Alpha)	Abcam	Cat# ab176560; RRID:AB_2860019
Actin (Beta)	Abcam	Cat# ab6276; RRID:AB_2223210
Vinculin	Sigma	Cat# SAB4200080; RRID:AB_10604160
Mouse IgG	Abcam	Cat# ab46540; RRID:AB_2614925
OCT6	Millipore	Cat# MABN738; RRID:AB_2876862
KLF4	R&D Systems	Cat# AF3158; RRID:AB_2130245
cTnT (Cardiac Troponin T)	Thermo Fisher	Cat# MS-295-P; RRID:AB_61806
H3	Abcam	Cat# ab1791; RRID:AB_302613
H3K36me2	Thermo Fisher	Cat# MA5-14867; RRID:AB_10983670
H3K36ac	Thermo Fisher	Cat# MA5-24672; RRID:AB_2661914
H3K36me3	Active Motif	Cat# 61021; RRID:AB_2614986
H3K4me1	Cell Signaling	Cat# 5326; RRID:AB_10695148
H3K4me3	Thermo Fisher	Cat# PA5-27029; RRID:AB_2544505
H3K27ac	Thermo Fisher	Cat# MA5-23516; RRID:AB_2608307
H3K18ac	Thermo Fisher	Cat# 703896; RRID:AB_2895888
H3K18acyl	EpiCypher	Cat# 13-0050; RRID:AB_2936815
H3K27me3	Cell Signaling	Cat# 9733; RRID:AB_2616029
H3K9me3	Abcam	Cat# ab176916; RRID:AB_2797591
H2Av (Spike-In Antibody)	Active Motif	Cat# 61686; RRID:AB_2737370
IRDye 680RD Goat Anti-Rabbit IgG	Li-Core	Cat# 925–68071; RRID:AB_2721181
IRDye 800CW Goat Anti-Rabbit IgG	Li-Core	Cat# 925–32211; RRID:AB_2651127
IRDye 680RD Goat Anti-Mouse IgG	Li-Core	Cat# 925–68070; RRID:AB_2651128
IRDye 800CW Goat Anti-Mouse IgG	Li-Core	Cat# 925–32210; RRID:AB_2687825

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Alexa Flour 488 Goat Anti-Rabbit IgG	Thermo Fisher	Cat# A-11034; RRID:AB_2576217
Alexa Flour 488 Goat Anti-Mouse IgG	Thermo Fisher	Cat# A-11029; RRID:AB_2534088
Alexa Fluor 594 Donkey Anti-goat IgG	Thermo Fisher	Cat# A-11058; RRID:AB_142540
Alexa Fluor 647 Goat Anti-Mouse IgG	Thermo Fisher	Cat# A-21236; RRID:AB_2535805
Bacterial and virus strains		
Stbl3 Chemically Competent Cells	Kristian Helin lab	N/A
Stellar Chemically Competent Cells	Takara	Cat# 636763
JM101 Chemically Competent Cells	Agilent Technologies	Cat# 200234
Chemicals, peptides, and recombinant proteins		
LIF	Kristian Helin lab	N/A
PD0325901 (MEKi)	Kristian Helin lab	N/A
CHIR99021 (GSK3i)	Kristian Helin lab	N/A
FGF2 (bFGF)	R&D Systems	Cat# 233-FB
Activin A	Peprotech	Cat# 120-14P
NVP-TNKS656 (TNKS2i)	Selleck Chemicals	Cat# S7238
dTAG-13	Tocris	Cat# 6605
dTAGv-1	Tocris	Cat# 6914
A-485 (p300/CBPi)	MedChem Express	Cat# HY-107455
4-Thiouridine (4sU)	Biosynth	Cat# NT06186
Iodoacetamide (IAA)	Sigma	Cat# I1149
MTSEA Biotin-XX	Biotium	Cat# BT90066
Digitonin	Millipore	Cat# 300410
pAG-MNase	This paper	N/A
Unmodified Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0006
H3K9ac Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0314
H3K14ac Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0343
H3K18ac Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0372
H3K27ac Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0365
H4K16ac Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0354
H3K4me1 Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0321
H3K4me2 Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0334
H3K4me3 Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0316
H3K36me1 Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0322
H3K36me2 Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0319
H3K36me3 Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0320
H3K27me3 Recombinant Nucleosome, Biotinylated	EpiCypher	Cat# 16-0317
Dynabeads MyOne Streptavidin T1	Thermo Fisher	Cat# 65601

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Dynabeads Protein G	Thermo Fisher	Cat# 10004D
BioMag Plus Concanavalin A Magnetic Beads	Bangs Laboratories	Cat# BP531
μMACS Streptavidin Microbeads	Miltenyi	Cat# 130-074-101
Agencourt AMPure XP Magnetic Beads	Beckman Coulter	Cat# A63881
Critical commercial assays		
In-Fusion HD Cloning Kit	Takara	Cat# 639650
Lipofectamine 3000 Transfection Reagent	Thermo Fisher	Cat# L3000001
NEBNext Ultra II DNA Library Prep Kit for Illumina	NEB	Cat# E7645L
Illumina Tagment DNA Enzyme and Buffer Small Kit	Illumina	Cat# 20034197
NEBNext High-Fidelity 2x PCR Master Mix	NEB	Cat# M0541S
NEXTFLEX Bisulfite Library Prep Kit	PerkinElmer	Cat# NOVA-5119-01
EZ DNA Methylation-Lightning Kit	Zymo Research	Cat# D5030T
KAPA HiFi HotStart Uracil+ ReadyMix Kit	Roche	Cat# KK2801
NEBNext Poly(A) mRNA Magnetic Isolation Module	NEB	Cat# E7490S
NEBNext Ultra II Directional RNA Library Prep Kit for Illumina	NEB	Cat# E7760S
QuantSeq 3′ mRNA-Seq Library Prep FWD Kit for Illumina	Lexogen	Cat# 015
Deposited Data		
Raw and Analyzed Sequencing Data	This paper	GEO: GSE208695
DECAP-Seq	Neri et al.[44]	GEO: GSE72856
Uncropped Western Blots and Immunofluorescence Images	Mendeley data	Mendeley data: https://doi.org/10.17632/ncy74b8tm7.1
Experimental models: Cell lines		
E14TG2a (E14)	Kristian Helin lab	RRID:CVCL_C320
HEK293T	Charles Sawyers lab	RRID:CVCL_0063
Schneider 2 (S2)	Kristian Helin lab	RRID:CVCL_Z232
Oligonucleotides		
sgRNAs	This paper	Table S2
Genotyping Primers	This paper	Table S4
ChIP-qPCR Primers	This paper	Table S6
RT-qPCR Primers	This paper	Table S7
Recombinant DNA		
pUC19-2xGGSG-dTAG-2xHA-P2A-Neo-Nsd1-CtKI-targeting	This paper	N/A
pUC19-2xGGSG-dTAG-2xHA-P2A-Puro-Nsd1-CtKI-targeting	This paper	N/A
pUC19-2xGGSG-dTAG-2xHA-P2A-Neo-Setd2-CtKI-targeting	This paper	N/A
pUC19-2xGGSG-dTAG-2xHA-P2A-Puro-Setd2-CtKI-targeting	This paper	N/A
U6-sgRNA-eSpCas9(1.1)-T2A-mCherry	Ian Chambers lab	N/A
U6-sgRNA-eSpCas9(1.1)-T2A-GFP	Ian Chambers lab	N/A
PBase	Thomas Vierbuchen lab	N/A
PB-CAG-MCS-HA-IRES-Neo	Thomas Vierbuchen lab	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
PB-CAG-MCS-3xFLAG-2xHA-IRES-Zeo	This paper	N/A
PB-CAG-MCS-3xFLAG-2xHA-IRES-Puro	This paper	N/A
PB-CAG-Nsd1-series-HA-IRES-Neo	This paper	N/A
PB-CAG-Nsd2-HA-IRES-Neo	This paper	N/A
PB-CAG-Nsd1-series-3xFLAG-2xHA-IRES-Zeo	This paper	N/A
PB-CAG-Nsd2-3xFLAG-2xHA-IRES-Puro	This paper	N/A
cDNA: Nsd1.1S	This paper	Table S3
cDNA: Nsd2	This paper	Table S3
pSPgRNA	Addgene	Cat# 47108
pcDNA-dCas9	Addgene	Cat# 61355
pcDNA-dCas9-VP64	Addgene	Cat# 47107
pcDNA-dCas9-p300^{core}	Addgene	Cat# 61357
pcDNA-dCas9-p300^{core} (D1399Y)	Addgene	Cat# 61358
pcDNA-dCas9-NSD1^{C}	This paper	N/A
pcDNA-dCas9-NSD1^{C} (N1751Q)	This paper	N/A
pAG-ERH-MNase-6xHis-HA	Steve Henikoff lab	N/A
Software and algorithms		
bcl2fastq2 (v2.20.0.422)	Illumina	N/A
FastQC (v0.11.8)	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	N/A
Bowtie2 (v2.3.4.1)	Langmead and Salzberg[88]	N/A
SAMtools (v1.9)	Danecek et al.[89]	N/A
MACS2 (v2.2.6)	Zhang et al.[90]	N/A
Bedtools (v2.27.1)	Quinlan and Hall[91]	N/A
Homer (v4.11)	Heinz et al.[92]	N/A
DeepTools (v3.2.1)	Ramírez et al.[93]	N/A
IGV (v2.5.1)	Robinson et al.[94]	N/A
R (v4.0.3)	www.r-project.org	N/A
ROSE (v1.0.0)	Whyte et al.[46]	N/A
Trim Galore (v0.6.4)	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/	N/A
Bismark (v0.19.0)	https://www.bioinformatics.babraham.ac.uk/projects/bismark/	N/A
UCSC Table Browser	https://genome.ucsc.edu/cgi-bin/hgTables	N/A
SlamDunk (v0.4.3)	Neumann et al.[95]	N/A
DESeq2 (v1.30.0)	Love et al.[96]	N/A
STAR (v2.6.0a)	Dobin et al.[97]	N/A
Enrichr	Chen et al.[98]	N/A
Cufflinks (v2.2.1)	Trapnell et al.[99]	N/A
BioVenn	Hulsen et al.[100]	N/A
TIDE	Brinkman et al.[101]	N/A
FlowJo	BD Bioscience	N/A
ImageJ	Schneider et al.[102]	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kristian Helin (kristian.helin@icr.ac.uk).

Materials availability

Plasmids and cell lines generated in this study will be available from the lead contact upon request.

Data and code availability

Next generation sequencing data have been deposited at GEO: GSE208695 and are publicly available as of the date of publication. Original western blot and immunofluorescence images have been deposited at Mendeley data: https://doi.org/10.17632/ncy74b8tm7.1 and are publicly available as of the date of publication. Accession numbers and DOI are also listed in the key resources table.

This study does not report original code. Analyses were performed with published tools as described in STAR Methods.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

E14TG2a (E14) mouse ESCs (male, RRID:CVCL_C320)[103] and derived stable cell lines were cultured on 0.1% gelatin-coated dishes in 2i/LIF medium containing 50% DMEM/F12 (Gibco), 50% Neurobasal (Gibco), 0.5% N2 supplement (Gibco), 1% B27 supplement (Gibco), 1% penicillin/streptomycin (MSKCC), 2 mM L-glutamine (MSKCC), 0.1 mM Non-essential amino acids (Gibco), 1 mM Sodium Pyruvate (Gibco), 50 μM b-mercaptoethanol (Gibco), 1 μM MEKi (PD0325901), 3 μM GSK3i (CHIR99021), and 1x leukemia inhibitory factor (LIF, produced in the Helin lab). For initial establishment of knockin, knockout, and transgenic expression cell lines, mESCs were cultured in serum/LIF medium containing GMEM (Sigma), 10% heat-inactivated FBS (HyClone), 1% penicillin/streptomycin (MSKCC), 2 mM L-glutamine (MSKCC), 0.1 mM Non-essential amino acids (Gibco), 1 mM Sodium Pyruvate (Gibco), 100 μM β-Mercaptoethanol (Gibco), and 1x LIF (produced in the Helin lab), and then adapted to the 2i/LIF culture condition for at least 3 passages. ESCs in 2i/LIF and serum/LIF conditions were passaged every two days using Accutase (Sigma) and 0.25% trypsin (MSKCC), respectively. HEK293T cells (female, RRID:CVCL_0063) were cultured in DMEM (high glucose, MSKCC), 10% heat-inactivated FBS (HyClone) and 1% penicillin/streptomycin (MSKCC). E14 and HEK293T cells were cultured at 37°C. *Drosophila* Schneider 2 (S2) cells (male, RRID:CVCL_Z232) were cultured in Schneider's Insect Medium (MSKCC), 10% heat inactivated FBS (HyClone), and 25 μg/mL primocin (InvivoGen) at room temperature without extra CO_2. Where indicated, cells were treated with 500 nM dTAG-13 (Tocris), 500 nM dTAGv-1 (Tocris), or 6 μM A-485 (MedChem Express). Where indicated antibiotic selection was performed with 500 μg/mL neomycin, 1 μg/mL puromycin, or 20 μg/mL zeocin. Cell lines were regularly tested for mycoplasma, but were not authenticated.

METHOD DETAILS

Plasmids and sequences

For construction of the $FKBP12^{F36V}$ degradation tag (dTAG) C-terminal knockin donor plasmids, locus-specific 700-1000 bp homology arms were PCR amplified from genomic DNA of E14 mESCs. The 2xGGSG-dTAG-2xHA-P2A-Neo cassette and 2xGGSG-dTAG-2xHA-P2A cassette was PCR amplified from an unpublished targeting vector (a gift from Dr. Helene Damhofer) and the Puro resistance gene was subcloned from pLenti-Puro (a gift from Dr. Jia Xu). The PCR products were assembled into pUC19 plasmid backbone (invitrogen) using In-Fusion cloning kit (Takara). The resulting targeting construct contains left homology arm, 2xGGSG, dTAG, 2xHA tag, P2A, neomycin or puromycin resistance gene, and right homology arm (Table S1).

For CRIPSR sgRNA (knockin and knockout) cloning into enhanced specificity Cas9 (eSpCas9)[104] vectors, oligos ordered from IDT were annealed in T4 ligation buffer (NEB) and phosphorylated using T4 PNK (NEB). The annealed oligos and BbsI (NEB)-linearized U6-sgRNA-eSpCas9(1.1)-T2A-mCherry or U6-sgRNA-eSpCas9(1.1)-T2A-GFP plasmid were ligated using T4 DNA ligase (NEB). sgRNAs of high specificity and efficiency were selected from the 'CRISPR 10K' track of the UCSC genome browser and editing efficiency was assessed using TIDE.[101] Sequences of sgRNAs used in this study are listed in Table S2.

For construction of expression vectors, the coding sequences of mouse *Nsd1* (NSD1.1S) and *Nsd2* (Table S3) were cloned from cDNA of E14 mESCs and assembled into a piggyBac vector (PB-CAG-MCS-HA-IRES-Neo) by In-Fusion cloning. Truncation mutants of NSD1 were generated by subcloning indicated fragments from the full-length *Nsd1* vector and assembled into the PB-CAG-MCS-HA-IRES-Neo vector by In-Fusion cloning. A nuclear localization signal (NLS) was added to ensure nuclear transport. For generating *Nsd1* with point mutations, regions containing Sotos syndrome missense mutations (mouse homologous amino acids) and the N1751Q (also known as N1918Q[24]) mutation were ordered as gBlocks from IDT. Positions of the point mutations were according to NSD1.1S protein sequence. gBlocks with point mutations and PCR-amplified flanking regions were assembled into the PB-CAG-MCS-HA-IRES-Neo vector by In-Fusion cloning. An additional FL NSD1 series (FL, FL ΔPP and FL N1751Q) and NSD2

were cloned into a PB-CAG-MCS-3xFLAG-2xHA-IRES-Zeo and PB-CAG-MCS-3xFLAG-2xHA-IRES-Puro vectors, respectively, by In-Fusion cloning.

pcDNA-dCas9, pcDNA-dCas9-VP64, pcDNA-dCas9-p300^{core}, pcDNA-dCas9-p300^{core} (D1399Y) and pSPgRNA plasmids were purchased from Addgene.[105] sgRNAs targeting the *HBG1* HS2 enhancer[105] (Table S2) were cloned into pSPgRNA plasmid via the BbsI restriction sites. pcDNA-dCas9-NSD1^{C} and pcDNA-dCas9-NSD1^{C} (N1751Q) were generated by subcloning coding sequences of wild type and N1751Q mutant NSD1 C fragments, respectively, into pcDNA-dCas9-VP64, replacing the VP64 coding sequence via the AscI/PacI restriction sites and using In-Fusion cloning.

Generation of degron knockin cell lines

For dTAG knockin at the C terminus of *Nsd1* and *Setd2*, mESCs were co-transfected with a U6-sgRNA-eSpCas9(1.1)-T2A-GFP vector encoding sgRNAs targeting regions near the stop codon, and two donor vectors (Neo and Puro) containing homology arms specific for *Nsd1* or *Setd2* genomic loci and designed for inserting the donor sequence immediately before the stop codon. Transfection was performed using Lipofectamine 3000 (Thermo Fisher) according to manufacturer's instructions. Transfected cells were selected with both neomycin and puromycin 48 h after transfection for 14 days and the double resistant cells were single-cell sorted using a SONY MA900 cell sorter. Resulting clonal cell lines were screened by PCR genotyping for biallelic integration (for primers see Table S4) and further confirmed by western blotting of tagged proteins in the presence or absence of dTAG-13.

Generation of knockout cell lines

Nsd1 knockout mESC lines were generated by CRISPR/Cas9 targeting using a combination of two sgRNAs to excise the *Nsd1* promoters or the indicated C-terminal region. mESCs were co-transfected with two eSpCas9 vectors (GFP and mCherry) encoding the sgRNA pair using Lipofectamine 3000 according to manufacturer's instructions. GFP^{+}/mCherry^{+} cells were single-cell sorted 48 h after transfection using a SONY MA900 cell sorter. Resulting clonal cell lines were screened by PCR genotyping for homozygous deletion (for primers see Table S4) and confirmed by RT-qPCR and western blotting. For p1/p2/p3 knockout combination, CRISPR/Cas9 editing was performed sequentially and established clonal knockout cells were used as parental cells for the next round of knockout.

Transgenic expression in mESCs

For NSD1 rescue expression, a vector encoding the piggyBac transposase (PBase) and a piggyBac vector encoding wild-type or mutant NSD1 (NSD1.1S) were transfected into *Nsd1* KO2 or NSD1-dTAG mESCs in a 2:1 ratio using Lipofectamine 3000 (Thermo Fisher) according to manufacturer's instructions. Cell lines with stable construct integration were generated through neomycin, zeocin or puromycin selection depending on vectors used. Additionally, for expression of FL NSD1 in *Nsd1* KO2 mESCs using the Neo vector, clonal cell lines were further derived from the neomycin resistant population and screened for ones with high transgene expression. For expression of the FL NSD1 series (FL, FL ΔPP and FL N1751Q) using the Zeo vector and NSD2 using the Puro vector in *Nsd1* KO2 mESCs (related to Figures 7I, 7J, 7M–7O, and S3I–S3M), zeocin or puromycin selection alone was performed, which selects for mESCs with relatively high transgene expression. For expression of NSD1 FL WT and N1751Q in NSD1-dTAG mESCs using the Zeo vector, zeocin selection was followed by clonal cell line derivation to achieve more homogenous transgene expression.

dCas9-mediated gene activation in HEK293T cells

Transient expression of the indicated combinations of dCas9 and/or dCas9 fusion proteins as well as pooled sgRNAs targeting the *HBG1* HS2 enhancer were achieved by transfection with Lipofectamine 2000 (Thermo Fisher) according to manufacturer's instructions. Transfections were performed in 12-well plates using 750 ng of respective dCas9 expression vector and 500 ng of equimolar pooled sgRNA expression vectors. For transfection of a single dCas9 expression vector, 1500 ng plasmid was used. RNA extraction and whole-cell extract preparation were done 48 hours after transfection.

Proliferation assay

Cells were seeded into 12-well plates with 2,000 cells per well in triplicate. Cells were counted using a hemocytometer every two days for six consecutive days.

ESC to EpiLC/EpiSC conversion

ESC-EpiLC-EpiSC conversion was performed as previously described[47,48] with modifications. 80,000 mouse ESCs were seeded in a well of 16.7 μg/mL fibronectin-coated 12-well plate in EpiLC medium (N2B27 medium supplemented with 12.5 ng/mL recombinant human basic FGF (FGF2, R&D Systems), 20 ng/mL Activin A (Peprotech), and 1% Knockout Serum Replacement (Gibco); N2B27 medium consists of 50% DMEM/F12 (Gibco), 50% Neurobasal (Gibco), 0.5% N2 supplement (Gibco), 1% B27 supplement (Gibco), 1% Penicillin/Streptomycin (MSKCC), 2 mM L-Glutamine (MSKCC), 50 μM β-mercaptoethanol (Gibco)). The medium was changed after 24 h and ESCs were converted to EpiLCs after 48 h. For further conversion to self-renewal EpiSCs, EpiLC colonies were dissociated using Accutase (Sigma) and 80,000 EpiLCs were seeded in a well of 16.7 μg/mL fibronectin-coated 12-well plate in EpiSC medium (N2B27 medium supplemented with 12.5 ng/mL recombinant human basic FGF (FGF2, R&D Systems), 20 ng/mL Activin A (PeproTech), and 175 nM Wnt inhibitor NVP-TNKS656 (Selleck Chemicals)). The medium was changed after 24 h and EpiLCs were converted to EpiSCs after 48 h. For maintenance of EpiSCs, medium was changed daily, and cells were passaged every

two days using Accutase (Sigma). For forebrain organoid differentiation, EpiSC conversion was done by plating ESCs on irradiated mouse embryonic fibroblast (MEF) feeder cells in EpiSC medium. EpiSC medium was changed daily, and cells were passaged every two days using 0.5 u/μL collagenase IV (STEMCELL Technologies) followed by dissociation with Accutase into small clumps of 3-5 cells. After 2 passages, the ESCs were converted to EpiSCs.

Multilineage EB differentiation

200,000 mESCs were seeded in a well of a non-TC-treated 6-well plate in serum/LIF medium without LIF to allow for aggregate formation and suspension culture. After two days, EBs were transferred to a 10 cm bacterial petri dish. Medium was changed every day and petri dish changed every other day thereafter. For analysis of contracting EBs, EBs in suspension culture on day 8 of differentiation were seeded onto 0.1% gelatin-coated TC-treated plates to promote attachment and stimulate cardiomyocyte differentiation. Contracting EBs were defined by presence of actively beating areas and were counted on day 10 of differentiation in adherent culture. All EB differentiation experiments were replicated at least twice, and a representative experiment was shown.

Forebrain organoid differentiation

Forebrain organoid differentiation was performed as previously described.[47] Briefly, on day 0, NSD1-dTAG EpiSC colonies cultured on irradiated MEFs were detached using 0.5 u/μL Collagenase IV (STEMCELL Technologies), washed once with PBS and dissociated into a single cell suspension using Accutase. Dissociated EpiSCs were seeded into Aggrewell plates (STEMCELL Technologies) pre-coated with anti-adherence rinsing solution (STEMCELL Technologies), such that ~1,000 EpiSCs were distributed into each microwell for EB formation for 24h in N2B27 (B27 without vitamin A) medium containing 100 nM LDN-193189 (Reprocell), 10 μM SB431542 (Tocris), 100 nM PD173074 (STEMCELL Technologies), 4 nM LGK974 (Selleck Chemicals), 50 nM Chroman-1 (MedChem Express), and 5 μM Emricasan (Selleck Chemicals). Day 1 EBs were embedded in Matrigel (Corning) and cultured in Neural Induction Medium (same medium as for EB formation but without Chroman-1 and Emricasan) for 24h. On day 2, medium was changed to Neuroepithelial Expansion Medium (N2B27 (B27 without vitamin A) medium supplemented with 100 ng/mL FGF8b (R&D Systems)) for additional 48h. Day 4 EBs were retrieved from Matrigel using Cell Recovery Solution (Corning), transferred to a bacterial petri dish and cultured in Neuronal Medium (N2B27 (B27 with vitamin A) medium supplemented with 20 ng/mL BDNF (PeproTech) and 20 ng/mL GDNF (PeproTech)) while mixing at 65 rpm. Medium was changed every other day until day 12. Images for day 12 forebrain organoids were acquired on a Leica Dmi1 microscope and processed in ImageJ.[102]

Immunofluoresence

Cells were cultured on gelatin-coated coverslips (NSD1-dTAG and NSD1 rescue mESCs) or directly on gelatin- or fibronectin-coated culture dishes (ESC to EpiSC transition). Cells were fixed with 4% paraformaldehyde in PBS containing Ca^{2+} and Mg^{2+} (PBS+/+) for 15 min and permeabilized with 0.25% Triton X-100 in PBS+/+ for 5 min. Cells were blocked with 5% BSA in PBS+/+ (blocking buffer) for 30 min and then incubated with primary antibodies in blocking buffer overnight at 4°C. Incubation with fluorophore-conjugated secondary antibodies was performed in blocking buffer for 1 h at room temperature. DNA were counterstained with 0.1 μg/mL DAPI for 10 min. Cells were washed with PBS+/+ after fixation (1x 5 min), permeabilization (2x 5 min) and antibody incubation steps (3x 10 min). All steps were performed at room temperature unless otherwise indicated. Antibodies used were listed in Table S5. Coverslips were mounted in Fluoro-Gel mounting medium (Electron Microscopy Sciences) and images were acquired on a Zeiss LSM 880 confocal microscope. For cells cultured directly on the dish, images were acquired using a Leica Dmi8 fluorescent microscope. Images were process in ImageJ.

Intracellular staining and flow cytometry

At day 10 of differentiation, EBs in adherent culture were dissociated into single cells using Accutase (Sigma) for 10 min at room temperature. Intracellular staining was performed using BD Cytofix/Cytoperm Fixation/Permeabilization Kit (BD Biosciences). 1 million cells were fixed and permeabilized with BD Fixation/Permeabilization Solution for 30 min on ice and then washed once with BD Perm/Wash Buffer. For evaluation of cardiomyocyte differentiation efficiency, cells were stained with a mouse monoclonal antibody for cardiac Troponin T (cTnT) at 1:100 dilution in 1x BD Perm/Wash Buffer for 1 h at room temperature. Cells were washed twice with 1x BD Perm/Wash Buffer and then stained with goat anti-mouse IgG Alexa 647 secondary antibody (Thermo Fisher, A-21236) at 1:200 dilution in 1x BD Perm/Wash Buffer for 1 h at room temperature. Cells were then washed twice with 1x BD Perm/Wash Buffer. At least 10,000 cells were analyzed using a BD LSRFortessa flow cytometer and the results were analyzed using FlowJo (BD Bioscience).

Whole-cell extract preparation and western blotting

Whole-cell extracts were prepared using total protein extraction (TOPEX) buffer[103] (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM $MgCl_2$, 0.5% Triton X-100, 1% SDS, 1 mM DTT, 125 U/ml Benzonase (Millipore) and 1x protease inhibitor cocktail). Cells were lysed at room temperature for 10 min and protein concentration was determined by BCA assay (Thermo Fisher). Samples were boiled in Laemmli buffer (Bio-rad) at 70°C for 10 min for immunoblotting following standard protocols using antibodies listed in Table S5. Fluorescence detection of protein bands were performed using the Odyssey CLx imaging system (LI-COR Biosciences).

Sequential salt extraction

Sequential salt extraction was performed as previously described[106] with modifications. Cells were lysed on ice in Buffer A (10 mM HEPES-KOH pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.1% NP-40, 10% glycerol, 1 mM DTT and 1x protease inhibitor cocktail). Supernatant was collected as the cytoplasmic and nuclear-leaked (CNL) fraction. Nuclei pellets were extracted sequentially with mRIPA buffer (50 mM Tris-HCl pH 8.0, 1% NP-40, 0.25% sodium deoxycholate, 1 mM DTT and 1x protease inhibitor cocktail) with 0, 100, 200, 300, 400, 500 and 600 mM NaCl as indicated, by pipetting followed by incubation (3 min) on ice. Supernatant was collected as nuclear salt extraction fractions after each round. The final pellets (insoluble chromatin) were lysed in TOPEX buffer at room temperature for 10 min as the pellet fraction. All fractions were boiled in Laemmli buffer at 70°C for 10 min for western blotting.

Nucleosome pull-down with nuclear extract

Nuclear extracts were prepared essentially as described[107] using *Nsd1* KO2 mESCs expressing NSD1 PP-SET, PP-SET ΔPHD1-4, PP-SET ΔPWWP2, and SET-PC fragments. Cells were washed with PBS and with hypotonic buffer (10 mM HEPES-KOH pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, and 1x protease inhibitor cocktail). Cells were then incubated with hypotonic buffer for 10 min at 4°C to swell and homogenized by 10 strokes through a 20G needle to isolate nuclei. Nuclei were extracted for 30 min at 4°C with head-to-head rotation using freshly prepared high salt buffer (20 mM HEPES-KOH pH 7.9, 420 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 25% glycerol, and 1x protease inhibitor cocktail) supplemented with 0.1% NP-40. The nuclei suspension was centrifuged at 20,000g for 30min at 4°C and the clear supernatants were taken as nuclear extracts. Protein concentration was determined by BCA assay (Thermo Fisher) and DTT was added to a final concentration of 1 mM. Nuclear extracts were snap frozen and stored at -80°C until nucleosome pull-down was performed.

For nucleosome pull-down, Dynabeads MyOne Streptavidin T1 beads (Thermo Fisher) were washed three times using nucleosome immobilization buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.1% NP-40, 1 mM DTT and 1x protease inhibitor cocktail). For each pull-down, 3 ug of biotinylated designer recombinant mononucleosomes with PTM (EpiCypher) was immobilized on 20 ul of Streptavidin T1 bead slurry in nucleosome immobilization buffer for 2 h at 4°C, followed by three washes using nucleosome immobilization buffer. 1%–4% of nuclear extracts were taken as input and boiled with Laemmli buffer for 10 min at 70°C for western blot analysis. The salt concentration of nuclear extracts was lowered to 150 mM by dilution with low salt buffer (20 mM HEPES-KOH pH 7.9, 0.2 mM EDTA, 20% glycerol, 1 mM DTT and 1x protease inhibitor cocktail). NP-40 was added to a final concentration of 0.1%. Nuclear extracts were cleared by centrifugation at 16,000xg for 10 min at 4°C. Supernatants were taken and precleared with 20 ul of Streptavidin T1 bead slurry for 1 h at 4°C. For each pull-down, 0.5 mg of precleared nuclear extracts was incubated with nucleosome-immobilized Streptavidin T1 beads for 4 h at 4°C. Beads were then washed five times with low salt buffer supplemented with 250 mM NaCl and 0.1% NP-40. Protein-bound beads were finally boiled in Laemmli buffer for 10 min at 70°C and pull-down products were analyzed by western blotting.

Quantification of global DNA methylation by mass spectrometry

DNA methylation quantification was performed in triplicates of independent cultures and in two batches, one with NSD1-dTAG mESCs upon 0 and 72h of dTAG-13 treatment and the other with WT and *Nsd1* KO1/2 mESCs. Genomic DNA was extracted using DNeasy Blood and Tissue Kit (Qiagen) with RNase A treatment according to manufacturer's protocol. Single nucleosides were generated by digesting 1 μg of genomic DNA at 37°C for 1 hour using the Nucleoside Digestion Mix (NEB). Nucleosides were cleaned up by using methanol. The digested mixture was mixed with 9x reaction volume of cold LC-MS grade methanol, vortexed vigorously for 5 min at 4°C, and spun at 14,000 rpm for 5 min at 4°C. Supernatants were dried using a speed vac for 30 min at room temperature. Dried nucleoside pellets were stored at -80°C until liquid chromatography-mass spectrometry (LC-MS) was performed as described.[108] Briefly, the dried nucleoside extracts were resuspended in 400 μL of 50% LC-MS grade acetonitrile, vortexed vigorously for 15 sec, and centrifuged at 20,000xg for 20 min at 4°C. The supernatant was transferred to an HPLC vial and 2 μl was injected onto a ZIC-pHILIC 150 x 2.1 mm (5 μm particle size) column (EMD Millipore). LC-MS analysis was conducted on a QExactive benchtop orbitrap mass spectrometer equipped with a heated electrospray ionization (HESI) probe coupled to a Vanquish UPLC System (Thermo Fisher). Integration of extracted ion chromatograms was performed using Skyline Daily (v 21.1)[109] using a 2 ppm mass tolerance and a <6 sec retention time deviation from known standards. 5-methyl-deoxycytidine peak areas were normalized to the 2-deoxycytidine peak area within each sample.

RRBS

RRBS was performed in duplicates of independent cultures and in two batches, one with NSD1-dTAG mESCs upon 0, 6 and 24h of dTAG-13 treatment and the other with WT and *Nsd1* KO2 mESCs. Genomic DNA was extracted using DNeasy Blood and Tissue Kit (Qiagen) with RNase A treatment according to manufacturer's protocol. Reduced representation bisulfite sequencing library was prepared using 2 μg of genomic DNA spiked with 0.1% (w/w) unmethylated λ DNA. Genomic DNA was digested by incubating overnight with Msp I (NEB) at 37°C and purified using MinElute PCR Purification Kit (Qiagen). 1 μg of Msp I-digested DNA was subjected to end repair, clean-up, 3′ adenylation, adapter ligation, and size selection clean-up using the NEXTFLEX Bisulfite Library Prep Kit (PerkinElmer) according to manufacturer's protocol. Bisulfite conversion of DNA was then performed using the EZ DNA Methylation-Lightning Kit (Zymo Research) according to manufacturer's protocol. DNA was amplified using the KAPA HiFi HotStart Uracil+ ReadyMix Kit (Roche) according to manufacturer's protocol and cleaned up using Agencourt AMPure XP magnetic beads (Beckman Coulter). Library quality control was performed with Qubit fluorometer and TapeStation. Barcoded libraries were multiplexed and subjected to 75-bp single-end sequencing with an Illumina NextSeq 550 instrument.

ChIP

Cells were crosslinked in single-cell suspension with 1% formaldehyde (Sigma) for 10 min at room temperature. Crosslinking was quenched with 0.125 M glycine for 5 min. All following steps until DNA purification were performed at 4°C unless otherwise indicated. Cells were washed twice with PBS and once with Cell Lysis Buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 1 mM EDTA, 0.2 % NP-40, 10 % Glycerol, 1 mM DTT and 1x protease inhibitor cocktail). For nuclei isolation, cells were incubated with Cell Lysis Buffer for 10 min. Nuclei pellets were resuspended with Nuclei Lysis Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.5 % SDS, 1 mM DTT and 1x protease inhibitor cocktail), incubated for 10 min, and sonicated for 15 cycles (30 s ON/30 s OFF per cycle) at high intensity using a Bioruptor (Diagenode). DNA was sheared to 200-500 bp fragments. Nuclear lysate was cleared by centrifugation at 18,000g for 10 min. Chromatin concentration of nuclear lysate was quantified spectroscopically (absorbance A260). 50-200 ug of chromatin was used for each immunoprecipitation and 1%–2% chromatin was taken as input. For ChIP, nuclear lysate was diluted 1:5 with ChIP Dilution Buffer (12.5 mM Tris-HCl pH 8.0, 187.5 mM NaCl, 1.25% Triton X-100, 1 mM DTT and 1x protease inhibitor cocktail). For RNA Pol II NTD, S2P and S5P ChIP, 5 μg of *Drosophila* spike-in chromatin was added per 100 μg of mESC chromatin. Antibodies (Table S5) were added, including anti-H2Av Spike-in antibody for RNA Pol II and pause-release/elongation factor ChIPs, and immunoprecipitation was performed overnight followed by incubation with Protein G Dynabeads (Thermo Fisher) for 2h. Beads were washed once with Low Salt Wash Buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 % Triton X-100, and 0.1% SDS), twice with High Salt Wash Buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 1 % Triton X-100, and 0.1% SDS), once with LiCl Wash Buffer (20 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 1 % NP-40, and 1% sodium deoxycholate), and once with TE Buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA). For elution, beads were incubated with fresh Elution Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1 % SDS, 100 mM NaHCO3, and 250 mM NaCl) twice, each for 30 min at 65°C while mixing in a thermomixer at 900 rpm, and eluates were pooled. To reverse crosslink, ChIP and input DNA were incubated at 65°C overnight in Elution Buffer with 100 μg/mL RNase A, followed by addition of Protease K to 400 μg/ml and incubation for 2h at 55°C. ChIP and input DNA were purified using a QIAquick PCR purification kit (Qiagen). qPCR was performed using PowerUp SYBR Green Master Mix (Applied Biosystems) with primers listed in Table S6. ChIP-seq libraries were prepared with 3 ng ChIP or input DNA using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) according to manufacturer's protocol. Bead based sized selection was performed to enrich for DNA with ~250 bp inserts. Library quality control was performed using Qubit fluorometer and TapeStation. Barcoded libraries were multiplexed and subjected to 35-bp paired-end sequencing with an Illumina NextSeq 550 instrument.

CUT&RUN

CUT&RUN was performed as previously described[45] with modifications. All wash and mixing steps were performed by gentle pipetting for around 5 times. 20 μl per sample of BioMag Plus Concanavalin A (Con A)-coated magnetic beads (Bangs Laboratories) were activated by washing three times each with 100 μl cold Binding Buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$). Activated Con A beads were resuspended with 10 μl cold Binding Buffer in 0.2 ml strip tubes. 0.5 x 10^6 cells were harvested per CUT&RUN and washed three times each with 100 μl Wash Buffer (20 mM HEPES-NaOH pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 1x proteosome inhibitor cocktail) at room temperature. Cells were resuspended in 100 μl Wash Buffer, mixed with 10 μl activated Con A beads and incubated for 10 min at room temperature to attach to Con A beads. Cell permeabilization and primary antibody binding was performed by adding 50 μl of cold Antibody Buffer (20 mM HEPES-NaOH pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 0.01% Digitonin, 2 mM EDTA, 1x proteosome inhibitor cocktail) to the beads and incubating overnight at 4°C on a nutator with cap side of the strips elevated. Antibodies used were listed in Table S5. The following steps were performed in a 4°C cold room until MNase digestion was stopped. To wash away unbound antibodies, beads were washed three times each with 200 μl cold Digitonin Buffer (20 mM HEPES-NaOH pH 7.5, 150 mM NaCl, 0.5 mM Spermidine, 0.01% Digitonin, 1x proteosome inhibitor cocktail). Beads were then resuspended with 50 μl cold Digitonin Buffer and pAG-MNase was added at a 1:1000 ratio, followed by incubation at 4°C for 1 h on a nutator. Unbound pAG-MNase was washed away by washing the beads three times each with 200 μl cold Digitonin Buffer. Beads were resuspended with 50 μl cold Digitonin Buffer and pAG-MNase was activated by adding $CaCl_2$ to 2 mM. Targeted chromatin cleavage was carried out by incubation at 4°C for 30 min on a nutator. The reaction was stopped by adding 50 μl cold 2x Stop Buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 0.01% Digitonin, 50 μg/ml RNase A, 1 pg/ul *Drosophila* spike-in DNA). Cleaved chromatin was released by incubating at 37°C for 30 min (mix every 10 min). Beads were magnetically collected, and CUT&RUN DNA (> 50 bp) was extracted from the supernatant using the Monarch DNA Cleanup Kit (NEB). CUT&RUN library was prepared with no more than 10 ng CUT&RUN DNA using NEBNext Ultra II DNA Library Prep Kit (NEB) according to manufacturer's protocol with modifications: end prep was performed for 30 min at 20°C followed by 1 h at 50°C to prevent melting of small fragments, adaptor-ligated DNA was cleaned up without size selection, and the anneal/extension step of PCR amplification was performed for 10 s at 65°C to favor exponential amplification of the desired CUT&RUN fragments. Library quality control was performed with Qubit fluorometer and TapeStation. Barcoded libraries were multiplexed and subjected to 35-bp paired-end sequencing with an Illumina NextSeq 550 instrument.

Purification of pAG-MNase

The pAG-ERH-MNase-6xHis-HA plasmid[110] (a gift from Dr. Steve Henikoff) was transformed into JM101 competent cells (Agilent Technologies) and a single transformed colony was picked up and grew in LB medium containing 50 μg/mL Kanamycin at 37°C. When OD_{600} reached 0.7, IPTG was added to 2 mM and cells were induced for 3 hours. Cells from 1 L of culture were pelleted and resuspended in 30 mL lysis buffer (10 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 mM imidazole) and lysed by sonication. After

cleared by centrifugation (18,000 rpm, 60 min, 4°C), the lysate supernatant was incubated with 1 mL pre-equilibrated Ni-NTA agarose (QIAGEN) for 60 min at 4°C. After washing the beads with 30 mL wash buffer (10 mM Tris-HCl pH 7.5, 300 mM NaCl, 0.03% Zwittergent, 20 mM imidazole), the bound proteins were eluted with elution buffer (10 mM Tris-HCl pH 7.5, 300 mM NaCl, 0.03% Zwittergent, 250 mm imidazole) by 1 mL fractions. Eluates were analyzed by SDS-PAGE, pooled, and dialyzed into storage buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1 mM PMSF). Glycerol was then added to 50% and aliquots were stored at -80°C.

Preparation of spike-in chromatin and DNA

Drosophila S2 cells were crosslinked, lysed, and sonicated and chromatin concentration was determined as described in the ChIP section. The resulting fragmented chromatin was used as ChIP-seq spike-in control. Crosslink of the sheared chromatin was reversed, and DNA purified using a QIAquick PCR purification kit (Qiagen), which was further subjected to double-sided bead purification with Agencourt AMPure XP magnetic beads (Beckman Coulter) to select for DNA of 100-500 bp. The resulting DNA was used as CUT&RUN spike-in control.

ATAC-seq

ATAC-seq was performed as previously described[111] with modifications. ATAC-seq was performed in duplicates of independent cultures. 50,000 cells were washed once with 50 μL cold PBS in tubes pre-coated with 1% BSA. Cells were lysed in 50 μL cold Lysis Buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 0.1% NP-40, and 0.01% Digitonin) for 5 min on ice. Lysis was stopped by adding 1 mL of cold Wash Buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, and 0.1% Tween-20) and nuclei were spun down by centrifuging at 600xg for 10 min at 4°C. Nuclei pellet was resuspended in 50 μL cold Transposition Mix (25 μL 2x TD buffer (Illumina), 2.5 μL Tn5 transposase (Illumina), 16.5 μL PBS, 0.5 μL 1% Digitonin, 0.5 μL 10% Tween-20, and 5 μL H_2O) and the reaction was incubated at 37°C for 30 min in a thermomixer with 1000 rpm mixing. Transposed genomic DNA was purified using a Zymo DNA Clean and Concentrator-5 Kit and subjected to PCR amplification using NEBNext High-Fidelity 2x PCR Master Mix with custom Nextera PCR primers[112] ordered from IDT. Optimal PCR cycles were determined by qPCR using partially amplified libraries. Libraries were cleaned up and size-selected using double-sided bead purification with Agencourt AMPure XP magnetic beads (Beckman Coulter) to remove primer dimers and fragments longer than 1 kb. Library quality control was performed with Qubit fluorometer and TapeStation. Barcoded libraries were multiplexed and subjected to 35-bp paired-end sequencing with an Illumina NextSeq 550 instrument.

RNA extraction, cDNA synthesis, and RT-qPCR

Total RNA was extracted using RNeasy Plus Mini Kit (Qiagen) according to manufacturer's protocol. 1 μg of total RNA was subjected to reverse transcription using First-Strand cDNA Synthesis System (ORIGENE). qPCR was performed using PowerUp SYBR Green Master Mix (Applied Biosystems) with primers listed in Table S7. Heatmaps of RT-qPCR results were generated using Morpheus (https://software.broadinstitute.org/morpheus).

RNA-seq

RNA-seq was performed in triplicates of independent cultures. Total RNA was extracted using RNeasy Plus Mini Kit (Qiagen). 1 μg of total RNA, spiked in with 2 μL 1:100 diluted ERCC RNA Mix 1 (Thermo Fisher), was used for poly(A) RNA selection using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB). Strand-specific RNA libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB) according to manufacturer's protocol. Quality of libraries was analyzed using Qubit fluorometer and Agilent TapeStation. Barcoded libraries were multiplexed and subjected to 75-bp single-end sequencing with an Illumina NextSeq 550 instrument.

SLAM-seq

SLAM-seq was performed as previously described[53] with modifications. SLAM-seq was performed in triplicates of independent cultures. mESCs at 60%–80% confluency were treated with medium containing 100 μM of 4-thiouridine (4sU, Biosynth) for 2 h to metabolically label newly synthesized RNA. Cells were harvested by direct lysis with Buffer RLT plus (Qiagen) supplemented with 40 mM DTT. Total RNA was extracted using RNeasy Plus Mini Kit (Qiagen) according to manufacturer's protocol. DTT was added to all buffers during RNA isolation to keep the samples under reducing conditions (0.1 mM DTT for Buffers RW1 and RPE and 1 mM DTT for RNase-free water used for elution). 1 μg of total RNA was spiked in with 2 μL 1:100 diluted ERCC RNA Mix 1 (Thermo Fisher) and subjected to thiol-linked alkylation by iodoacetamide (IAA, Sigma) for 15 min at 50°C in a mixture containing 10 mM IAA, 50 mM $NaPO_4$ pH 8.0, 50% DMSO. Alkylation was stopped by adding DTT to 20 mM (samples were protected from white light until this step) and RNA was re-purified by ethanol precipitation. 300 ng of alkylated RNA was used for 3′-end polyA RNA sequencing library preparation with the QuantSeq 3′ mRNA-Seq Library Prep FWD Kit for Illumina (Lexogen) according to manufacturer's protocol. Quality of libraries was analyzed using Qubit fluorometer and Agilent TapeStation. Barcoded libraries were multiplexed and subjected to 75-bp single-end sequencing with an Illumina NextSeq 550 instrument.

TT_{chem}-seq

TT_{chem}-seq was performed as previously described.[113] TT-seq was performed in duplicates of independent cultures. Briefly, one 10-dish of mESCs at 60%–80% confluency was used per replicate. 4sU (Biosynth) was added directly to the culture medium to a

final concentration of 1 mM and incubated with the cells for 15 min to label newly transcribed RNA. Total RNA was purified using TRIzol/chloroform extraction followed by isopropanol precipitation. 1% (w/w) 4sU-labeled *Drosophila* S2 cell RNA (a gift from Dr. Hua Wang) was spiked in to 4sU-labeled mESC RNA. RNA was fragmented using NaOH and cleaned up with Micro Bio-Spin P-30 gel columns (Bio-Rad). Fragmented 4sU-RNA was biotinylated with MTSEA biotin-XX linker (Biotium) and purified using phenol/chloroform extraction followed by isopropanol precipitation. Biotinylated 4sU-RNA was pulled down using μMACS streptavidin MicroBeads in combination with μColumns from the μMACS Streptavidin Kit (Miltenyi) and cleaned up using RNeasy MinElute Cleanup Kit (Qiagen) (1.5x (v/v) ethanol was added to the RLT buffer to retain <200-nt RNA fragments). 4sU-RNA was quantified by Qubit fluorometer and quality checked by Agilent TapeStation. ~60-90 ng of 4sU-RNA was used to prepare strand-specific RNA libraries using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB) according to manufacturer's protocol for use with rRNA depleted FFPE RNA. Quality of libraries was analyzed using Qubit fluorometer and Agilent TapeStation. Barcoded libraries were multiplexed and subjected to 75-bp single-end sequencing with an Illumina NextSeq 550 instrument.

CUT&RUN, ChIP-seq, and ATAC-seq analysis

Bcl files containing raw sequencing data were converted to fastq format, adaptor trimmed, and demultiplexed using bcl2fastq2 (v2.20.0.422, Illumina). Quality of the sequencing data was verified using FastQC (v0.11.8) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Paired-end sequencing reads were aligned to the mouse genome (GRCm38.p6/mm10) using Bowtie2 (v2.3.4.1)[88] with "–local –very-sensitive-local –phred33 –no-unal –no-mixed –no-discordant -I 10 -X 700 -S". For CUT&RUN and ChIP-seq, reads were also aligned to the *drosophila* genome (dm6) with additional "–no-overlap" to map spike-in reads. For downstream spike-in normalization, the total number of mapped reads to each genome was used to generate scale factors assuming that the *Drosophila* spike-ins are equally present in each sample. Sam files were converted to bam files, sorted, and indexed using samtools (v1.9).[89] Peaks were called using MACS2 (v2.2.6)[90] callpeak function with "-f BAMPE –keep-dup 1 -q 0.05". For CUT&RUN peak calling, datasets generated using NSD1 degraded (72h) cells, *Nsd1* KO cells, or IgG was used as controls. Peaks from NSD1 CUT&RUN in NSD1-dTAG mESCs were additionally filtered with a stringent cutoff at $q < 0.000001$. For ChIP-seq peak calling, input was used as control. Intersection of peaksets were performed using the intersectBed program from bedtools (v2.27.1).[91] Motif analysis was performed using the findMotifsGenome program from HOMER (v4.11)[92] with "-size 500 -S 20 -mask". Bigwig pileup files were generated using the bamCoverage program from DeepTools (v3.2.1)[93] with "-bs 10 –extendReads –normalizeUsing RPKM –exactScaling –skipNonCoveredRegions" and filtered for mouseENCODE blacklisted genomic regions (v2)[114] using the "-bl" option. For NSD1 and H3K36me2 CUT&RUN in NSD1-dTAG cells, SETD2 and H3K36me3 CUT&RUN in SETD2-dTAG cells, and NSD1 C series and H3K36me2 CUT&RUN in rescue cells, scale factors calculated using the percentage of spike-in reads were applied to the "–scaleFactor" option of bamCoverage for cell number based normalization. In all other cases, the spike-in read percentage was very close and a scaling factor of 1 was applied, i.e. only sequencing depth based normalization was performed. Log_2 fold change bigwig files were further generated using the bigwigCompare program from DeepTools with "-bs 10 –operation log_2 –skipZeroOverZero –skipNonCoveredRegions".

Genome browser tracks of bigwig pileups were generated using IGV (v2.5.1).[94] Genome-wide Spearman's correlation analysis was performed using multiBigwigSummary bins program from DeepTools using 10 kb bins. The results were plotted using the plotCorrelation program from DeepTools either as heatmaps of correlation coefficients or as scatterplots depicting natural log-transformed read density. Heatmaps and average profiles of bigwig pileups at peaksets, enhancers or genes were generated using the computeMatrix, plotProfile and plotHeatmap programs from DeepTools. For analysis with genes, the protein coding genes from mouse Ensembl genes 98 version (GRCm38.p6/mm10) were used. For boxplot analysis with R (v4.0.3) (www.r-project.org), CUT&RUN, ChIP-seq and ATAC-seq read density or log_2 fold change of read density at individual genomic features were calculated using multiBigwigSummary BED-file program from DeepTools, with each genomic feature as a bin. Pausing index was defined as the ratio of RNA Pol II read density in the promoter proximal region bin (-100 bp to +300 bp around the TSS) to RNA Pol II read density in the transcribed region (gene body) bin (+300 bp to the TES). RNA Pol II read density was calculated using multiBigwigSummary BED-file program from DeepTools. ECDF plot of pausing indexes was generated using the stat_ecdf() function from the ggplot2 package in R.

For analysis of localization of NSD1 peaks relative to distinct genomic features through intersection, promoters were defined as 1 kb regions centered at the TSS, gene bodies as +500 bp to the TES, intergenic regions as regions not overlapping with promoters or gene bodies, and promoter-distal regions as all non-promoter regions, i.e. gene bodies and intergenic regions. Only protein coding genes were analyzed. Active promoters were defined by the presence of H3K4me3 and absence of H3K27me3, and bivalent promoters by the presence of both H3K4me3 and H3K27me3. Active enhancers were defined as promoter-distal regions with both H3K4me1 and H3K27ac, and poised enhancers as promoter-distal regions with H3K4me1 but without H3K27ac. Random 10 kb genomic regions were selected as controls. For all other analysis, active enhancers were defined as 1 kb intervals centered on ATAC-seq peak summits, overlapped with H3K27ac CUT&RUN peaks, and more than 500 bp away from TSS. Super-enhancers were called from H3K27ac CUT&RUN signal using ROSE (v1.0.0)[46] with "-s 12500 -t 2500", so that H3K27ac peaks within 12.5 kb were stitched together and regions within 2.5 kb of TSSs were excluded. Enhancer-gene association were performed by assigning enhancers to the nearest TSS of an active gene (RPKM >1 by RNA-seq) using the closestBed program from bedtools and restricting promoter-enhancer proximity to 100 kb.[115] NSD1-bound genes were defined as genes with NSD1-bound active enhancer or promoter. NSD1-bound active enhancers or promoters were those overlapping with NSD1 peaks. To specifically associate NSD1

and H3K36me2 peaks with genes, peaks were assigned to the nearest TSS of an active gene within 100 kb. *De novo* enhancers from day 6 EBs were defined as unique H3K27ac peaks in day 6 EBs vs. ESCs. *De novo* enhancers associated with EB d6 dTAG-13 down genes were further identified by restricting promoter-enhancer proximity to 100 kb.

RRBS analysis

Raw sequencing reads were 3′ trimmed for quality (<20) and adapter content using Trim Galore (v0.6.4) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The trimmed sequence reads were C(G) to T(A) converted and mapped using Bismark (v0.19.0) (https://www.bioinformatics.babraham.ac.uk/projects/bismark/) to similarly converted reference mouse genome mm10 using default Bowtie2 settings. Uniquely aligned reads were retained and duplicated reads were discarded. The remaining alignments were then used for cytosine methylation calling by Bismark methylation extractor. CHG and CHH were discarded to focus on CpG methylation. CpGs were further filtered by C coverage >= 5 reads in all samples and 1kb windows with >= 3 CpGs were selected for further analysis. Genome-wide correlation analysis was performed using 1kb bins.

SLAM-seq analysis

Raw sequencing data were processed using bcl2fastq as described for CUT&RUN and quality checked using FastQC. 3′ UTR annotations of the GRCm38.p6/mm10 assembly were obtained from the UCSC table browser (https://genome.ucsc.edu/cgi-bin/hgTables). 3′ UTR annotations were assigned to Ensembl gene IDs and collapsed using the mergeBed program from bedtools. For alignment to the mouse genome (GRCm38.p6/mm10) and ERCC spike-in sequences and read counting at 3′ UTRs, single-end sequencing reads were processed using the slamdunk all program from SlamDunk (v0.4.3)[95] with ''-n 100 -m'' and default parameters. For gene-level analysis, reads mapped to different 3′ UTRs of the same gene were summed up using the alleyoop collapse program from SlamDunk. Newly synthesized RNA was measured in the context of total RNA by detection of 4sU through thymine-to-cytosine (T>C) conversion. Differential gene expression analysis was performed using DESeq2 (v1.30.0)[96] with reads that contain more than one T>C conversions (newly synthesized RNA) or all reads (total RNA). Size factors were calculated using corresponding total RNA reads for global normalization. Genes with $q < 0.05$ were considered to have significant expression changes. MA plots were generated using DESeq2 with shrinkage of fold changes by incorporating zero-centered normal prior. Venn diagrams were generated using BioVenn (https://www.biovenn.nl/index.php).[100]

TT_{chem}-seq analysis

Raw sequencing data were processed using bcl2fastq as described for CUT&RUN and quality checked using FastQC. Single-end sequencing reads were aligned to both the mouse genome (GRCm38.p6/mm10) and the *drosophila* genome (dm6) using STAR (v2.6.0a)[97] with ''–outFilterMismatchNoverLmax 0.03'' and default parameters. The number of reads mapped to each gene was counted using STAR with ''–quantMode GeneCounts''. For cell number based normalization, a *drosophila* gene-level count matrix was generated and passed to the estimateSizeFactors function of DESeq2 to calculate scale factors. Scaled bigwig pileup files of mouse alignments were generated using the bamCoverage program from DeepTools with ''-bs 10 –normalizeUsing CPM –scaleFactor 1/(DESeq2 size factors calculated using *Drosophila* spike-in alignments) –exactScaling –minMappingQuality 255 –skipNonCoveredRegions''. Average profiles were generated using the computeMatrix and plotProfile programs from DeepTools.

RNA-seq analysis

Raw sequencing data were processed using bcl2fastq as described for CUT&RUN and quality checked using FastQC. Single-end sequencing reads were aligned to both the mouse genome (GRCm38.p6/mm10) and the ERCC spike-in sequences using STAR with default parameters. The number of reads mapped to each gene was counted using STAR with ''–quantMode GeneCounts''. The percentage of ERCC spike-in reads was very similar across conditions and ERCC reads were therefore disregarded for downstream analysis. Differential gene expression analysis was performed using DESeq2 with default settings and cutoffs indicated in figure legends. MA plots and PCA plots were generated using DESeq2. Z-score heatmap was generated using the heatmap.2() function from the gplots package in R. Gene ontology analysis was performed using Enrichr (https://maayanlab.cloud/Enrichr).[98] RPKM (reads per kilobase per million mapped reads) gene expression values were calculated using Cufflinks (v2.2.1).[99]

QUANTIFICATION AND STATISTICAL ANALYSIS

Calculations and statistical analyses were performed using Microsoft Excel, GraphPad Prism and R. The statistical details of the experiments are provided in the figure legends and method details wherever applicable. Welch's t-test was used in Figures 2C, 6A, 6E, 7H, and S5E. Student's t-test was used in Figures 7E, 7F, 7J, 7L, and S7K. Boxplots were generated using R. Boxes span the lower to upper quartile, median is indicated with a center line, notches display the confidence interval around the median, and whiskers show 1.5 times the interquartile range.