# DOT1L bridges transcription and heterochromatin formation at mammalian pericentromeres

Aushaq B Malla[1] (ID), Haoming Yu[1] (ID), Delaney Farris[1] (ID), Srilekha Kadimi[1], TuKiet T Lam[2,3],
Andy L Cox[1] (ID), Zachary D Smith[1,4] & Bluma J Lesch[1,5,*] (ID)

Abstract

Repetitive DNA elements are packaged in heterochromatin, but many require bursts of transcription to initiate and maintain long-term silencing. The mechanisms by which these heterochromatic genome features are transcribed remain largely unknown. Here, we show that DOT1L, a conserved histone methyltransferase that modifies lysine 79 of histone H3 (H3K79), has a specialized role in transcription of major satellite repeats to maintain pericentromeric heterochromatin and genome stability. We find that H3K79me3 is selectively enriched relative to H3K79me2 at repetitive elements in mouse embryonic stem cells (mESCs), that DOT1L loss compromises pericentromeric satellite transcription, and that this activity involves possible coordination between DOT1L and the chromatin remodeler SMARCA5. Stimulation of transcript production from pericentromeric repeats by DOT1L participates in stabilization of heterochromatin structures in mESCs and cleavage-stage embryos and is required for preimplantation viability. Our findings uncover an important role for DOT1L as a bridge between transcriptional activation of repeat elements and heterochromatin stability, advancing our understanding of how genome integrity is maintained and how chromatin state is set up during early development.

Introduction

Control of transcription by chromatin is central to metazoan development and frequently disrupted in disease. Most regions of the genome are packaged by either open, transcriptionally active euchromatin, or compact, transcriptionally repressed heterochromatin. Repeat sequences such as satellite elements and retrotransposons are classic examples of constitutively repressed heterochromatic elements. Paradoxically, RNA transcripts from repeat sequences are often required to establish or maintain the heterochromatic state, meaning that these regions must periodically accommodate active transcription (Fukagawa *et al*, 2004; Zofall & Grewal, 2006; Grewal & Jia, 2007; Frescas *et al*, 2008; Bulut-Karslioglu *et al*, 2012; Millanes-Romero *et al*, 2013; Velazquez Camacho *et al*, 2017; Seczynska *et al*, 2022). To date, most research on chromatin regulation at repetitive elements has focused on factors that set up and maintain transcriptional repression. A major unsolved problem is how coordinated pulses of transcriptional activity are enacted in the context of heterochromatin at repetitive sequences, and how this transcriptionally active state can be rapidly reversed to restore epigenetic silencing.

Chromatin composition and function is determined in part by covalent modifications to histone proteins, which modulate interactions between nucleosomes and chromatin binding factors (Hyun *et al*, 2017). DOT1L (Disruptor of telomeric silencing 1 like, also called KMT4) is an evolutionarily conserved histone methyltransferase that catalyzes mono-, di-, and tri-methylation of histone H3 at lysine (K) 79 (H3K79me1/2/3; Feng *et al*, 2002). DOT1L and H3K79 methylation are typically associated with transcriptional activation at single-copy genes (Guenther *et al*, 2008; Steger *et al*, 2008; Wood *et al*, 2018). H3K79me2 accumulates downstream of the promoter region at actively transcribed genes, and DOT1L interacts directly with the C-terminal domain of RNA Polymerase II (PolII) and with several PolII-associated elongation complexes (Mohan *et al*, 2010; Kim *et al*, 2012; Jonkers *et al*, 2014; Veloso *et al*, 2014; Wood *et al*, 2018). In leukemia, DOT1L interacts with elongation factors fused to the transcriptional coactivator MLL1 (mixed lineage leukemia 1) and promotes constitutive transcriptional activation of genes responsible for leukemogenesis (Okada *et al*, 2005; Guenther *et al*, 2008; Bernt *et al*, 2011).

However, other data have suggested a contrasting role for DOT1L in heterochromatin regulation and formation. Loss of *Dot1* or mutation of the H3K79 residue impair telomeric silencing in budding yeast (Ng *et al*, 2002). In mouse embryonic stem cells (mESCs), *Dot1l* mutation reduces the constitutive heterochromatin

1   Department of Genetics, Yale School of Medicine, New Haven, CT, USA
2   Keck MS & Proteomics Resource, Yale School of Medicine, New Haven, CT, USA
3   Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA
4   Yale Stem Cell Center, Yale School of Medicine, New Haven, CT, USA
5   Yale Cancer Center, Yale School of Medicine, New Haven, CT, USA
    *Corresponding author. Tel: +1 203 737 1074; E-mail: bluma.lesch@yale.edu

modifications H3K9me2 and H4K20me3 at large arrays of tandem repeats that make up the bulk of centromeric and pericentromeric regions, with reciprocal gains in the euchromatin-associated modification H3K9ac (Jones et al, 2008). H3K79me3 has also been reported to localize to pericentromeric heterochromatin by immunofluorescence (Ooga et al, 2008; Abe et al, 2011) and overexpression of full-length and mutant DOT1L in zygotes was reported to disrupt organization of pericentromeric heterochromatin in preimplantation embryos (Ooga et al, 2013), raising the possibility that DOT1L is required for regulation of transcriptional activity at these regions.

Pericentromeric heterochromatin (PCH) is required for transposon suppression and chromosome segregation, and coordinates heterochromatin formation and maintenance across the genome (Janssen et al, 2018). Pericentromeric heterochromatin is composed of tandem arrays of specific repeating sequences called satellites; in mouse, this 234 base pair repeating sequence element is called the major satellite. Although packaged in heterochromatin and generally silent, major satellite DNA is transcribed at specific times in the cell cycle and during development. Major satellite transcription is cell cycle dependent in mouse fibroblasts, with transcript levels peaking near the G1/S boundary and persisting through mitosis (Lu & Gilbert, 2007). In mESCs, transcription factors such as NANOG, SALL1, and YY1 are recruited to promoters within pericentromeric regions and may contribute to their transcriptional activity (Shestakova et al, 2004; Novo et al, 2016). A sudden burst of major satellite transcription occurs in the early cleavage stages following fertilization and is thought to be essential for reorganizing heterochromatin into densely stained structures called chromocenters prior to further developmental progression (Probst et al, 2010; Casanova et al, 2013). In later development, major satellite transcripts accumulate in the mouse central nervous system from embryonic Day 11.5 (E11.5) to E15.5, and are abundant in adult liver and testis (Rudert et al, 1995). Transcriptional activity at major satellite sequences is proposed to recruit heterochromatin protein 1 (HP1) and stabilize binding of SUV39H enzymes, suggesting that major satellite transcription is essential for heterochromatinization at PCH (Frescas et al, 2008; Probst et al, 2010; Johnson et al, 2017; Shirai et al, 2017; Velazquez Camacho et al, 2017). In keeping with a requirement for appropriate major satellite regulation in maintaining normal cellular function, aberrant upregulation of satellite transcripts also occurs in human pathological states, including several cancers (Ting et al, 2011; Bersani et al, 2015).

Despite its importance in genome regulation and integrity, our understanding of how intermittent transcriptional activation is accomplished in heterochromatin elements such as PCH remains limited. Here, we reveal a critical role for DOT1L in promoting pericentromeric transcription in mESCs and mouse preimplantation embryos. We show that DOT1L selectively enhances transcriptional activity at major satellite elements and highlight the chromatin regulator SMARCA5 as a potential DOT1L partner that contributes to major satellite expression. In cleavage-stage embryos, DOT1L inhibition leads to loss of major satellite expression, cell cycle arrest, and lethality. Our work introduces DOT1L as a transcriptional activator at PCH, helps to explain past observations that DOT1L contributes to heterochromatin regulation, and suggests that DOT1L activity is required for establishment of heterochromatin structure and embryo viability.

Results

H3K79me3 is enriched at repetitive elements

To understand the distribution of H3K79 methylation, we performed co-staining of H3K79me2 or H3K79me3 with markers for mitosis (H3S10P) and pericentromeric heterochromatin (HMGA1, High Mobility Group protein A1; Jagannathan et al, 2018) in mouse embryonic stem cells (mESCs). The distribution of H3K79me2 is cell cycle dependent: it localized to chromosome arms during mitosis, and to euchromatin in interphase cells (Fig 1A and B). In contrast, H3K79me3 strongly localized to HMGA1-enriched pericentromeric regions both in mitotic cells and at interphase (Fig 1A and B). We observed the same pattern in the human colorectal cancer cell line HCT116, indicating that this phenomenon is generalizable beyond mESCs (Fig EV1A and B). These results suggest that H3K79me2 and H3K79me3 are differentially enriched within chromatin, with H3K79me2 accumulating in euchromatic and H3K79me3 in more heterochromatic regions, a conclusion supported by previous observations in fibroblasts (Ooga et al, 2008).

To understand their genomic distribution in greater detail, we mapped both H3K79me2 and H3K79me3 in mESCs using chromatin immunoprecipitation followed by sequencing (ChIP-seq). To verify specificity of the ChIP antibody, we generated Dot1l knockout (KO) mESCs using CRISPR/Cas9 and confirmed strong depletion of DOT1L, H3K79me2, and H3K79me3 by immunoblot and immunofluorescence in Dot1l KO cells (Table EV1 and Fig EV1C–E). ChIP-seq signal was strongly correlated between biological replicates and well correlated with published data for both H3K79me2 and H3K79me3 (Fig EV1F and G). We detected very few peaks (2 H3K79me2 and 67 H3K79me3 peaks) in Dot1l KO cells, validating ChIP specificity. We eliminated these peaks as artifacts, defining a final set of 46,409 H3K79me2 peaks and 5,167 H3K79me3 peaks in mESCs (Fig EV1H). Enriched gene ontology (GO) categories for genes near H3K79me2 and H3K79me3 peaks were similar and consistent with known roles for DOT1L, including functions related to chromatin organization, cellular response to DNA damage, and cell cycle regulation (Fig 1C and Dataset EV1).

In keeping with our immunofluorescence data, we observed that the genomic distributions of H3K79me2 and H3K79me3 were distinct (Fig EV2A), and a large fraction of H3K79me2 and H3K79me3 peaks do not overlap (41% of H3K79me3 and 93% of H3K79me2 peaks; Fig 1D and E), and the sets of peaks are enriched for different sequence motifs (Fig EV2B). As expected, H3K79me2 signal was found almost exclusively in gene bodies, with the strongest enrichment immediately downstream of the transcription start site (TSS; Figs 1F, and EV2C and D). In contrast, one-quarter of H3K79me3 peaks are located at distal intergenic sites (Figs 1F and EV2C). Correspondingly, many more single-copy genes were associated with H3K79me2 than H3K79me3 peaks (8,726 vs. 2,129 genes; Dataset EV1). H3K79me2 signal appeared stronger even at the subset of genes that were also enriched for H3K79me3 (Fig EV2D and E), consistent with a preferential enrichment for H3K79me2 over H3K79me3 near the TSS of single-copy genes.

On the contrary, our data suggested that H3K79me3 was more likely than H3K79me2 to be enriched at repeat elements. Twice as many reads in our H3K79me3 libraries were non-uniquely mapped (mean 29.6% of reads for H3K79me3 compared to 15.5% for
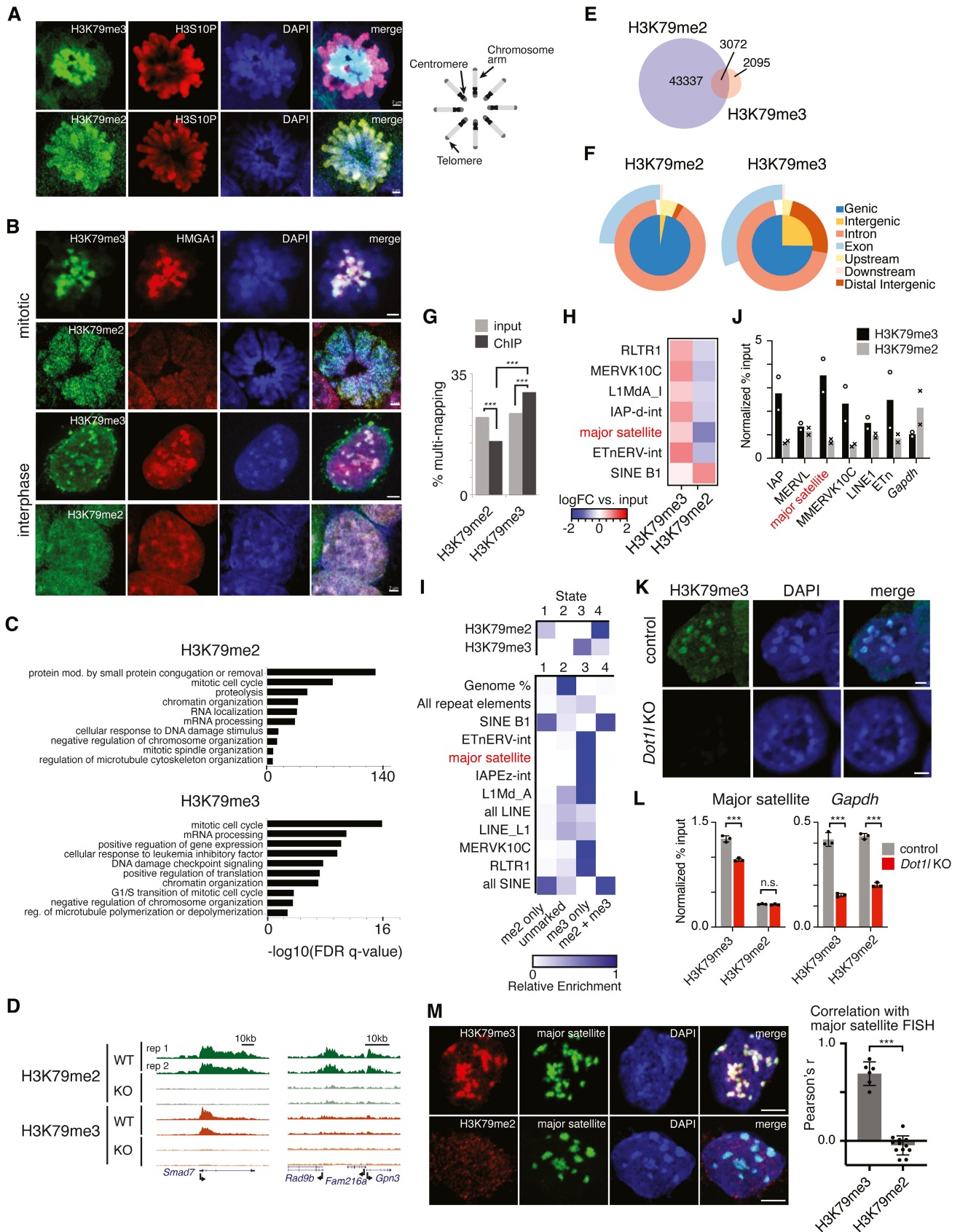
Figure 1.

Figure 1. Differential distribution of H3K79me2 and H3K79me3 on the repetitive genome in mESCs.

A   Immunofluorescence staining of single prometaphase nuclei showing H3K79me3 or H3K79me2 (green) and mitotic marker phospho-H3S10 (H3S10P, red). DNA is stained with DAPI (blue). Right, schematic of chromosomes at prometaphase showing arrangement of centromeres at the center of the metaphase plate. Scale bar, 2 μm.

B   Immunofluorescence staining of single nuclei for H3K79me3 or H3K79me2 (green) and the pericentromeric heterochromatin marker HMGA1 (red). Upper panels, prometaphase; lower panels, interphase. DNA is stained with DAPI (blue). Scale bar, 2 μm.

C   Selected Gene Ontology (GO) terms enriched among genes overlapping H3K79me2 or K3K79me3 peaks. See Dataset EV1 for complete term lists.

D   Representative ChIP-seq genome browser tracks (mm10) for H3K79me2 and H3K79me3 at a locus enriched for both marks (left) and a locus enriched for H3K79me2 only (right) in wild-type and *Dot1l* KO mESCs. Data from two biological replicates is shown.

E   Overlap between H3K79me2 and H3K79me3 consensus peaks.

F   Genomic feature distribution of H3K79me2 and H3K79me3 consensus peaks.

G   Fraction of ChIP and input libraries composed of non-uniquely mapping reads for H3K79me2 and H3K79me3 ChIP-seq data. Bars represent fraction of multi-mapping reads out of total library size after pooling reads across two replicates (Table EV1). ***$P < 10^{-15}$, two-sided test of two proportions using pooled data.

H   H3K79me2 and H3K79me3 ChIP-seq enrichment relative to input at selected repetitive element sequences.

I   ChromHMM emission states (top) obtained from H3K79me2 and H3K79me3 ChIP-seq data showing H3K79me2-only (state 1), H3K79me3-only (state 3), and H3K79me2 + H3K79me3 (state 4) states. Enrichment for selected repeat elements in regions assigned to each state is shown (bottom).

J   ChIP-qPCR for H3K79me3 and H3K79me2 at selected repeat element consensus sequences and *Gapdh* in wild-type mESCs. Data are normalized to a control intergenic region (mm10 chr19:51758385–58758494). Bars represent mean of $n = 2$ biological replicates.

K   Immunofluorescence staining for H3K79me3 in control and *Dot1l* KO nuclei. Scale bar, 2 μm.

L   ChIP-qPCR for H3K79me3 and H3K79me2 in wild-type and *Dot1l* KO cells at the major satellite consensus sequence (left panel) or *Gapdh* (right panel). Data are normalized to a control intergenic region (mm10 chr19:51758385–58758494). Bars represent mean and error bars represent standard deviation of $n = 3$ biological replicates. ***$P < 0.001$, two-tailed unpaired *t*-test.

M   DNA Fluorescence *in situ* hybridization (DNA FISH) for the mouse major satellite element with immunofluorescence staining for H3K79me3 (top) or H3K79me2 (bottom) in single mESC nuclei. Scale bar, 5 μm. Plot shows Pearson's correlation values between FISH and H3K79me2 or H3K79me3 signal, each point representing an individual nucleus. ***$P < 0.0001$, two-tailed unpaired *t*-test.

Source data are available online for this figure.

H3K79me2, $P < 10^{-5}$, two-sided test of two proportions), implying that they aligned to repetitive regions (Fig 1G). This disparity was not apparent in nonimmunoprecipitated input control libraries (mean 23.5% for H3K79me3 compared to 22.5% for H3K79me2). More formal analysis of enrichment at specific repetitive element consensus sequences using a dedicated analytic pipeline (Criscione *et al*, 2014), revealed that H3K79me3 was enriched (FDR *q*-value < 0.05) at 203 repeat elements in mESCs, whereas H3K79me2 was depleted (Figs 1H and EV2F). In contrast, H3K79me2 was enriched at SINEs, consistent with the association between SINEs and active regulatory elements (Faulkner *et al*, 2009; Tashiro *et al*, 2011; Nakajima *et al*, 2017; Policarpi *et al*, 2017). Parallel analysis of these data using a different analytic approach, chromHMM (Ernst & Kellis, 2012) identified distinct H3K79me2-only, H3K79me3-only, and H3K79me2 + H3K79me3 states across the genome and confirmed that the H3K79me3-only state was selectively enriched at repetitive elements overall including major satellite sequences, while the H3K79me2-only and H3K79me2 + H3K79me3 states were enriched at SINEs but not at other repeat elements (Figs 1I and EV2G). ChIP-qPCR at selected repeat element consensus sequences further validated stronger enrichment for H3K79me3 relative to H3K79me2 at the major satellite sequence and other repeat elements (Fig 1J).

Because immunofluorescence staining showed H3K79me3 localization to chromocenters, which are composed largely of major satellite sequence, we followed up specifically on the apparent enrichment of H3K79me3 to major satellite sequences. We confirmed that H3K79me3 at chromocenters is dependent on DOT1L by immunofluorescence in *Dot1l* KO cells (Fig 1K) and that H3K79me3, but not H3K79me2, was reduced at major satellite elements in *Dot1l* KO mESCs by ChIP-qPCR (Fig 1L). Joint immunofluorescence for H3K79me3 and DNA fluorescence *in situ* hybridization (DNA FISH) for major satellite sequences confirmed their colocalization (Fig

1M). These results suggest selective functions for H3K79me3 within the repetitive genome, including at major satellite repeats.

DOT1L is a transcriptional activator at major satellite repeats

The strong immunofluorescence-based localization of H3K79me3 to major satellites at pericentromeric regions (Fig 1A, B, L and M) and enrichment for major satellite sequences by ChIP-seq (GSAT_MM, $q = 4.52 \times 10^{-6}$) and DNA FISH prompted us to examine a potential role for DOT1L in regulation of transcription at pericentromeric heterochromatin. To see whether major satellite expression was altered in the absence of DOT1L, we performed transcriptional profiling using ribosome-depletion RNA-seq in *Dot1l* KO mESCs (Dataset EV2). Overall, DOT1L ablation had a modest effect on protein-coding gene expression. Differentially expressed genes were not strongly biased toward up- or downregulation (662 genes (58%) upregulated and 487 genes (42%) downregulated; Fig 2A and Dataset EV2). As a group, downregulated genes were enriched for H3K79me2 but not H3K79me3, (Figs 2B and EV3A), consistent with the known association between H3K79me2 and transcriptional elongation at single-copy genes (Guenther *et al*, 2008; Steger *et al*, 2008; Veloso *et al*, 2014), while upregulated genes were slightly enriched for H3K79me3 but not H3K79me2 (Figs 2B and EV3A). Most differentially expressed genes were not associated with H3K79 methylation (190/487 = 39.0% of downregulated and 56/662 = 8.5% of upregulated genes associated with any peaks), likely due to indirect effects of DOT1L loss. We did not detect any significant functional enrichment or obvious pattern in the set of downregulated genes, although upregulated genes were enriched for some pathways known to be associated with DOT1L function, such as developmental regulation and control of cell death (Fig 2C and Dataset EV2). Together, these results support a role for DOT1L and H3K79 methylation, especially H3K79me2, in the regulation of single-copy genes,

but confirm previous findings that loss of DOT1L has only modest effects on single-copy gene expresson in mESCs (Cao *et al*, 2020a).

We then examined expression of repetitive elements. Although H3K79me3 was enriched at multiple repetitive element classes (Figs 1H and I, and EV2F and G), we found that at the transcriptional level DOT1L loss primarily resulted in downregulation of satellite elements, including major satellites (Fig 2D and E). To confirm that expression of major satellites requires DOT1L, we assayed major satellite transcript levels in control and *Dot1l* KO mESCs by both RNA FISH and RT-qPCR. By RNA FISH, major satellite transcript levels were enriched at H3K79me3-positive chromocenters in control mESCs, and transcript levels were substantially reduced and mislocalized away from chromocenters in *Dot1l* KO cells (Fig 2F). Major satellite transcript levels were likewise reduced when assayed by RT-qPCR (Fig 2G). Treatment of wild-type mESCs with the DOT1L inhibitors SGC0946 (Yu *et al*, 2012) or EPZ5676 (Daigle *et al*, 2013) reduced both global H3K79 methylation and major satellite transcript levels, reinforcing the importance of DOT1L catalytic activity in this effect (Figs 2G and EV3B). The stronger effect of SGC0946 relative to either KO or EPZ5676 treatment may be a result of incomplete KO, higher potency of SGC0946 (Stauffer *et al*, 2019), additional off-target effects of SGC0946, or variability between experiments. In contrast to major satellites, expression of other classes of repetitive elements, including LTRs, LINES, and SINEs, was either unchanged or slightly upregulated in *Dot1l* KO cells (Figs 2D and EV3C). Inhibition of DOT1L also reduced expression of human pericentromeric alpha satellites in HCT116 cells, confirming that the role of DOT1L in promoting pericentromeric satellite expression is generalizable beyond mESCs (Fig EV3D). Together, these results indicate that DOT1L is selectively required for promoting expression of major satellite transcripts.

At single copy genes, DOT1L cooperates with the Super Elongation Complex (SEC) to promote transcriptional elongation. We therefore asked whether the SEC is also required for DOT1L function at major satellites. We assessed engaged RNA polymerase II (PolII) at repeat elements by reanalyzing published Precision Run-On sequencing (PRO-seq) data from wild-type, *Dot1l* KO, and catalytically inactive mutant (CI) mESCs treated with the SEC inhibitor KL2

(Cao *et al*, 2020a; Data ref: Cao *et al*, 2020b). Unexpectedly, *Dot1l* CI, *Dot1l* KO, and wild-type KL2-treated cells all exhibited increased PolII engagement at major satellites, contrasting with the reduced steady-state major satellite transcript levels in KO and inhibitor-treated cells. This effect was not enhanced by the combination of either *Dot1l* KO or CI mutants and SEC inhibitor, was stronger for CI compared to KO mutants, and was selective for major satellites compared with other repeat elements (Fig EV3E). These results suggest that DOT1L may be required for elongation, but not initiation, of transcription at major satellites: in the absence of either DOT1L or SEC, PolII may initiate partial transcripts at major satellite sequences without proceeding to elongation. We evaluated PolII enrichment at major satellites by ChIP-qPCR and found that the form of PolII associated with transcriptional elongation (PolII-S2P, phosphorylated at Serine 2 of the C-terminal domain) was depleted at major satellites in *Dot1l* KO mESCs, consistent with an elongation defect (Fig EV3F). In contrast, the initiation-associated form PolII-S5P was neither reduced nor elevated (Fig EV3G). Loss of DOT1L therefore has a stronger effect on PolII elongation than initiation, although it is surprising that PolII-S5P ChIP-qPCR signal is not enriched similar to PRO-seq signal. Transcription can initiate at any point in the concatenated major satellite repeat rather than at a discrete transcription start site (Rudert *et al*, 1995; Lu & Gilbert, 2007; Valgardsdottir *et al*, 2008), meaning that a distributed increase in PolII-S5P signal may not be detectable by qPCR. Together, these results indicate that DOT1L promotes expression at major satellites, possibly by promoting PolII elongation.

Association of DOT1L with SMARCA5 and heterochromatin factors facilitates its role in major satellite transcription at PCH

Major satellite transcripts derived from pericentromeric repeats play an essential role in the *de novo* reestablishment of heterochromatin after each mitotic division (Nair *et al*, 2020). In mice, major satellite transcripts themselves appear to recruit the H3K9 methyltransferase SUV39H1 to facilitate H3K9me3 deposition, and DOT1L has been shown to regulate deposition of heterochromatin marks at centromeres and telomeres (Jones *et al*, 2008; Johnson *et al*, 2017; Shirai

Figure 2. DOT1L promotes transcription at major satellites in mESCs.

A Volcano plot of differentially expressed single-copy genes in *Dot1l* KO mESCs. Each point represents one gene. Differentially expressed genes (log2 fold change ≥ 1, adjusted *P*-value ≤ 0.05) are shown in red. *P*-values were corrected for multiple hypothesis testing using the Benjamini–Hochberg method.

B Metagene plots showing mean ChIP signal for H3K79me2 and H3K79me3 at downregulated and upregulated genes identified in (A).

C Selected GO terms enriched among significantly upregulated single-copy genes in *Dot1l* KO mESCs. *P*-values were adjusted for multiple hypothesis testing using the Benjamini–Hochberg method. See Dataset EV2 for complete term lists.

D Differential expression of repeat element classes in *Dot1l* KO relative to control mESCs. Red dot indicates relative expression of major satellite transcripts within the broader set of annotated satellite sequences. Boxes represent interquartile range (25th–75th percentile) of uniquely annotated repeat classes within each broad category; center band represents the median; whiskers represent maximum and minimum values. Total number of repeat element annotations included in each class is shown in parentheses.

E Strip plot showing relative expression of selected specific repeat element annotations in *Dot1l* KO mESCs compared to control.

F RNA FISH for major satellite transcripts and simultaneous staining for H3K79me3 in control and *Dot1l* KO mESCs. Scale bar, 5 μm. Below left, quantitation of fluorescence intensity along a line bisecting a representative nucleus. Bisecting lines used to generate the plots are shown in the merged images. Below right, quantitation of total FISH fluorescence signal for *n* = 46 cells of each genotype. Line summaries show mean ± standard deviation. ***P* < 0.0001, unpaired Welch's *t*-test.

G Quantitative real-time PCR (RT-qPCR) of major satellite transcripts in *Dot1l* KO mESCs and wild-type mESCs treated with DOT1L inhibitors SGC0946 or EPZ5676 relative to untreated wild-type mESCs (dashed line). Bars represent mean of *n* = 4 biological replicates for control and *Dot1l* KO and *n* = 2 biological replicates for inhibitor-treated cells. Control region indicates an intergenic locus (mm10 chr19:51758385–58758494) not strongly associated with either active or inactive chromatin. Values were normalized to *Gapdh* as an internal control. Bars represent mean and error bars represent ± SEM. N.d., not done. ***P* < 0.001, two-tailed unpaired *t*-test.
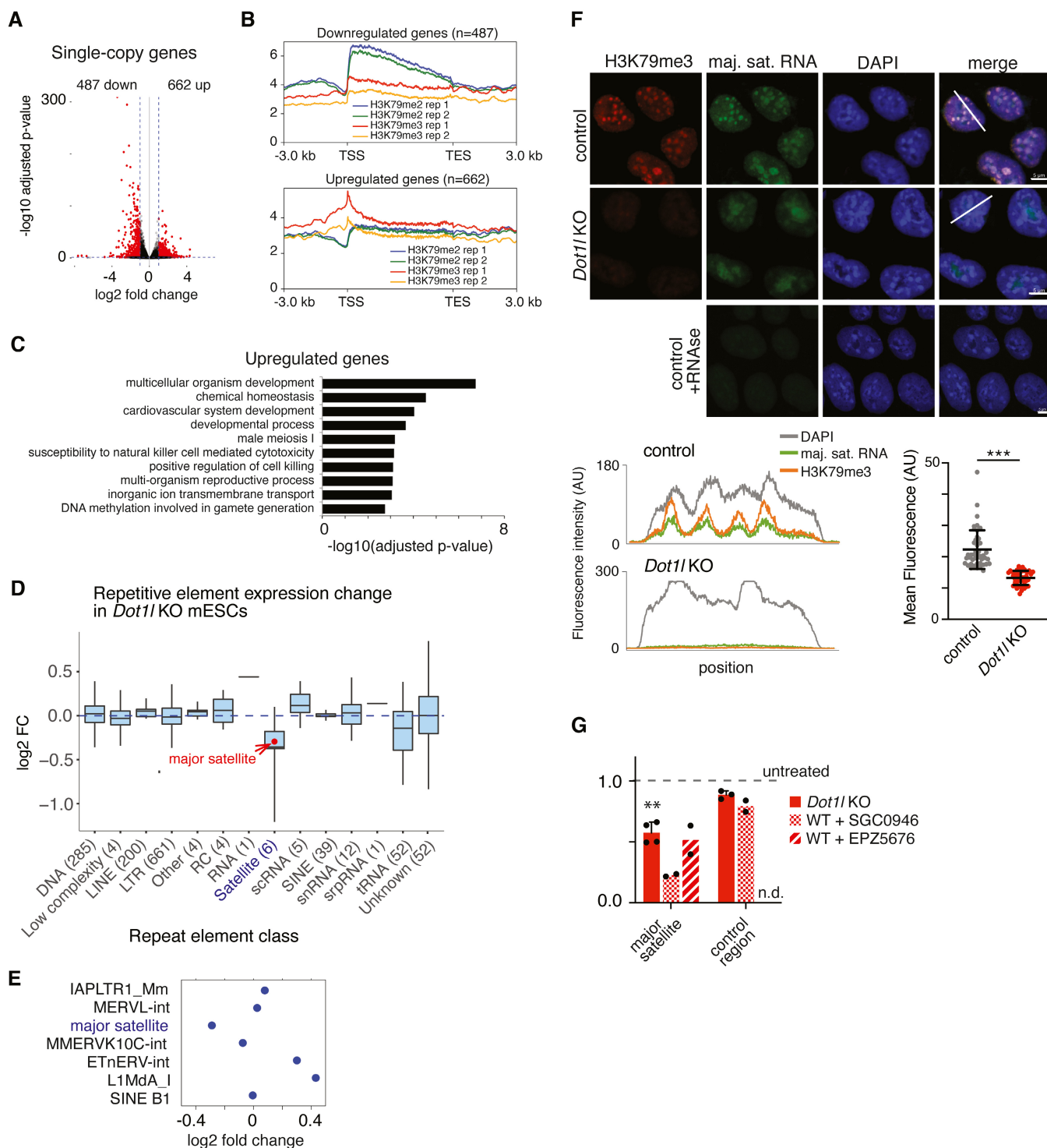
Source data are available online for this figure.

Figure 2.

major satellites in both assays (Fig 3A and B), as was chromocenter number (Fig 3C). The heterochromatin-associated histone modification H3K9me2 was also reduced at major satellites in Dot1l KO cells (Fig 3D). Consistent with previous reports, we did not detect a reduction in H3K9me3 by ChIP-qPCR, although this modification is moderately reduced at the global level in Dot1l KO mESCs (Fig
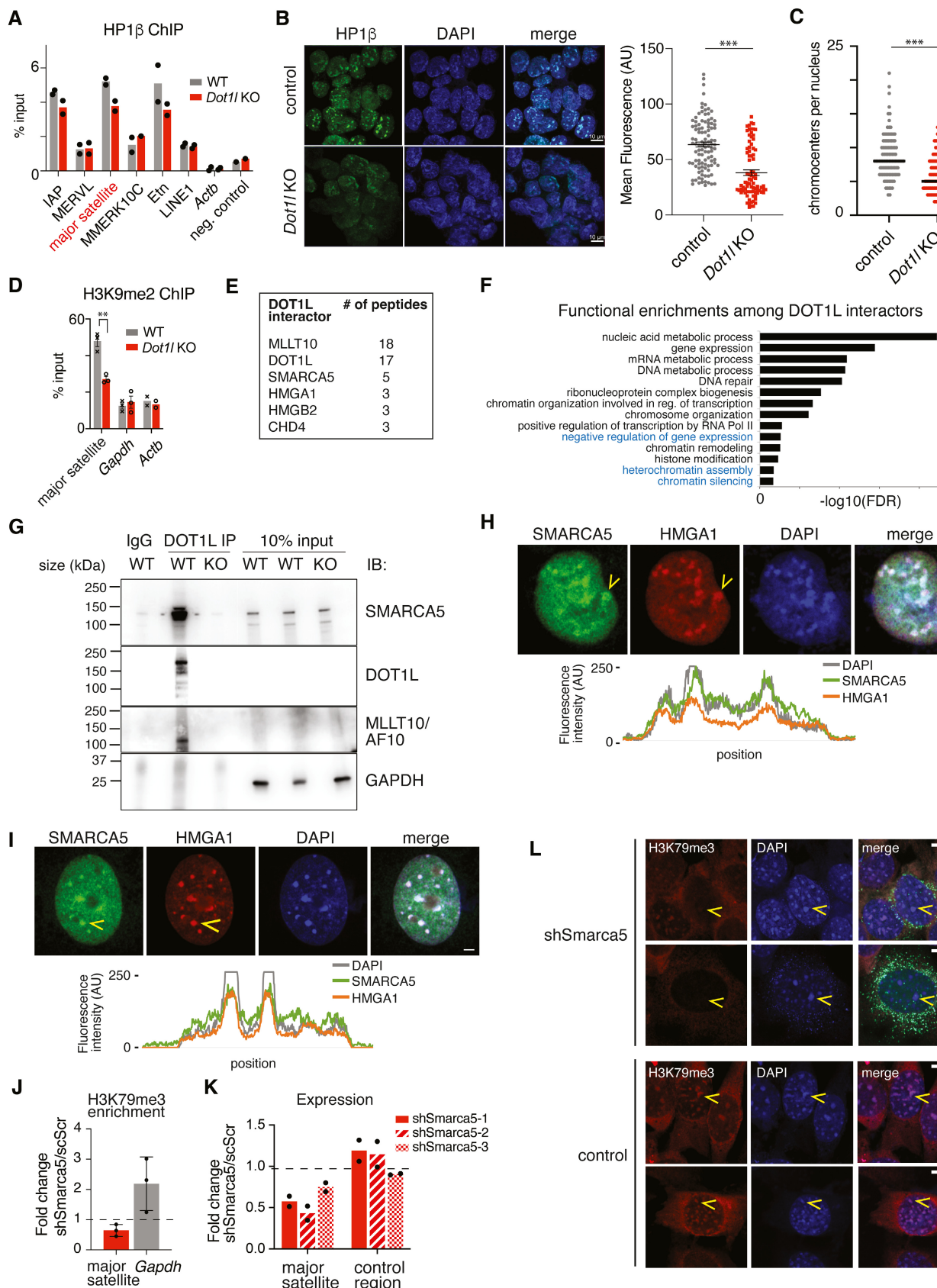
Figure 3.

◄

EV4A and B). These results suggest that DOT1L promotes major satellite transcription and also contributes to maintenance of heterochromatin at pericentromeres.

The complex interaction between DOT1L, transcriptional activation, and heterochromatin reflects a longstanding challenge in understanding the molecular regulation of PCH. PCH fluctuates between transcriptionally active and heavily repressed states depending on the biological context, making it difficult to define straightforward regulatory pathways using genetic knockouts. To better define the function of DOT1L at PCH, we next sought to discover the protein factors that partner with DOT1L at these regions. We overexpressed HA-tagged full-length DOT1L in mESCs and performed immunoprecipitation followed by mass spectrometry (IP-MS). A total of 70 protein interactors were identified in transfected compared to nontransfected cells (Dataset EV3). As expected, the top hit was MLLT10 (AF10), a transcriptional activator and known DOT1L interactor (Fig 3E; Okada *et al*, 2005; Bitoun *et al*, 2007; Mohan *et al*, 2010). Overall, the list of interactors was enriched for pathways known to be associated with DOT1L function, including gene expression, DNA repair, chromatin remodeling, and histone modification (Fig 3F and Dataset EV3). There was also significant enrichment for pathways related to transcriptional repression, such as negative regulation of gene expression, hetero-chromatin assembly, and chromatin silencing. Several of these interactors are known regulators of heterochromatin but have not previously been shown to interact with DOT1L. One such protein was HMGA1, a marker for PCH (Jagannathan *et al*, 2018), supporting the model that DOT1L is recruited to PCH to mediate major satellite transcription. Another intriguing interactor was SMARCA5 (SNF2H). SMARCA5 is the mammalian homolog of ISWI, an ATP-dependent chromatin remodeler that promotes

transcription by disrupting DNA-histone contacts to slide or evict histones (Aihara *et al*, 1998; Stopka *et al*, 2000; Clapier & Cairns, 2009), and has been reported to localize to PCH (Vargova *et al*, 2009). We validated the interactions between endogenous DOT1L and AF10, SMARCA5, HMGA1, and CHD4, another heterochromatin-associated factor, in wild-type mESCs by co-immunoprecipitation, and confirmed that co-precipitations were abrogated in *Dot1l* KO mESCs (Figs 3G and EV4C). We examined localization of SMARCA5 to PCH by immunofluorescence and found that it co-stains with HMGA1 at DAPI-dense chromocenters in both mESCs (Fig 3H). In mouse 3T3 fibroblasts, DAPI-dense foci similarly stained for H3K9me3, H3K79me3, HMGA1, and SMARCA5, further supporting preferential localization of SMARCA5 to chromocenters (Figs 3I and EV4D).

We then asked whether SMARCA5 might contribute to the function of DOT1L in promoting major satellite expression. We gener-ated mESC cell lines stably expressing *Smarca5* shRNAs and recovered multiple *Smarca5* KD clones, with levels of SMARCA5 protein ranging from 1 to 40% of control (Fig EV4E). H3K79me3 sig-nal was reduced at major satellites in *Smarca5* KD mESCs as assayed by ChIP-qPCR (Fig 3J) and major satellite transcript levels were reduced by 50% (Fig 3K). Similarly, knockdown of SMARCA5 in 3T3 fibroblasts led to reduced H3K79me3 accumulation at chro-mocenters by immunofluorescence (Figs 3L and EV4F). In mESCs, knockdown of SMARCA5 led to a significant reduction in DOT1L protein levels by immunofluorescence, but had little effect on DOT1L localization (Fig EV4G). SMARCA5 levels were also mildly reduced in *Dot1l* KO mESCs by immunofluorescence with no nota-ble change in SMARCA5 localization (Fig EV4H). These data indi-cate that DOT1L and SMARCA5 influence each others' expression either directly or indirectly and have related roles in promoting
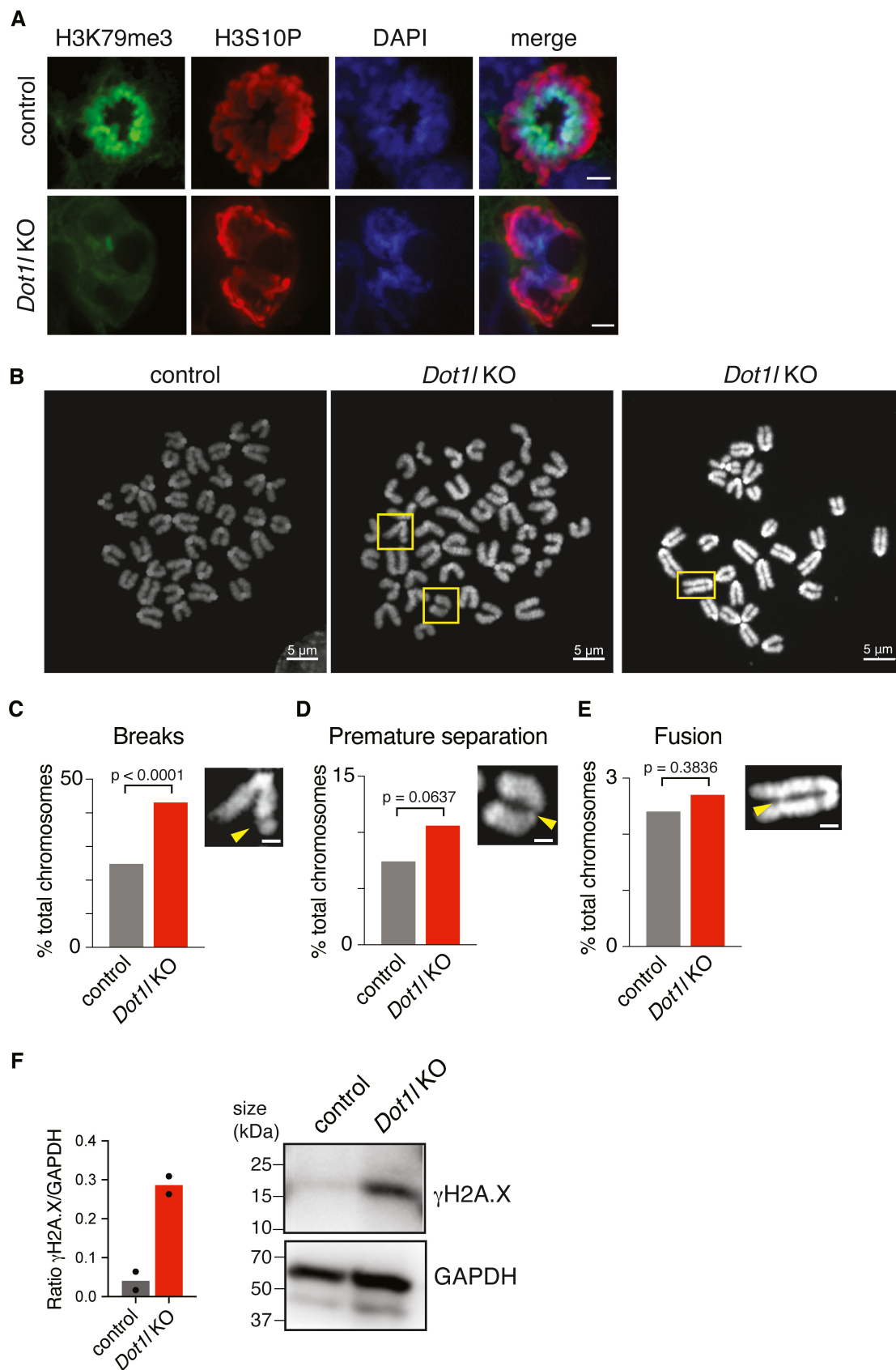
Figure 4.

Figure 4. Loss of DOT1L causes mitotic defects and chromosomal abnormalities in mESCs.

A   Immunofluorescence staining of H3K79me3 (green) and H3S10P (red) in a single nucleus of control and *Dot1l* KO mESCs. DNA is stained with DAPI (blue). Scale bar, 5 μm.

B   Example metaphase spreads in wild-type and *Dot1l* KO mESCs showing examples of a chromosome break, premature centromere separation, and chromosome fusion (yellow boxes). Scale bar, 5 μm.

C–E   Quantitation of chromosome defects in *Dot1l* KO mESCs, including chromosomal breaks and fragments (C), premature centromere separation (D), and chromosome fusions (E). Insets show representative examples of each chromosome defect, highlighted by an arrowhead; examples in (C) through (E) correspond to the boxes in (B). Scale bars, 1 μm. $N = 201$ (wild-type) or $n = 445$ (KO) chromosomes from $n = 46$ (wild-type) or $n = 65$ (KO) individual nuclei in two independent experiments; chromosomes considered ambiguous were excluded. *P*-values were calculated using Fisher's Exact test.

F   Western blot for the DNA damage marker γH2A.X in control and *Dot1l* KO mESCs. Quantitation of signal on the blot for $n = 2$ biological replicates is shown.

Source data are available online for this figure.

accumulation of H3K79me3 and transcription at major satellite sequences, but that an additional unknown factor is responsible for their recruitment to chromocenters.

Loss of DOT1L leads to chromosomal abnormalities in mESCs

Pericentromeric heterochromatin stabilizes the centromeric core during mitosis (Yi *et al*, 2018), and transcription of major satellites has been proposed to reinforce heterochromatinization at PCH during anaphase, suggesting that it is required for mitotic progression (Lu & Gilbert, 2007; Saksouk *et al*, 2015). Since we found that DOT1L promotes major satellite expression, we predicted that DOT1L loss would result in defects in chromosome segregation during mitotic division. From our immunofluorescence data (Fig 1A), we observed that a fraction of Dot1l KO cells exhibited aberrant chromosome alignment at the metaphase plate resulting in disorganized chromosome congression (Fig 4A). To further characterize this defect, we examined metaphase chromosome spreads in *Dot1l* KO and control mESCs, and found that *Dot1l* KO cells had significantly higher levels of chromosome breakage (Fig 4B and C) as well as a trend toward higher frequency of premature centromere separation (Fig 4D). We did not observe a significant increase in chromosome fusions in *Dot1l* KO cells (Fig 4E). Consistent with increased levels of chromosome breakage, we found that levels of the DNA damage marker γH2A.X were elevated in *Dot1l* KO mESCs (Fig 4F). Collectively, these results demonstrate that DOT1L loss predisposes cells to chromosomal rearrangements and genomic instability, as predicted when major satellite transcription is impaired. Chromosome damage may contribute to the cell

proliferation defects previously reported in mESCs lacking DOT1L (Jones *et al*, 2008).

DOT1L activity is required for cell cycle progression and viability in mouse preimplantation embryos

We next asked whether the loss of DOT1L activity at PCH has *in vivo* significance. Major satellite expression during early cleavage stages in mouse preimplantation embryos is required for heterochromatin establishment and formation of chromocenters (Probst *et al*, 2010; Almouzni & Probst, 2011). We therefore asked whether DOT1L is required for major satellite expression in early embryogenesis. First, we characterized the dynamics of H3K79me2 and H3K79me3 in preimplantation mouse embryos by immunostaining. For comparison, we co-stained with H3K9me3, which is present specifically in the maternal pronucleus in the zygote and localizes to heterochromatin as it forms in the two-cell stage (Liu *et al*, 2004; Burton & Torres-Padilla, 2010). H3K79me2 was absent from the zygote through at least the eight-cell stage (Fig EV5A). In contrast, H3K79me3 was present at low levels in the early female pronucleus at 4-h postfertilization (hpf), depleted in the late zygote by 10 hpf, and reestablished at chromocenters at the four-cell stage (48 hpf; Figs 5A, and EV5B and C). H3K79me3 was not observed in 8-cell embryos, but reappeared at chromocenters at blastocyst stage, consistent with our observations in mESCs (Fig EV5D). These results suggest that H3K79me2 and H3K79me3 are dynamically and differentially regulated during early stages of preimplantation development, and that DOT1L methyltransferase activity is present at early chromocenters by the four-cell stage.

Figure 5. Dot1l activity is essential for cell cycle progression and viability during mouse preimplantation development.

A   Immunofluorescence staining of H3K79me3 (green) and H3K9me3 (red) in wild-type mouse embryos during the first 48 h of preimplantation development. Bottom images show an embryo treated with DOT1L inhibitor at 48 h post fertilization (hpf). DNA is stained with DAPI (blue). Scale bars, 10 μm (main image) or 4 μm (inset). Quantitation of pixel-by-pixel fluorescence intensity for H3K9me3 and H3K79me3 for a representative embryo is shown at right. AU, arbitrary units.

B   Quantitation of numbers of total, H3K9me3-positive, and H3K79me3-positive chromocenters in control and inhibitor-treated embryos at 48 hpf. Chromocenters were defined as DAPI-dense foci. Each point represents one nucleus. $N = 69$ (control) or $n = 20$ nuclei (inhibitor treated) in $n = 19$ (control) or $n = 5$ (inhibitor treated) embryos *$P < 0.05$, ***$P < 0.001$, two-tailed unpaired *t*-test.

C   Quantification of the fraction of fertilized zygotes that progress to a given stage at each time point in control (vehicle) and inhibitor (SGC0946) treated embryos. Numbers of embryos that successfully progressed to a given stage were counted at each time point and the fraction of total is shown. Points represent the mean of $n = 52$ (control) or $n = 59$ (inhibitor treated) embryos across 5–6 independent experiments and error bars represent standard deviation. **$q < 0.01$, ***$q < 0.001$; false discovery rate (FDR) *q*-values calculated by unpaired *t*-test with two-state step-up correction for multiple comparisons.

D   Quantification of cell numbers per embryo at indicated time points in control (DMSO) and inhibitor (SGC0946) treated embryos. Numbers of nuclei per embryo were counted at each time point for $n = (14, 20, 26, \text{and } 12)$ or $n = (14, 4, 19, \text{and } 23)$ embryos at time $= (24, 48, 56, \text{and } 72 \text{ hpf})$ for control and inhibitor-treated embryos, respectively across three independent experiments. Bars represent mean and error bars represent standard deviation. **$q < 0.01$; false discovery rate (FDR) *q*-values calculated by unpaired t-test with two-state step-up correction for multiple comparisons. n.s., not significant at a threshold of $q < 0.05$.
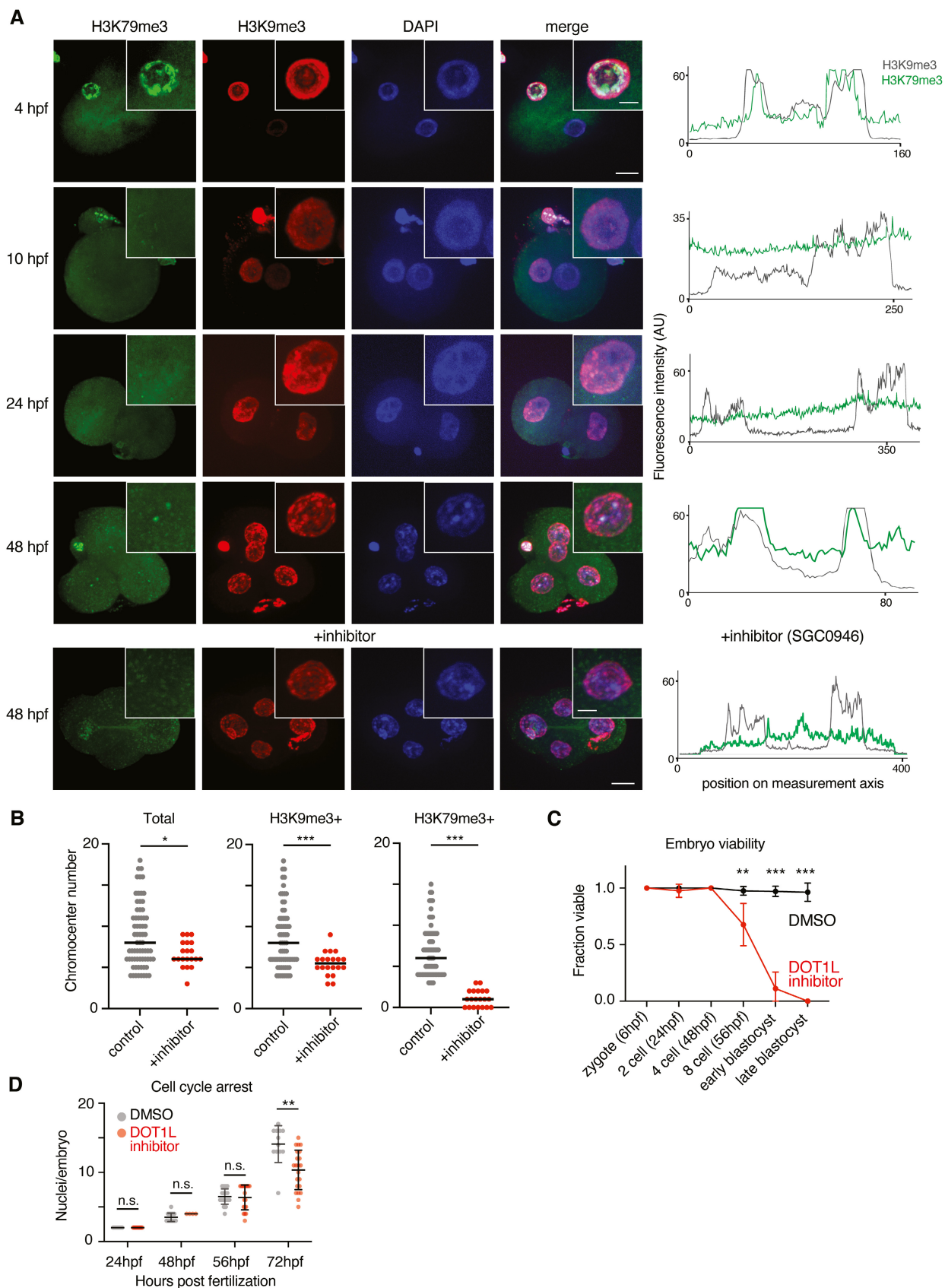
Source data are available online for this figure.

Figure 5.

We next investigated whether DOT1L activity was required for preimplantation development by treating embryos *in vitro* with the DOT1L inhibitor SGC0946. Fertilized zygotes were cultured in 10 μM inhibitor and examined for up to 96 h to evaluate the effect of DOT1L inhibition on embryonic progression to blastocyst. Immunostaining at intermediate time points confirmed that H3K79me3 was absent or substantially reduced in inhibitor-treated embryos (Fig 5A), although some embryos displayed residual staining that could reflect incomplete inhibition. In addition to loss of H3K79me3 from chromocenters at 48 hpf, inhibitor treatment also resulted in reduced numbers of chromocenters at 48 hpf (Fig 5B), implying that chromocenter stabilization is impaired in the absence of DOT1L activity. DOT1L inhibition resulted in significant embryo death starting at 56 hpf, with no embryos progressing past the early morula stage or to fully cavitated blastocysts (Figs 5C and EV5E). A similar defect beginning at 56 hpf was seen with a second DOT1L inhibitor, EPZ5676, although the magnitude of the effect was lower (Fig EV5F). While morphologically indistinguishable from controls in early cleavage, by 72 hpf inhibitor-treated embryos failed to establish a cohesive and compact morphology typical of morula-stage embryos, suggesting pervasive developmental defects had accumulated prior to this initial differentiation event (Fig EV5G).

To better understand the reason for the embryonic lethality, we evaluated progression through early cell cycle divisions by counting the number of nuclei per embryo at each stage. Normal embryo progression through the third cleavage (8-cell stage) was followed by a significant decrease in cell number in inhibitor-treated embryos by 72 hpf, accompanied by elevated numbers of mitotic figures (Fig 5D). We also observed lagging chromosomes in anaphase cells of inhibitor-treated embryos beginning at the eight-cell stage, although there was insufficient data to quantify this effect (Fig EV5H). Mitotic dysfunction and progressive cell cycle arrest therefore begin soon after the induction of major satellite transcription during late cleavage. Together with immunostaining data indicating that DOT1L enzymatic activity begins by 48 hpf (Fig 5A), these results support a requirement for DOT1L in cell cycle progression beginning at the four- to eight-cell stage of preimplantation development.

DOT1L activity in preimplantation embryos operates through major satellite transcription

A burst of major satellite transcription occurs at the two-cell stage, preceding the effects of DOT1L inhibition in early mouse embryos (Burton & Torres-Padilla, 2014). Given our finding that DOT1L promotes major satellite transcription at PCH in mESCs (Fig 2), we asked whether DOT1L inhibition impairs the dynamics of major satellite transcription in preimplantation embryos. We performed ribosome depletion RNA-seq in eight-cell stage (56 hpf) embryos cultured in the presence or absence of DOT1L inhibitor and assessed transcription of single-copy genes and repetitive elements (Table EV1). We detected limited changes in expression of euchromatic single-copy genes following DOT1L inhibition: In total, 125 genes exhibited low-magnitude but statistically significant changes in expression in inhibitor-treated embryos (adjusted *P*-value < 0.05; Fig 6A and Dataset EV4), with a slight bias toward decreased expression

as we saw in *Dot1l* KO mESCs (Fig 6A). Downregulated genes were significantly enriched for transcriptional and developmental functions, while upregulated genes were associated with cell cycle arrest and activation of apoptotic pathways that could reflect emerging secondary effects (Fig 6B and Dataset EV4).

Among repetitive elements, we detected significant changes exclusively for major satellite transcripts (Fig 6C and D). Major satellite transcripts were significantly downregulated following drug treatment (log2 fold change −1.50, false discovery rate *q*-value 0.000450), matching localization of H3K79me3 to chromocenters in control embryos (Fig 5A) as well as the onset of cell cycle defects observed in inhibitor-treated embryos (Fig 5C). Notably, major satellite transcripts were more strongly downregulated than all but one single-copy gene (*Slc30a2*, log2 fold change −1.51, adjusted *P*-value 0.00153). These results suggest that major satellite transcription is stimulated by DOT1L at this time point and represents a critical moment in PCH stabilization immediately following its establishment in early embryogenesis, gating subsequent cell cycle progression and embryo viability.

Discussion

We uncover a critical role for DOT1L in activation of major satellite expression at pericentromeric heterochromatin, and find that this function is important for stabilization of early heterochromatin structures in preimplantation embryos and for heterochromatin maintenance and cell cycle progression in ESCs. While DOT1L is well known to be involved in transcriptional elongation at single-copy genes and has previously been shown to regulate heterochromatin, selective transcriptional activation of major satellite repeats by DOT1L in PCH has not previously been reported and represents a novel regulatory function that unites its roles in transcriptional activation and heterochromatin maintenance.

DOT1L is an evolutionarily ancient and unique histone methyltransferase with no cognate demethylase, making it the sole known regulator of H3K79 methylation. We find that the DOT1L substrates H3K79me2 and H3K79me3 are differentially distributed within the repetitive genome, with H3K79me3 preferentially enriched at major satellites, raising the question of how these two modifications are differentially directed by a single enzyme. At single-copy loci, increased valency of H3K79 methylation correlates with reduced nucleosome turnover, suggesting a model where increased H3K79me3 at PCH is secondary to reduced nucleosome turnover in heterochromatin (Chory *et al*, 2019). On the contrary, sequential conversion of H3K79me1 to H3K79me2 and H3K79me3 also correlates with enhanced rates of transcription and increased turnover (Steger *et al*, 2008). The level of H3K79me3 enrichment at PCH is therefore likely to be a function of these two opposing processes. We propose that DOT1L participates in two distinct functional complexes: one that supports transcriptional elongation at single-copy genes and preferentially deposits H3K79me2 (Mohan *et al*, 2010), and a second, previously unknown complex that may include the nucleosome remodeler SMARCA5, operates to open chromatin and drive transcription at PCH in a highly controlled fashion, and ultimately promotes complete
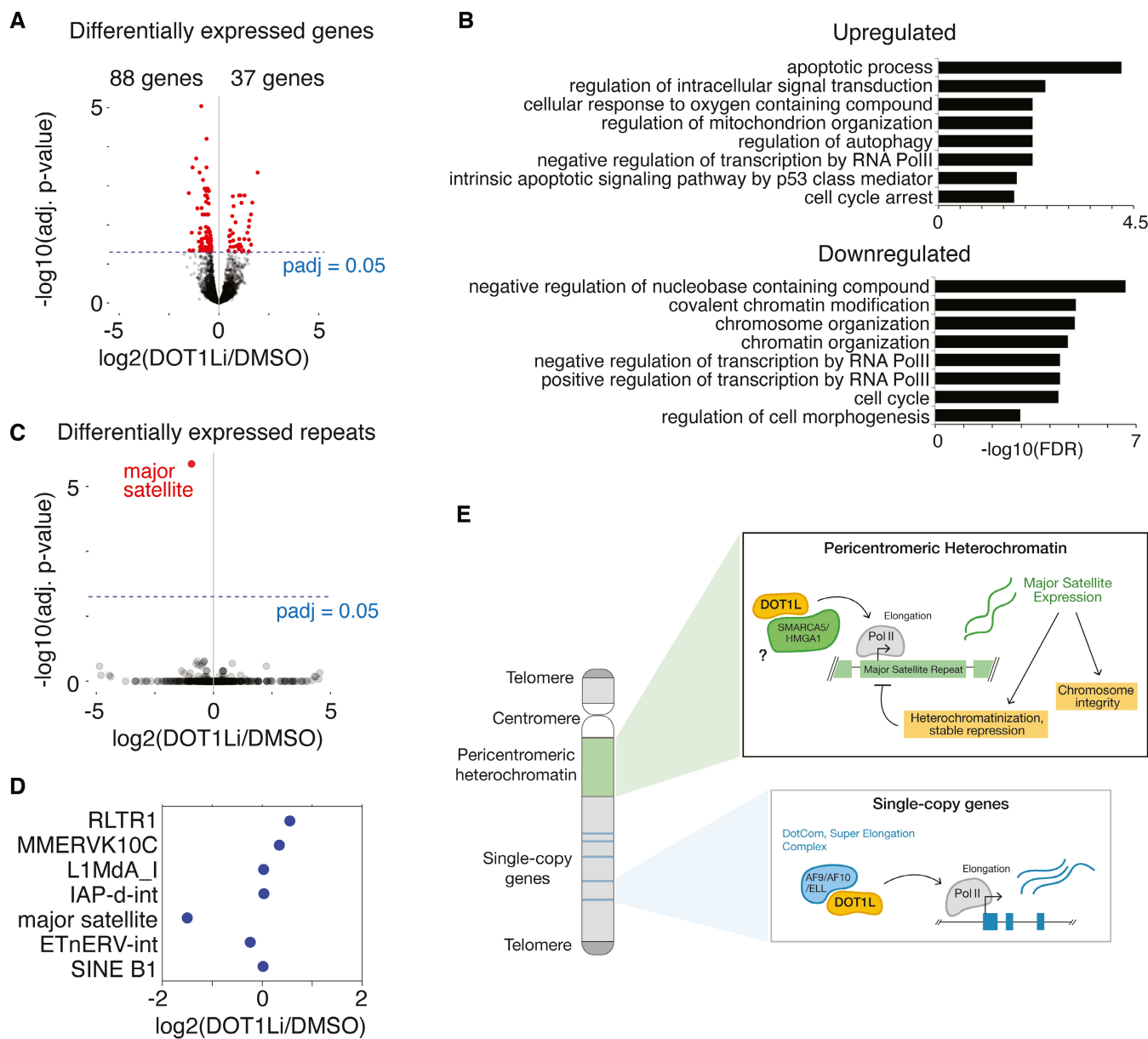
Figure 6. DOT1L promotes major satellite transcription in preimplantation mouse embryos.

A Volcano plot showing differentially expressed single-copy genes in DOT1L inhibitor-treated vs control embryos at 56 hpf (8 cell stage). Each dot signifies one gene. Differentially expressed genes (adjusted *P*-value ≤ 0.05; no threshold for log2 fold change) are shown in red. *P*-values were corrected for multiple hypothesis testing using the Benjamini–Hochberg method.

B Selected enriched GO terms among significantly upregulated and downregulated single-copy genes in inhibitor-treated vs. vehicle-treated embryos.

C Volcano plot of annotated repeat elements in DOT1L inhibitor-treated vs. vehicle-treated eight-cell embryos (56 hpf). Each dot signifies one annotated repetitive element category as indicated by RepeatMasker (Smit *et al*). *P*-values were corrected for multiple hypothesis testing using the Benjamini–Hochberg method. Differentially expressed repetitive elements (adjusted *P*-value ≤ 0.05) are shown in red. The major satellite element is indicated.

D Strip plot showing log2 fold change of selected repeat elements in eight-cell embryos as shown in (C).

E Model for DOT1L function at PCH and euchromatic single-copy genes. DOT1L coordinates with SMARCA5 and other heterochromatin-associated factors for its recruitment to PCH to promote major satellite transcription, but interacts with elongation factors including AF9, AF10, and ELL at transcribed single copy genes.

Source data are available online for this figure.

methylation to H3K79me3 (Fig 6E). DOT1L was recently reported to interact with other ATP-dependent chromatin remodelers, including SMARCA4, SMARCD2, SMARCC2, in K562 cells (Wu *et al*, 2021), and levels of H3K79 methylation were reduced in spermatogenic cells of mice heterozygous for a *Smarca5* knockout allele (Vargova *et al*, 2009), further

supporting a functional interaction between DOT1L and SMARCA5.

A recent analysis of *Dot1l* KO mESCs found minimal impact on expression and no change in PolII-mediated elongation at single-copy genes, concluding that the catalytic activity of DOT1L is dispensable for its function in transcription (Cao *et al*, 2020a). We identify an alternative methyltransferase-dependent function for DOT1L in promoting expression of major satellites, implying a stronger contribution of DOT1L catalytic activity to transcriptional activity at PCH compared to euchromatic genes. Our work also highlights DOT1L as one of the first epigenetic regulators associated with transcriptional activation that stimulates proper PCH initiation and maintenance. We find that DOT1L is important for major satellite transcription as well as retention of heterochromatin factors at pericentromeric regions, adding to a growing body of evidence that major satellite transcripts are important for recruitment and retention of heterochromatin factors at PCH (Probst *et al*, 2010; Santenard *et al*, 2010; Bulut-Karslioglu *et al*, 2012; Millanes-Romero *et al*, 2013; Johnson *et al*, 2017; Shirai *et al*, 2017; Velazquez Camacho *et al*, 2017; Burton *et al*, 2020). Notably, many of the phenotypes we observe following DOT1L loss or inhibition appear to phenocopy the loss of canonical repressive factors associated with PCH and other repetitive heterochromatic elements (Peters *et al*, 2001; Abe *et al*, 2016).

We also identified DOT1L as a critical regulator of early development, where inhibition of DOT1L activity caused embryo arrest and death at preimplantation stages. In embryos, we found that H3K79me3 is enriched in chromocenters at the four-cell stage, when somatic heterochromatin structures, including chromocenters, first become prominent and immediately following the interval when major satellite expression is upregulated and chromocenters begin to form at the two-cell stage (Almouzni & Probst, 2011). We also found that DOT1L promotes major satellite transcription during early preimplantation development; the offset between initial upregulation of major satellite transcription (Probst *et al*, 2010; Casanova *et al*, 2013; Burton & Torres-Padilla, 2014) and the appearance of H3K79me3 at the four-cell stage may indicate a preferential role for DOT1L in reinforcement compared to initiation of major satellite expression. The specificity of the H3K79me3 signal is supported by its loss following inhibitor treatment; while H3K79me3 was previously reported to be absent between the two-cell and blastocyst stages, this prior study did not report knockout or inhibitor-treated controls (Ooga *et al*, 2008).

The additional delay between the first appearance of H3K79me3 at chromocenters at the four-cell stage and the phenotype we observe after the third cleavage is consistent with a mitotic defect that emerges from failed interphase regulation beginning at the four-cell stage. Pericentromeric transcription occurs at the end of the G1 phase of the cell cycle (Lu & Gilbert, 2007), but the phenotype is delayed to the next mitosis. Major satellite transcripts and DOT1L are both independently known to regulate mitotic progression (Kim *et al*, 2014; Biscotti *et al*, 2015; Müller & Almouzni, 2017). Pericentromeric heterochromatin stabilizes the centromeric core during mitosis (Yi *et al*, 2018), and transcriptional activity at major satellites has been proposed to reinforce heterochromatinization at PCH (Frescas *et al*, 2008) and

stabilize mitotic progression (Lu & Gilbert, 2007; Saksouk *et al*, 2015). The increased frequency of mitotic defects we observed in *Dot1l* KO mESCs suggests that DOT1L may promote mitotic progression via upregulation of major satellite transcripts. If so, the severe lethality phenotype in embryos following DOT1L inhibition could be the cumulative effect of reduced major satellite transcription on chromosome condensation, alignment, and segregation.

The cell cycle arrest and embryo lethality along with the reduced number of chromocenters we observed following inhibitor treatment support a critical role for DOT1L in chromocenter establishment following fertilization and complement past work in which overexpression of truncated DOT1L in newly fertilized zygotes resulted in precocious chromocenter formation and developmental arrest (Ooga *et al*, 2013). In contrast, there is a discrepancy between the phenotype of inhibitor-treated embryos and that of *Dot1l* zygotic null embryos, which survive until E10.5 (Jones *et al*, 2008). It is possible that maternal DOT1L protein from the oocyte is sufficient to establish PCH during preimplantation development and that this signature is maintained epigenetically in knockout embryos. In our experiments, maternal DOT1L would also be inhibited, effectively mirroring a combined maternal and zygotic disruption. Our RNA-seq data in DOT1L embryos was generated from F1 offspring of polymorphic mouse strains (see Materials and Methods), allowing us to distinguish paternal from maternal transcripts, and we found no bias in parent of origin for *Dot1l* transcripts; however, maternally inherited protein may still contribute to preimplantation epigenetic state. Recently, a maternal *Dot1l* conditional knockout was found to be dispensable for mouse development (Liao & Szabó, 2020), but a combined maternal-zygotic knockout has not been evaluated. Future experiments using maternal-zygotic *Dot1l* knockouts or acute antibody-mediated depletion of DOT1L (Clift *et al*, 2017) will be important to further understand this effect.

In addition to early embryogenesis, the role of DOT1L in major satellite expression may have *in vivo* significance in cancer. Several cancers are associated with DOT1L upregulation, including leukemia, lung adenocarcinoma, and colorectal cancer, and are also prone to overexpression of major satellite transcripts (Ting *et al*, 2011; Bersani *et al*, 2015). Heightened major satellite transcription due to DOT1L upregulation may contribute to tumor severity or progression in these cancers. Notably, SMARCA5 has also been implicated in pathogenesis of acute myeloid leukemia (AML), and like DOT1L is required for hematopoiesis (Stopka & Skoultchi, 2003; Kokavec *et al*, 2017; Zikmund *et al*, 2020). Future work will be important to better understand the potential role of DOT1L-mediated regulation of major satellite transcription in cancer.

In summary, our study defines a role for the H3K79 methyltransferase DOT1L in major satellite transcription and PCH regulation, illuminating a new mechanism for transient transcriptional activation in heterochromatin and uncovering the events surrounding first establishment of heterochromatin in the early embryo. Future studies investigating this function will be important for understanding the fundamental molecular mechanisms controlling heterochromatin dynamics across the cell cycle and throughout development.

Materials and Methods

Reagents and Tools table

Reagent/resource	Source	Identifier	Usage notes
Antibodies			
H3K79me1	Abcam	ab2886 RRID:AB_303383	WB 1:1,000
H3K79me2	Abcam	ab3594 RRID: AB_303937	WB 1:1,000 IF 1:500 ChIP 1 μg/10^6 cells
H3K79me3	Abcam	ab2621 RRID:AB_303215	WB 1:1,000 IF 1:500 ChIP 1 μg/10^6 cells
DOT1L	Abcam	ab64077 RRID:AB_2095412	WB 1:400 IF 1:250
H3S10P	Abcam	ab14955 RRID:AB_443110	IF 1:500
AF10	Abcam	ab208016	WB 1:1,000
CHD4	Abcam	ab70469 RRID:AB_2229454	WB 1:500
HMGA1	Santa Cruz Biotechnologies	sc-393213, clone D-12	IF 1:200
SMARCA5	Thermo Scientific	PA5-52601 RRID:AB_2647608	WB 1:1,000 IF 1:500
RNA Polymerase II S2P	Abcam	ab5095 RRID:AB_304749	ChIP 1 μg/10^6 cells
RNA Polymerase II S5P	Cell Signaling Technologies	13523 RRID:AB_2798246	ChIP 1 μg/10^6 cells
HP1β	Cell Signaling Technologies	2613 RRID:AB_561072	IF 1:400 ChIP 2 μg/10^6 cells
H3K9me3	Novus Biologicals	NBP1-30141 RRID:AB_1987267	IF 1:500
H3K9me2	Cell Signaling Technologies	4658 RRID:AB_10544405	ChIP 2 μg/10^6 cells
γH2A.X	Abcam	ab2893 RRID:AB_303388	WB 1:1,000
GAPDH	Santa Cruz	sc32233 RRID:AB_627679	WB 1:200
Oligonucleotides			
PNA major satellite probe	PNABio	Alexa488-o-ttgccatattccacgtcc	DNA FISH RNA FISH
Dot1l exon 5 guide RNA 1 F	This study	caccgGGTCTCCCCATACACCTCAG	CRISPR
Dot1l exon 5 guide RNA 1 R	This study	aaacCTGAGGTGTATGGGGAGACCc	CRISPR
Dot1l exon 5 guide RNA 2 F	This study	caccgCCCAGATGATTGATGAGATC	CRISPR
Dot1l exon 5 guide RNA 2 R	This study	aaacGATCTCATCAATCATCTGGGc	CRISPR
IAP_F	Berrens et al (2017)	AAGCAGCAATCACCCACTTTGG	qPCR
IAP_R	Berrens et al (2017)	CAATCATTAGATGTGGCTGCCAAG	qPCR
MERVL_F	Berrens et al (2017)	TTCTTCTAGACCTGTAACCAGACTCA	qPCR
MERVL_R	Berrens et al (2017)	TCCTTAGTAGTGTAGCGAATTTCCTC	qPCR
GSAT_F	Berrens et al (2017)	GACGACTTGAAAAATGACGAAATC	qPCR
GSAT_R	Berrens et al (2017)	CATATTCCAGGTCCTTCAGTGTGC	qPCR
Etn_F	Berrens et al (2017)	GTGGTATCTCAGGAGGAGTGCC	qPCR
Etn_R	Berrens et al (2017)	GGGCAGCTCCTCTATCTGAGTG	qPCR
MMERVK10C_F	Berrens et al (2017)	ATGTGAGCTAGCTGTTAAAGAAGGAC	qPCR

Reagents and Tools table   (continued)

Reagent/resource	Source	Identifier	Usage notes
MMERVK10C_R	Berrens et al (2017)	CTCTCTGTTTCTGACATACTTTCCTGT	qPCR
LINE1_F	Berrens et al (2017)	GACATAGACTAACAAACTGGCTACACAAAC	qPCR
LINE1_R	Berrens et al (2017)	GGTAGTGTCTATCTTTTTCTCTGAGATGAG	qPCR
Gapdh_ChIP_F	This study	TACCTGATGAACCTAAGCTGGG	qPCR (ChIP)
Gapdh_ChIP_R	This study	CAGGTTCCGAGGAGGGATAC	qPCR (ChIP)
Gapdh_RT_F	This study	CCATCAACGACCCCTTCATTGACC	qPCR (RT)
Gapdh_RT_R	This study	TGGTTCACACCCATCACAAACATG	qPCR (RT)
Actb_mm_F	This study	CCACTGTCGAGTCGCGTCC	qPCR
Actb_mm_R	This study	GCCCACGATGGAGGGGAATA	qPCR
Chr19_F	Lesch et al (2016)	TGGTTCCACACACATCTCCG	qPCR (ChIP)
Chr19_R	Lesch et al (2016)	CAGCCGAACCAGGAACTCAT	qPCR (ChIP)
HSATIIChr10_F	Bersani et al (2015)	gcattcaattcattagatgacgg	qPCR
HSATIIChr10_R	Bersani et al (2015)	ccttgacccgaatgcaatca	qPCR
HSATIIChr16a_F	Bersani et al (2015)	ccattcgttaatgcctttcg	qPCR
HSATIIChr16a_R	Bersani et al (2015)	cacgaatggaatcattgtcg	qPCR
HSATIIChr16b_F	Bersani et al (2015)	tccattcgaggattccactc	qPCR
HSATIIChr16b_R	Bersani et al (2015)	caaagggaagcaaaggaaatc	qPCR
HSATIII_M13	Valgardsdottir et al (2008)	CCGTAAAACGACGGCCAG	qPCR
HSATIII_Hur98-F	Valgardsdottir et al (2008)	TTCCATTCCAATCCTGACTCG	qPCR
HSATIII_Hur98-R	Valgardsdottir et al (2008)	AATCAACCCGAGTGCAATCG	qPCR
chr16q_F	Bersani et al (2015)	GGGGTAAAAGGGCATGTTTT	qPCR
chr16q_R	Bersani et al (2015)	CTGAAGAAGCCCACTCAAGG	qPCR
Actb_hs_F	Bersani et al (2015)	ctcttccagccttccttcct	qPCR
Actb_hs_R	Bersani et al (2015)	agcactgtgttggcgtacag	qPCR
Cell lines			
v6.5 mouse embryonic stem cells	Novus Biologicals	RRID:CVCL_C865	
Dot1l KO mESCs	This study	Available on request	
NIH/3T3 mouse fibroblasts	ATCC	RRID:CVCL_0594	
HEK293T cells	ATCC	RRID:CVCL_0063	
HCT116 cells	ATCC	RRID:CVCL_0291	
Mouse strains			
B6D2F1 mouse strain	Jackson Labs	RRID:IMSR_JAX:100006	
CAST/EiJ mouse strain	Jackson Labs	RRID:IMSR_JAX:000928	
Other reagents			
DOT1L inhibitor	Sigma Aldrich	SGC0946	
pLenti-CRISPR-V2-Puro	Addgene	52961 RRID:Addgene_52961	
MSCB-hDot1lwt plasmid	Addgene	74173 RRID:Addgene_74173	
Nontargeting shRNA	transOMIC Technologies	TLNSU4440	
Smarca5 shRNAs	transOMIC Technologies	TLMSU14222	
Software			
E-CRISP	Heigwer et al (2014)	RRID:SCR_019088	gRNA design
Proteome Discoverer v2.4	Thermo Fisher	RRID:SCR_014477	Proteomics
Mascot Search Engine v2.7	Matrix Science LLC		Proteomics
Scaffold v4.0	Proteome Software	RRID:SCR_014345	Proteomics

Reagents and Tools table (continued)

Reagent/resource	Source	Identifier	Usage notes
kallisto	Bray et al (2016)	RRID:SCR_016582	RNA-seq
STAR	Dobin et al (2013)	RRID:SCR_004463	RNA-seq
HTSeq v0.9.1	Anders et al (2014)	RRID:SCR_005514	RNA-seq
DESeq2	Love et al (2014)	RRID:SCR_015687	RNA-seq
Bowtie2	Langmead and Salzberg (2012)	RRID:SCR_016368	RNA-seq, ChIP-seq
RepEnrich2	Criscione et al (2014)		RNA-seq, ChIP-seq
edgeR	Robinson et al (2010)	RRID:SCR_012802	RNA-seq, ChIP-seq
ggplot2	Wickham (2016)	RRID:SCR_014601	Visualization
lattice (R package)	Sarkar (2008)	RRID:SCR_015662	Visualization
FASTX-Toolkit	Hannon (2010)	RRID:SCR_005534	RNA-seq, ChIP-seq
MACS2	Zhang et al (2008)	RRID:SCR_013291	ChIP-seq
BEDTools	Quinlan and Hall (2010)	RRID:SCR_006646	ChIP-seq
deepTools	Ramírez et al (2016)	RRID:SCR_016366	ChIP-seq
ChIPseeker	Yu et al (2015)	RRID:SCR_021322	ChIP-seq
Molecular Signatures Database (MSigDB)	Liberzon et al (2011), Subramanian et al (2005)	RRID:SCR_016863	Annotation

Methods and Protocols

Cell lines

Mouse embryonic stem cells (mESCs)

V6.5 male mouse embryonic stem cells (RRID:CVCL_C865) were grown in complete DMEM (Gibco, 11965-092) supplemented with 15% fetal bovine serum (Sigma-Aldrich, F2442), 1:100 pen-strep (Gibco, 15140-122), 1:100 GlutaMax (Gibco, 35050-061), 1:100 MEM NEAA (Gibco, 11140-050), Sodium Pyruvate 1:100 (Gibco, 11360-070), HEPES (Gibco, 15630-080), β-mercaptoethanol (Sigma-Aldrich, M6250), and 1:10,000 LIF (Millipore, ESG1106). Cells were grown on a monolayer of mitotically arrested mouse embryonic fibroblasts in plates coated with 0.1% gelatin, maintained at 5% CO_2 and passaged when they reached 70 to 80% confluence. Media was changed daily. No additional validation of the cell line was performed. Cells were screened monthly for mycoplasma and results were excluded if cells were found to be mycoplasma positive.

Mouse fibroblasts

NIH/3T3 cells (ATCC #CRL-1658, RRID:CVCL_0594) were grown in Dulbecco's Modified Eagle's Medium (DMEM, Life Technologies) supplemented with 10% fetal bovine serum (Life Technologies), 1 mM L-glutamine (Life Technologies), 100 U/ml penicillin, and 100 μg/ml streptomycin (Life Technologies). Cells were grown at 37°C at 5% CO_2, and passaged when they reached 70–80% confluence. No additional validation of the cell line was performed. Cells were screened monthly for mycoplasma and results were excluded if cells were found to be mycoplasma positive.

HCT116 cells

HCT116 cells (ATCC #CCL-247, RRID:CVCL_0291) were grown in Dulbecco's Modified Eagle's Medium (DMEM, Life Technologies) supplemented with 10% fetal bovine serum (Life Technologies), 1 mM L-glutamine (Life Technologies), 100 U/ml penicillin, and 100 μg/ml streptomycin (Life Technologies). Cells were grown at 37°C at 5% CO_2, and passaged when they reached 70–80% confluence. No additional validation of the cell line was performed. Cells were screened monthly for mycoplasma and results were excluded if cells were found to be mycoplasma positive.

Mice

All mice used in these studies were maintained and euthanized according to the principles and procedures described in the National Institutes of Health Guide for the Care and Use of Laboratory Animals. These studies were approved by the Yale University Institutional Animal Care and Use Committee under protocols 2020-20169 and 2020-20357 and conducted in accordance with the specific guidelines and standards of the Society for the Study of Reproduction. B6D2F1 strain (RRID:IMSR_JAX:100006) female mice (age 6–8 weeks, Jackson Labs) were superovulated for oocyte collection (see below) and zygotes were generated by ICSI using thawed CAST/EiJ strain sperm (RRID:IMSR_JAX:000928). For natural mating, superovulated B6D2F1 females were crossed with B6D2F1 males and zygotes isolated the following morning.

Generation of Dot1l knockout mESC lines

Dot1l knockout (KO) mESCs were generated using CRISPR/Cas9. Briefly, paired gRNAs were designed using E-CRISP (Heigwer et al, 2014) to target exon 5 of *Dot1l*, and cloned into pLenti-CRISPR-V2-Puro (Addgene # 52961) using the BsmBI restriction site as described previously (Shalem et al, 2014). gRNA sequences are provided in the Reagents and Tools table above. v6.5 mouse ESCs were seeded on gelatinized (0.2% gelatin) plates and transfected with gRNA vectors using the Fugene transfection reagent (BioRad F6-1000). 14–18 h after transfection, media was changed and puromycin-resistant feeder cells (Thermo Fisher #A34959) were added. 4–6 h after adding feeders (~24 h after transfection), puromycin (2 μg/ml) was added and transfected mESCs were selected in puromycin for 48 h. After 7 days of selection, individual colonies

were picked into 96-well plates and allowed to regrow for another 2–3 days. Cells were lysed overnight at 37°C in lysis buffer (10 mM Tris–HCl, pH 7.5, 10 mM EDTA, 10 mM NaCl, 0.5% sarcosyl, 1 mg/ml proteinase K) and genomic DNA was isolated. The targeted region (exon 5) of Dot1l was PCR amplified and the deletion was validated using Sanger sequencing.

Generation of Smarca5 knockdown mESC lines

Lentiviral vectors encoding either a nontargeting shRNA (TLNSU4440) or two specific shRNAs directed against mouse *Smarca5* (ULTRA3453660 and ULTRA3454387, transOMIC technologies, Cat#TLMSU14222), were co-transfected with VSV-G and psPAX2 (Addgene #12260) plasmids into HEK293T cells (RRID: CVLC_0063) using Lipofectamine 3000 transfection reagent (Invitrogen, #L3000008) and incubated at 37°C, 5% CO_2 to generate lentiviral particles. After 48 h, viral particles were harvested from the culture supernatant by filtering through a 0.45 μm syringe filter unit. Viral particles harboring either non-targeting control or *Smarca5* directed shRNA were used to infect v6.5 mESCs following treatment with polybrene (8 μg/ml, Millipore Sigma, # TR-1003-G) for 2 h. After 48 h, transduced cells were selected with 2 μg/ml puromycin (Gibco, #A1113803) with daily media changes. Once cells reached optimal growth and formed clearly distinguishable colonies, individual colonies were picked and expanded in a 96-well plate. All colonies were maintained in complete mESC media as described above. Selected clones were expanded into a 24-well plate and knockdown of SMARCA5 was confirmed by immunoblot analysis.

Mouse embryo collection

B6D2F1 strain female mice (age 6–8 weeks, Jackson Labs) were superovulated by serial Pregnant Mare Serum Gonadotropin (5 IU per mouse, Prospec Protein Specialists) and human chorionic gonadotropin (5 IU, MilliporeSigma) injections 47 h apart. The following day, MII stage oocytes were isolated in M2 media supplemented with hyaluronidase (MilliporeSigma) and stored in 25 μl drops of pre-gassed KSOM with half-strength concentration of amino acids (MilliporeSigma) under mineral oil (Irvine Scientific). Zygotes were generated by piezo-actuated intracytoplasmic sperm injection (ICSI; see Grosswendt *et al*, 2020) using thawed CAST/EiJ strain sperm in batches of 30–50 oocytes and standard micromanipulation equipment, including a Hamilton Thorne XY Infrared laser, Eppendorf Transferman NK2 micromanipulators, and a Zeiss Axio Observer inverted microscope. Alternatively, superovulated B6D2F1 females were crossed with B6D2F1 males and zygotes isolated the following morning.

For DOT1L inhibitor treatment experiments, 10 μM SGC0946 or 25 μM EPZ5676 in KSOM was plated the night before and isolated oocytes were deposited directly into these drops prior to ICSI, after which they were returned to new drug containing drops and washed several times to serially dilute the M2 media. Then, progression to set morphological stages was conducted by visual inspection at 24 hpf (2-cell), 48 hpf (4-cell), 56 hpf (8-cell), 72 hpf (morula/early blastocyst), 84 hpf (blastocyst), and 96 hpf (late blastocyst).

Embryos were prepared for fixing in 4% paraformaldehyde by removal of the zona pellucida using Acid Tyrode's Solution (MilliporeSigma) and directly plating onto Poly-L-lysine coated dishes (MatTek) in PBS.

DNA fluorescence In situ hybridization (DNA FISH)

The FISH protocol was modified from Cesare *et al* (2015). mESCs were grown as stated earlier in 35-mm glass bottom MatTek dishes (MatTek, #P35GC-1.5-14-C) overnight. The next day, the culture medium was aspirated and the cells were washed twice with 1.5 ml of cold PBS. Cells were incubated with 500 μl diluted pre-extraction buffer (0.05% Triton X-100, 2 mM HEPES-KOH, pH 7.9, 5 mM NaCl, 0.3 mM $MgCl_2$, 30 mM sucrose) on ice for up to 3 min. Cells were then fixed with 4% paraformaldehyde at room temperature for 10 min. Following fixation, cells were rinsed twice with 1.5 ml distilled, deionized water and then permeabilized with 2 ml KCM permeabilizing solution (120 mM KCl, 20 mM NaCl, 10 mM Tris·Cl, pH 7.5, 0.1% (v/v) Triton X-100) at room temperature for 10 min. Cells were blocked with ABDIL blocking solution (20 mM Tris·Cl, pH 7.5, 2% (w/v) bovine serum albumin (BSA), 2% (v/v) normal goat serum, 0.2% (w/v) gelatin (Sigma), 150 mM NaCl, 0.1% (v/v) Triton X-100, 0.1% (w/v), and sodium azide) containing 100 μg/ml DNase-free RNase A at room temperature for 1 h. Primary antibodies diluted in ABDIL solution were added to cells and incubated overnight at 4°C in a humidified chamber. Following incubations, cells were washed three times, each time with PBST for 10 min at room temperature with gentle shaking. AlexaFluor568-conjugated secondary antibodies diluted in ABDIL solution were added to cells and incubated at room temperature for 1 h in a humidified chamber protected from direct light. From here onwards, all steps were performed in dark. Cells were washed three times for 10 min in PBST and fixed with 4% paraformaldehyde for 10 min at room temperature. Cells were rinsed with distilled deionized water as mentioned and then ethanol dehydrated with a graded series of ethanol: 3 ml 70% ethanol for 3 min, 3 ml 90% ethanol for 3 min, then 3 ml 100% ethanol for 3 min. Cells were allowed to air dry to completion for approximately 30 min. In the meantime, incubator was preheated to 85°C, and FISH hybridization buffer (20 Mm Tris, pH 7.4, 60% formamide, 0.5% of blocking reagent (Roche 11096176001)), and cells were prewarmed in the incubator for 5 min. Mouse major satellite probe conjugated to Alexa488 (PNA Bio, Alexa488-o-ttgccatattccacgtcc) was diluted in prewarmed hybridization buffer at a final concentration of 500 nM and then added to cells. Cells were incubated at 85°C for 12 min for denaturation and then allowed to hybridize for 8–10 h at room temperature in the dark in a humidified chamber. Cells were washed twice with gentle shaking, each time for 10 min first with 1.5 ml PNA wash solution A (70% (v/v) formamide (deionized) 10 mM Tris·Cl, pH 7.5) and then with PNA wash solution B (50 mM Tris·Cl, pH 7.5,150 mM NaCl 0.8% (v/v) Tween 20). Cells were counterstained with 50 ng/ml DAPI and rinsed twice with deionized water. Cells were dehydrated through a graded ethanol series and air dried as described above, and finally mounted with Prolong Gold and imaged on LSM 880 confocal microscope (Zeiss) using ZEN acquisition software.

Co-localization analysis was performed in ZEN Blue software (Zeiss). Briefly, a fixed threshold was set across all images to exclude background fluorescence and the percentage of co-localization of H3K79me2 or H3K79me3 with Alexa488 Maj-Sat in the entire cell was calculated by the ratio of the number of co-localized pixels of H3K79me2 or H3K79me3 to the total H3K79me2 or H3K79me3 pixels.

Immuno-RNA FISH

Immuno-RNA FISH was performed as previously reported (Ogawa & Ogawa, 2021), with slight modifications. Briefly, ESCs were grown overnight on 0.2% gelatin and gently washed with ice-cold PBS on ice for 5 min. Cells were permeabilized with ice-cold 0.5% Triton X-100/CSK buffer (10 mM Pipes, pH 7, 100 mM NaCl, 300 mM sucrose, 3 mM MgCl$_2$) supplemented with 2 mM vanadyl ribonucleoside complex (VRC; Sigma) for 5 min on ice, washed twice with cold CSK buffer. For the RNase A-treated control, there was an additional incubation with 1 mg/ml RNase A (Millipore) for 20 min at room temperature. After briefly washing with ice-cold PBS for 1 min, cells were fixed with 4% paraformaldehyde for 12 min at room temperature followed by two PBS washes. Cells were blocked in blocking solution containing 5% normal goat serum, 2% BSA, 0.4 U RNaseOUT (Life Technologies) in PBS for 45 min at room temperature. Cells were then incubated with the anti-H3K79me3 primary antibody (Reagents and Tools table) in blocking solution overnight at 4°C. After washes in PBS, the cells were incubated with the fluorophore conjugated secondary antibody (Alexa Fluor 568 donkey anti-rabbit: Molecular Probes, Reagents and Tools table). The cells were then postfixed in 4% PFA for 10 min and washed twice with PBS. Following dehydration with 70, 90, and 100% ethanol, cells were incubated with FISH hybridization buffer (50% formamide, 0.25 mg/ml salmon sperm DNA, 0.25 mg/ml yeast tRNA, 10% dextran sulfate, 1 mg/ml BSA, 10 mM VRC, 4× SSC) containing 0.5 mM MajSat PNA fluorescent probe (PNA Bio) for 4 h at 37°C in the dark. Following hybridization, cells were washed with prewarmed 2 × SSC for 20 min at 37°C, 1 × SSC for 20 min at room temperature on a shaker, and 4 × SSC for 2 min at room temperature. Finally, cells were stained with 0.5 ng/ml DAPI for 5 min in 2 × SSC, and slides were mounted with Vectashield antifade mounting medium and sealed around the edges of the coverslip with nail polish.

Metaphase spreads

V6.5 mESCs were plated on gelatinized six-well dishes and allowed to grow for 1–2 days. One hour prior to addition of Colcemid (Karyomax, Gibco #15210-040), cells were fed with fresh ESC media. Cells were treated with Colcemid to a final concentration 1 μg/ml and allowed to grow at 37°C for 1 h. After rinsing with PBS, cells were trypsinized and pelleted down at 200 g for 5 min. The supernatant was aspirated, leaving about 200 μl media in the tube. The cell pellet was resuspended by gently tapping the bottom of the tube, and 5 ml of ice cold 0.56% KCl solution was added slowly and mixed by inverting the tube once. Cells were incubated at room temperature for 6 min and then spun at 200 g for 4 min. The supernatant was aspirated and the pellet was resuspended in the remaining drops of liquid by gently tapping the bottom of the tube. Cells were fixed with 5 ml of fixative (methanol: glacial acetic acid, ratio 3:1). Fixative was added dropwise with continuous mixing. After 5 min cells were centrifuged at 200 g for 4 min. Finally, cells were resuspended in 500 μl of remaining fixative and released one drop at a time onto an alcohol cleaned slide with a 20 μl pipettor by holding the slide slightly tilted over a jar of water. Slides were air dried for 2 h at room temperature. After adding 1 or 2 drops of antifade DAPI mounting media (Invitrogen), slides were coverslipped and imaged on LSM 980 confocal microscope (Zeiss) as mentioned above using ZEN acquisition software.

Immunofluorescence staining and imaging of mESCs

Wild-type and *Dot1l* KO ESCs were plated on 0.2% gelatin (10 μg/ml) coated 35 mm coverslip dishes (MatTek corp. #P35GC-1.5-14-C.S) for overnight growth. Cells were fixed in 4% paraformaldehyde for 15 min, washed with DPBS and permeabilized with 0.2% Triton X-100 in PBS for 10 min. Cells were then blocked in blocking buffer (5% BSA, 0.1% Triton X-100 or 4% FBS, 0.1% Triton X-100) for 1 h at room temperature (RT) and then incubated overnight at 4°C with primary antibodies diluted in blocking buffer. The dishes were washed with 0.1% PBST and stained with fluorophore-conjugated secondary antibodies: AlexaFluor 488-conjugated goat anti-rabbit-IgG (Molecular Probes), AlexaFluor 568-conjugated goat anti-rabbit-IgG (Molecular Probes), AlexaFluor 568-conjugated goat anti-mouse-IgG (Molecular Probes). All secondary antibodies were used at 1:700 dilution in blocking buffer and incubated at RT for 1 h. Finally, the slides were mounted in antifade mounting medium containing DAPI (Invitrogen). Images were acquired with LSM 710 or LSM 980 confocal microscope (Zeiss) equipped with 405, 488, 555/561 nm and two-photon lasers, and fitted with a 63× 1.4 NA objective using ZEN acquisition software.

Immunofluorescence staining and imaging of embryos

Embryos were grown *in vitro* from 4 to 96 h with or without treatment of DOT1L inhibitor SGC0946. Embryos were collected at the appropriate time point and fixed with 4% paraformaldehyde for 20 min at 4°C. After three washes with PBS-1% BSA for 5 min each, embryos were permeabilized with 0.2% Triton X-100 for 20 min at room temperature and blocked with blocking solution (5% BSA, 0.15% Triton X-100) for 1 h at room temperature. Embryos were washed three times with wash buffer (0.1% PBS-Triton X-100) and incubated with primary antibodies overnight at 4°C in blocking solution. The next day, embryos were washed with wash buffer and stained with fluorophore-conjugated secondary antibodies: AlexaFluor 488-conjugated goat anti-mouse-IgG (Molecular Probes) or Alexa Fluor 568-conjugated goat anti-mouse-IgG (Molecular Probes). All secondary antibodies were used at 1:700 dilution in blocking buffer and incubated at RT for 1 h. Finally, the slides were mounted in antifade mounting medium containing DAPI (Invitrogen). Images were acquired with LSM 710 or LSM 980 confocal microscope (Zeiss) equipped with 405, 488, 555/561 nm and two-photon lasers and fitted with a 63× 1.4 NA objective using ZEN acquisition software.

Real-time quantitative PCR

Reverse transcription of 1 μg of total RNA was performed with oligo dT primers and iScript reverse transcriptase (Bio-Rad #1708896) in a total volume of 20 μl according to the manufacturer's instructions. Reaction mixtures were incubated in a thermocycler at 25°C for 5 min then 42°C for 60 min before stopping the reaction at 95°C for 1 min. 1 μl of undiluted cDNA was used for a 20 μl reaction volume consisting of 4 μl of 10 μM forward plus reverse primer mix, 10 μl of Power SYBR Green PCR Master Mix (Applied Biosystems #4367659) and 5 μl nuclease-free water. Primer sequences used for qPCR are listed in the "Primers" subsection above. Reactions for each target gene were performed in duplicate in a 96 well plate loaded into an Applied Biosystems QuantStudio 3 Real-Time PCR System. Standard cycling conditions were used: Hold stage (×1): 50°C for 2 min, 95°C for 10 min; PCR stage (×40): 95°C for 15 s,

60°C for 1 min. Melt curve stage conditions were: 95°C for 15 s, 60°C for 1 min, 95°C for 15 s. Relative fold change in transcript abundance was calculated using the delta–delta Ct method by normalizing target gene expression levels to either *Gapdh* or *Actb*.

Western blot analysis

Cells were lysed in RIPA buffer (50 mM Tris–HCl pH 7.4, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, 1 mM EDTA, 10 mM NaF, 1 mM PMSF) containing fresh cOmplete EDTA-free protease inhibitor cocktail (Roche). Cell extracts were subjected to sonication using a Bioruptor bath sonicator (Diagenode) and centrifuged at 20,000 × g for 15 min at 4°C and supernatants were immediately mixed with 4× SDS sample buffer to a final dilution of 1×. The samples were heated at 95°C for 5 min, resolved on a Mini-PROTEAN TGX gel (Bio-Rad, 456-8093) by SDS–PAGE for 2 h at 80 V and transferred onto PVDF membranes (GE Healthcare) in 20 mM Tris–HCl (pH 8.0), 150 mM glycine, 20% methanol. Following transfer, membranes were blocked with 5% skim (nonfat) milk for 1 h at RT, followed by incubation with the primary antibody overnight at 4°C. After washing three times with TBST buffer (20 mM Tris–HCl pH 7.6, 150 mM NaCl and 0.1% Tween 20), membranes were incubated at RT for 1 h with HRP-conjugated secondary antibodies as follows: goat anti-rabbit IgG conjugated to HRP (Jackson Immuno Research, 111-035-003, 1:20,000) for H3K79me1, H3K79me2, H3K79me3, SMARCA5, DOT1L; and goat anti-mouse IgG conjugated to HRP (Jackson Immuno Research, 111-035-003, 1:20,000) for CHD4, HMGA1 and GAPDH. After washing the membrane three times with TBST, specific protein bands were detected using a SuperSignal West Pico PLUS Chemiluminescent Substrate (ThermoFisher #34577). Chemiluminescence was detected using the FluorChem E (Protein Simple) documentation system. Densitometry analysis of bands was performed by using ImageJ software (Fiji) or the multiplex band analysis tool in AlphaView software (Protein Simple). The band intensity of the protein of interest was normalized to the loading control.

Co-immunoprecipitation of endogenous DOT1L

Wild-type and *Dot1l* KO mESCs were grown in 10 cm dishes. To harvest, cells were washed with ice cold PBS twice and lysed in ice-cold lysis buffer (25 mM Tris.HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP40, 0.5% Triton X-100, 5% glycerol, 1 mM PMSF, 2 mM NaF, 1× Protease inhibitor cocktail (Roche)) for 5 min on ice. Cells were harvested by scraping, and lysate was incubated at 4°C for 30 min with constant agitation. The lysate was centrifuged at 14,000 × g for 10 min at 4°C and the supernatant was transferred to a fresh 1.5-ml tube. Preclearing was performed for 1 h at 4°C by adding 20 μl of Protein G Dynabeads to the lysate. After a brief spin, supernatant was collected on a magnetic stand. 10% of the lysate was set aside as input. The remaining lysate was mixed with 10 μg of primary antibody and incubated overnight at 4°C with end-over-end mixing. 20 μl of fresh beads was added to the lysate and incubated at 4°C for 2–4 h with end-over-end rotation. Beads were collected on a magnetic stand and washed three times in wash buffer (10 mM Tris.HCl pH 7.4, 1 mM EDTA, 1 mM EGTA pH 8.0, 150 mM NaCl, 1% Triton X-100, 0.2 mM sodium orthovanadate, Protease inhibitor cocktail). Finally, the antigen–antibody complex was eluted by heating the beads in 4× SDS buffer at 95°C for 10 min. Co-immunoprecipitated proteins were analyzed by SDS–PAGE.

Co-immunoprecipitation of HA-tagged DOT1L for mass spectrometry

Dot1l KO mESCs were transfected with MSCB-hDot1lwt plasmid (Addgene #74173) expressing HA-tagged human DOT1L. HA-tagged DOT1L was immunoprecipitated from a confluent 10 cm dish using the Pierce HA-Tag Magnetic IP/Co-IP Kit (Pierce™ HA-Tag Magnetic IP/Co-IP Kit, Cat # 88838) following the manufacturer's instructions with some modifications. A corresponding untransfected dish was kept as a negative control. Briefly, transfected and untransfected *Dot1l* KO ESCs were washed with ice-cold PBS twice and lysed in ice-cold lysis buffer (25 mM Tris.HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP40, 0.5% Triton X-100, 5% glycerol, 1 mM PMSF, 2 mM NaF, 1× Protease inhibitor cocktail (Roche)) for 5 min on ice. Cells were harvested by scraping and lysate was incubated at 4°C for 30 min with constant agitation. Lysate was centrifuged at 14,000 × g for 10 min at 4°C, and the supernatant was transferred to a fresh 1.5-ml tube. Preclearing was performed for 1 h at 4°C by adding 20 μl of Protein G Dynabeads (Thermo Scientific #10007D) to the lysate. After a brief spin, supernatant was collected on a magnetic stand. 30 μl (0.30 mg) of Pierce Anti-HA magnetic beads was added to a fresh 1.5-ml microcentrifuge tube and washed with lysis buffer several times. Lysate containing the HA-tagged protein was added to the pre-washed anti-HA magnetic beads and incubated at 4°C for 8 h with constant mixing. After two washes with lysis buffer and one wash with ultrapure water, beads were collected and eluted with elution buffer according to the manufacturer's instructions. Eluted sample was mixed with 5× reducing sample buffer, heated at 95°C for 10 min, and anaylzed by SDS–PAGE. For mass spec analysis, immunoprecipitated proteins were resolved for up to 1 cm on an SDS–PAGE gel. The gel was fixed in 30% ethanol and 10% acetic acid in water. The gel piece was excised and submitted to the Keck MS & Proteomics Resource at Yale University for analysis.

Mass spectrometry sample processing and analysis

Excised 1D SDS polyacrylamide gel band/plug corresponding to co-IP eluted proteins were subjected to *in situ* enzymatic digestion. The gels are washed with 250 μl 50% acetonitrile/50% water for 5 min followed by 250 μl of 50 mM ammonium bicarbonate/50% acetonitrile/50% water for 30 min. One final wash is done using 10 mM ammonium bicarbonate/50% acetonitrile/50% water for 30 min. After washing, the gel plugs are dried in a Speedvac and rehydrated with 0.1 μg of either modified trypsin (Promega), per (approximately) 15 mm³ of gel in 15 μl 10 mM ammonium bicarbonate. Samples are digested at 37°C for 16 h (overnight). Digest is then centrifuge at supernatant/solution is transferred to injection vial to be injected onto a Waters NanoACQUITY UPLC coupled to a Q-Exactive Plus mass spectrometer. The Waters nanoACQUITY UPLC system uses a Waters Symmetry® C18 180 μm × 20 mm trap column and a 1.7 μm, 75 μm × 250 mm nanoAcquity™ UPLC™ column (35°C) for peptide separation. Trapping is done at 5 μl/min, 99% Buffer A (100% water, 0.1% formic acid) for 3 min. Peptide separation is performed at 300 nl/min with Buffer A: 100% water, 0.1% formic acid and Buffer B: 100% CH₃CN, 0.075% formic acid. For the MS and MS/MS Q-Exactive Plus data collection, High-energy Collisional Dissociation (HCD) MS/MS spectra filtered by dynamic exclusion was acquired over a 3 s duty cycle for charge states 2–8 with m/z isolation window of 1.6. All MS (Profile) and MS/MS (centroid) peaks were detected in the Orbitrap. Trapping was carried out

for 3 min at 5 μl/min in 97% Buffer A (0.1% FA in water) and 3% Buffer B [0.075% FA in acetonitrile (ACN)] prior to eluting with linear gradients that will reach 5% B at 1 min, 255% B at 90 min, and 50% B at 110 min, and 90% B at 115 min for 5 min; then dropdown to 3% B at 125 min for 5 min.

LC MS/MS data were analyzed utilizing Proteome Discoverer 2.4 (Thermo Fisher Scientific) with Mascot search engine (v. 2.7 Matrix Science LLC.). Resulting PD analyses were imported into Scaffold (v. 4.0, Proteome Software) for further interrogation. Positive protein identification were based on hits with two or more unique peptides per protein, and peptides were considered significant if the Mascot Score is better than the 95% confidence level.

Chromatin immunoprecipitation (ChIP)

Cells were cross-linked with 1% formaldehyde at room temperature for 10 min. Formaldehyde was quenched with 0.1375 M glycine at room temperature for 10 min. Fixed cells were spun down at 200 g for 5 min at 4°C. Cells were washed with cold PBS twice and resuspended in 50 μl ChIP lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tri-HCl at pH 8.1) and frozen in −80°C. Antibody-bound Dynabeads (Thermo Fisher #10007D) were prepared by mixing 10 μl aliquots of beads with 100 μl block solution (0.5% BSA in PBS). Beads were washed twice with 150 μl of block solution and resuspended in 30 μl block solution. 1 μg antibody per million cells for H3K79me2, H3K79me3, HP1β, or RNA PolIIS2P antibodies was added to each aliquot of beads and incubated for 8 h rotating at 4°C. Between 2 and 4×10^6 cells were used for ChIP-seq and ChIP-qPCR. Frozen cells were thawed on ice and ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl, 16.7 mM Tris–HCl at pH 8.1) was added to reach a total volume of 150 μl. Cells were sonicated at 4°C for 30 cycles (30 s on/off) using a Bioruptor bath sonicator (Diagenode). Aliquots of the same samples were pooled and spun down at $12,000 \times g$ for 5 min. The supernatant was moved to a new 1.5 ml Eppendorf tube and 600 μl of dilution buffer and 100 μl protease inhibitor cocktail (Roche #11836153001) were added. 50 μl of each sample was set aside as input before an aliquot of primary antibody-bound Dynabeads was added to the lysate and incubated overnight rotating at 4°C. After overnight incubation, beads were washed twice with low-salt immune complex wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl, and 20 mM Tris–HCL at pH 8.1), twice with LiCl wash buffer (0.25 M LiCl, 1% NP40, 1% deoxycholate, 1 mM EDTA, and 10 mM Tris–HCl at pH 8.1), and twice with TE (1 mM EDTA and 10 mM Tris–HCl at pH 8.0). Bound DNA was eluted twice with 125 μl elution buffer (0.2% SDS, 0.1 M NaHCO3, and 5 mM DTT in TE) at 65°C and crosslink reversal was performed by incubating at 65°C for 8–15 h. ChIP and input samples were incubated for 3 h at 37°C with 0.2 mg/ml RNAse A (Millipore 70856-3), and 4–10 h at 55°C with 0.1 mg/ml Proteinase K (NEB P8107S). Sample DNA was prepared using a Zymo ChIP DNA Clean & Concentrator kit (Zymo Research #D5201) according to the manufacturer's instructions. Columns were washed twice with 200 μl wash buffer and DNA was eluted in 20 μl Elution Buffer into fresh Eppendorf tubes, then re-eluted with the same eluate to enhance the yield.

RNA isolation and sequencing library preparation

For mESCs, the cell pellet was resuspended in 1 ml of TRIzol and cells were homogenized by drawing up and down in a 21G needle with 3-ml syringe. 200ul chloroform was added, then samples were vortexed and incubated briefly at room temperature and centrifuged at $12,000 \times g$, 15 min, 4°C. The aqueous phase was transferred to a gDNA eliminator column from the RNEasy Micro Plus kit (Qiagen), and total RNA was isolated according to manufacturer's instructions. For embryos, two replicates of 30–35 mouse embryos at 56-h postfertilization were each disrupted in 75 μl Buffer RLT, and total RNA was prepared using the RNeasy Micro kit (Qiagen #74004) according to the manufacturer's instructions. RNA was eluted in 12 μl RNase-free water.

Sequencing library preparation

DNA or RNA integrity and fragment size were confirmed on a BioAnalyzer (Agilent #G2939BA). For ChIP-seq, approximately 5–10 ng of DNA was end-repaired, A-tailed, adapter-ligated, and PCR enriched (8–10 cycles) using the KAPA Hyper Library Preparation kit (KAPA Biosystems, #KK8504) according to the manufacturers' instructions. For RNA-seq, libraries were prepared by ribosome depletion using the KAPA RNA Hyperprep Kit with RiboErase (Roche #08098140702) (mESCs) or the SMARTer Stranded Total RNA-seq kit v3 – Pico Input (Takara Bio #634485) (embryos) according to the manufacturers' instructions. Indexed libraries were quantitated by qPCR using the KAPA Library Quantification Kit (KAPA Biosystems #KK4854). All samples were sequenced on an Illumina NovaSeq using 100 bp paired-end sequencing. Demultiplexing was performed using CASAVA 1.8.2 (Illumina).

RNA-seq data analysis

For single-copy genes, data were aligned and assembled with kallisto (Bray et al, 2016) or STAR (Dobin et al, 2013) using default parameters and mouse genome assembly mm10. Read counts from STAR alignments were generated using HTSeq (Anders et al, 2014). Differential gene expression was called using DESeq2 (Love et al, 2014), with a cutoff P-value of < 0.05 after adjusting for multiple comparisons. For analysis of repetitive elements, data were first aligned to the mm10 genome assembly using Bowtie2 (Langmead & Salzberg, 2012) and multi-aligning reads were parsed using RepEnrich2 (Criscione et al, 2014). Differential repeat expression was called using EdgeR (Robinson et al, 2010) according to the RepEnrich2 pipeline. Plotting was done in R using the ggplot2 (Wickham, 2016) and lattice (Sarkar, 2008) packages.

ChIP-seq data analysis

Data were filtered for high-quality reads using the fastq_quality_filter tool from FASTX-Toolkit (Hannon, 2010) with parameters -q 20 -p 80 and aligned to the mm10 genome assembly using Bowtie2 with –end-to-end --fast parameters. Peaks were called using MACS2 with default parameters (Zhang et al, 2008). Peak intersections were evaluated using BEDTools (Quinlan & Hall, 2010). Analysis of repetitive elements was performed using RepEnrich2, and repeats were called as significantly enriched at false discovery rate < 0.05 for ChIP compared to input control. Heatmap and metagene profiles were generated using the deepTools software (Ramírez et al, 2016). H3K4me3, H3K36me3, and H3K27ac bigwig files used to generate Fig EV2D were downloaded from ENCODE (files ENCFF397UZX, ENCFF048JVJ, and ENCFF610BPO; ENCODE Consortium, 2011, 2012). Genome and genomic feature distributions were analyzed using the ChIPseeker package in R (Yu et al,

2015). Metagene and correlation plots were generated using Deep-Tools plotProfile and plotCorrelation, respectively (Ramírez *et al*, 2016).

Motif analysis

Motif enrichment analysis was performed using HOMER (findMotifsGenome.pl with -mask and -size given; Heinz *et al*, 2010).

ChromHMM analysis

Chromatin state annotations were generated using ChromHMM (version 1.23; Ernst & Kellis, 2012), following the protocol from the authors (Ernst & Kellis, 2017). Two replicates of H3K79me2 and H3K79me3 ChIP-seq in wild-type mESCs and an associated input control file were used as input (8 files total). Processed .bam files aligned to mm10 were binarized using "BinarizeBam" and output was used to call "LearnModel" to generate a four-state model. The standard emissions output is shown in Fig 1I, with the transition probabilities matrix shown in Fig EV2G. The RefSeq annotation fold enrichment was generated automatically from the LearnModel call (Fig EV2G), using mm10 annotations. The fold enrichment for repeat element classes was generated using "OverlapEnrichment" (Fig 1I). External coordinates for enrichment in repeat elements was generated from the mm10 RepeatMaskerViz.bed (Smit *et al*, 2013). The number of elements in each class of repeat elements is listed in Fig EV2G.

Peak-gene assignments and gene ontology analysis

We assigned peaks to genes if the peak overlapped with the gene body as annotated by Ensembl v101 (Cunningham *et al*, 2022) by at least one base pair, using bedtools intersect (Quinlan & Hall, 2010) to identify overlaps. For ChIP-seq peaks, gene ontology (GO) enrichment analysis was conducted using the R package GOStats (Falcon & Gentleman, 2007). GOStats calculates significant enrichment using a two-step significance test. First, statistical significance is conditioned on the GO graph structure. Then, enrichment is calculated based on a hypergeometric test. P-values were corrected for multiple hypothesis testing using the Benjamini–Hochberg method. Analysis of functional enrichments for RNA-seq and proteomics data was performed using the Molecular Signatures Database tool (MSigDB; Subramanian *et al*, 2005; Liberzon *et al*, 2011) to search the Gene Ontology Biological Function database (Ashburner *et al*, 2000; Gene Ontology Consortium, 2021).

Statistical analysis

Assays were done in triplicate ($n = 3$) unless stated otherwise in the figure legend, and the meaning of n is described in each figure legend. Error bars represent standard deviation or standard error of the mean as specified in the figure legend. Statistical comparisons were performed using two-tailed unpaired t-test for continuous variables and using Fisher's Exact test for proportions except where specified. For comparison of proportions in Fig 1G, two-sided tests of proportions were performed using the prop.test() function in R. When multiple comparisons were performed, P-values were corrected as described in the figure legends or relevant Methods sections. Differences were considered statistically significant at $P < 0.05$ (for single comparisons) or false discovery rate (FDR) < 0.05 (for multiple comparisons). Statistics for

genomics and proteomics data analysis are described in the relevant Methods sections above.

Exclusion criteria and blinding

In qPCR assays, melting curves were assessed and individual samples were excluded if there was evidence of off-target amplification. No blinding was done for this study.

Data availability

RNA-seq and ChIP-seq datasets have been deposited at GEO and are publicly available as of the date of publication under accession number GSE182744. Mass spectrometry datasets have been deposited at the ProteomeXchange Consortium via the PRIDE respository and are publically available as of the date of publication under accession number PXD028824. Original Western blot images, raw microscopy data, and raw qPCR data reported in this paper are available upon request.

Expanded View for this article is available online.

Author contributions

Aushaq B Malla: Conceptualization; formal analysis; validation; investigation; visualization; writing – original draft. **Haoming Yu:** Formal analysis; validation; investigation. **Delaney Farris:** Formal analysis; visualization. **Srilekha Kadimi:** Formal analysis; investigation. **TuKiet T Lam:** Formal analysis; investigation. **Andy L Cox:** Writing – review and editing. **Zachary D Smith:** Resources; investigation; writing – review and editing. **Bluma J Lesch:** Conceptualization; resources; formal analysis; supervision; funding acquisition; visualization; project administration; writing – review and editing.

Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

References

Abe M, Tsai SY, Jin S-G, Pfeifer GP, Szabó PE (2011) Sex-specific dynamics of global chromatin changes in fetal mouse germ cells. *PLoS ONE* 6: e23848

Abe Y, Sako K, Takagaki K, Hirayama Y, Uchida KSK, Herman JA, DeLuca JG, Hirota T (2016) HP1-assisted Aurora B kinase activity prevents chromosome segregation errors. *Dev Cell* 36: 487 – 497

Aihara T, Miyoshi Y, Koyama K, Suzuki M, Takahashi E, Monden M, Nakamura Y (1998) Cloning and mapping of SMARCA5 encoding hSNF2H, a novel human homologue of Drosophila ISWI. *Cytogenet Cell Genet* 81: 191–193

Almouzni G, Probst AV (2011) Heterochromatin maintenance and establishment: lessons from the mouse pericentromere. *Nucleus* 2: 332–338

Anders S, Pyl PT, Huber W (2014) HTSeq-a python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166–169

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al* (2000) Gene ontology: tool for the unification of biology. *Nat Genet* 25: 25–29

Bernt KM, Zhu N, Sinha AU, Vempati S, Faber J, Krivtsov AV, Feng Z, Punt N, Daigle A, Bullinger L *et al* (2011) MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell* 20: 66–78

Berrens RV, Andrews S, Spensberger D, Santos F, Dean W, Gould P, Sharif J, Olova N, Chandra T, Koseki H *et al* (2017) An endosiRNA-based repression mechanism counteracts transposon activation during global DNA demethylation in embryonic stem cells. *Cell Stem Cell* 21: 694–703

Bersani F, Lee E, Kharchenko PV, Xu AW, Liu M, Xega K, MacKenzie OC, Brannigan BW, Wittner BS, Jung H *et al* (2015) Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. *Proc Natl Acad Sci USA* 112: 15148–15153

Biscotti MA, Canapa A, Forconi M, Olmo E, Barucca M (2015) Transcription of tandemly repetitive DNA: functional roles. *Chromosome Res* 23: 463–477

Bitoun E, Oliver PL, Davies KE (2007) The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Hum Mol Genet* 16: 92–106

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34: 525–527

Bulut-Karslioglu A, Perrera V, Scaranaro M, de la Rosa-Velazquez IA, van de Nobelen S, Shukeir N, Popow J, Gerle B, Opravil S, Pagani M *et al* (2012) A transcription factor–based mechanism for mouse heterochromatin formation. *Nat Struct Mol Biol* 19: 1023–1030

Burton A, Torres-Padilla M-E (2010) Epigenetic reprogramming and development: a unique heterochromatin organization in the preimplantation mouse embryo. *Brief Funct Genomics* 9: 444–454

Burton A, Torres-Padilla M-E (2014) Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat Rev Mol Cell Biol* 15: 723–735

Burton A, Brochard V, Galan C, Ruiz-Morales ER, Rovira Q, Rodriguez-Terrones D, Kruse K, Le Gras S, Udayakumar VS, Chin HG *et al* (2020) Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive H3K9me3. *Nat Cell Biol* 22: 767–778

Cao K, Ugarenko M, Ozark PA, Wang J, Marshall SA, Rendleman EJ, Liang K, Wang L, Zou L, Smith ER *et al* (2020a) DOT1L-controlled cell-fate determination and transcription elongation are independent of H3K79 methylation. *Proc Natl Acad Sci USA* 117: 27365–27373

Cao K, Ozark PA, Shilatifard A (2020b) NCBI Gene Expression Omnibus GSE134083. (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134083). [DATASET]

Casanova M, Pasternak M, El Marjou F, Le Baccon P, Probst AV, Almouzni G (2013) Heterochromatin reorganization during early mouse development requires a single-stranded noncoding transcript. *Cell Rep* 4: 1156–1167

Cesare AJ, Heaphy CM, O'Sullivan RJ (2015) Visualization of telomere integrity and function In vitro and In vivo using immunofluorescence techniques. *Curr Protoc Cytom* 73: 12.40.1–12.40.31

Chory EJ, Calarco JP, Hathaway NA, Bell O, Neel DS, Crabtree GR (2019) Nucleosome turnover regulates histone methylation patterns over the genome. *Mol Cell* 73: 61–72

Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem* 78: 273–304

Clift D, McEwan WA, Labzin LI, Konieczny V, Mogessie B, James LC, Schuh M (2017) A method for the acute and rapid degradation of endogenous proteins. *Cell* 171: 1692–1706

Criscione SW, Zhang Y, Thompson W, Sedivy JM, Neretti N (2014) Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* 15: 583

Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R *et al* (2022) Ensembl 2022. *Nucleic Acids Res* 50: D988–D995

Daigle SR, Olhava EJ, Therkelsen CA, Basavapathruni A, Jin L, Boriack-Sjodin PA, Allain CJ, Klaus CR, Raimondi A, Scott MP *et al* (2013) Potent inhibition of DOT1L as treatment of MLL-fusion leukemia. *Blood* 122: 1017–1025

Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, Hilton JA, Jain K, Baymuradov UK, Narayanan AK *et al* (2018) The encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* 46: D794–D801

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21

ENCODE Project Consortium (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol* 9: e1001046

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74

Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 9: 215–216

Ernst J, Kellis M (2017) Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 12: 2478–2492

Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics* 23: 257–258

Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T *et al* (2009) The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* 41: 563–571

Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y (2002) Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol* 12: 1052–1058

Frescas D, Guardavaccaro D, Kuchay SM, Kato H, Poleshko A, Basrur V, Elenitoba-Johnson KS, Katz RA, Pagano M (2008) KDM2A represses transcription of centromeric satellite repeats and maintains the heterochromatic state. *Cell Cycle* 7: 3539–3547

Fukagawa T, Nogami M, Yoshikawa M, Ikeno M, Okazaki T, Takami Y, Nakayama T, Oshimura M (2004) Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat Cell Biol* 6: 784–791

Gene Ontology Consortium (2021) The gene ontology resource: enriching a GOld mine. *Nucleic Acids Res* 49: D325–D334

Grewal SIS, Jia S (2007) Heterochromatin revisited. *Nat Rev Genet* 8: 35–46

Grosswendt S, Kretzmer H, Smith ZD, Kumar AS, Hetzel S, Wittler L, Klages S, Timmermann B, Mukherji S, Meissner A (2020) Epigenetic regulator function through mouse gastrulation. *Nature* 584: 102–108

Guenther MG, Lawton LN, Rozovskaia T, Frampton GM, Levine SS, Volkert TL, Croce CM, Nakamura T, Canaani E, Young RA (2008) Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes Dev* 22: 3403–3408

Hannon GJ (2010) FASTX-Toolkit.

Heigwer F, Kerr G, Boutros M (2014) E-CRISP: fast CRISPR target site identification. *Nat Methods* 11: 122–123

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38: 576–589

Hyun K, Jeon J, Park K, Kim J (2017) Writing, erasing and reading histone lysine methylations. *Exp Mol Med* 49: e324

Jagannathan M, Cummings R, Yamashita YM (2018) A conserved function for pericentromeric satellite DNA. *eLife* 7: e34122

Janssen A, Colmenares SU, Karpen GH (2018) Heterochromatin: Guardian of the genome. *Annu Rev Cell Dev Biol* 34: 265–288

Johnson WL, Yewdell WT, Bell JC, McNulty SM, Duda Z, O'Neill RJ, Sullivan BA, Straight AF (2017) RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. *eLife* 6: e25299

Jones B, Su H, Bhat A, Lei H, Bajko J, Hevi S, Baltus GA, Kadam S, Zhai H, Valdez R *et al* (2008) The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. *PLoS Genet* 4: e1000190

Jonkers I, Kwak H, Lis JT (2014) Genome-wide dynamics of pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* 3: e02407

Kim SK, Jung I, Lee H, Kang K, Kim M, Jeong K, Kwon CS, Han YM, Kim YS, Kim D *et al* (2012) Human histone H3K79 methyltransferase DOT1L protein [corrected] binds actively transcribing RNA polymerase II to regulate gene expression. *J Biol Chem* 287: 39698–39709

Kim W, Choi M, Kim J-E (2014) The histone methyltransferase Dot1/DOT1L as a critical regulator of the cell cycle. *Cell Cycle* 13: 726–738

Kokavec J, Zikmund T, Savvulidi F, Kulvait V, Edelmann W, Skoultchi AI, Stopka T (2017) The ISWI ATPase Smarca5 (Snf2h) is required for proliferation and differentiation of hematopoietic stem and progenitor cells. *Stem Cells* 35: 1614–1623

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. *Nat Methods* 9: 357–359

Lesch BJ, Silber SJ, McCarrey JR, Page DC (2016) Parallel evolution of male germline epigenetic poising and somatic development in animals. *Nat Genet* 48: 888–894

Liao J, Szabó PE (2020) Maternal DOT1L is dispensable for mouse development. *Sci Rep* 10: 20636

Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27: 1739–1740

Liu H, Kim J-M, Aoki F (2004) Regulation of histone H3 lysine 9 methylation in oocytes and early pre-implantation embryos. *Development* 131: 2269–2280

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550

Lu J, Gilbert DM (2007) Proliferation-dependent and cell cycle–regulated transcription of mouse pericentric heterochromatin. *J Cell Biol* 179: 411–421

Millanes-Romero A, Herranz N, Perrera V, Iturbide A, Loubat-Casanovas J, Gil J, Jenuwein T, García de Herreros A, Peiró S (2013) Regulation of heterochromatin transcription by Snail1/LOXL2 during epithelial-to-mesenchymal transition. *Mol Cell* 52: 746–757

Mohan M, Herz HM, Takahashi YH, Lin C, Lai KC, Zhang Y, Washburn MP, Florens L, Shilatifard A (2010) Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). *Genes Dev* 24: 574–589

Müller S, Almouzni G (2017) Chromatin dynamics during the cell cycle at centromeres. *Nat Rev Genet* 18: 192–208

Nair L, Chung H, Basu U (2020) Regulation of long non-coding RNAs and genome dynamics by the RNA surveillance machinery. *Nat Rev Mol Cell Biol* 21: 123–136

Nakajima R, Sato T, Ogawa T, Okano H, Noce T (2017) A noncoding RNA containing a SINE-B1 motif associates with meiotic metaphase chromatin and has an indispensable function during spermatogenesis. *PLoS ONE* 12: e0179585

Ng HH, Feng Q, Wang H, Erdjument-Bromage H, Tempst P, Zhang Y, Struhl K (2002) Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and sir protein association. *Genes Dev* 16: 1518–1527

Novo CL, Tang C, Ahmed K, Djuric U, Fussner E, Mullin NP, Morgan NP, Hayre J, Sienerth AR, Elderkin S *et al* (2016) The pluripotency factor Nanog regulates pericentromeric heterochromatin organization in mouse embryonic stem cells. *Genes Dev* 30: 1101–1115

Ogawa A, Ogawa Y (2021) A quick Immuno-FISH protocol for detecting RNAs, proteins, and chromatin modifications. In *Functional Analysis of Long Non-Coding RNAs: Methods and Protocols*, Cao H (ed), pp 251–257. New York, NY: Springer US

Okada Y, Feng Q, Lin Y, Jiang Q, Li Y, Coffield VM, Su L, Xu G, Zhang Y (2005) hDOT1L links histone methylation to leukemogenesis. *Cell* 121: 167–178

Ooga M, Inoue A, Kageyama S, Akiyama T, Nagata M, Aoki F (2008) Changes in H3K79 methylation during preimplantation development in mice. *Biol Reprod* 78: 413–424

Ooga M, Suzuki MG, Aoki F (2013) Involvement of DOT1L in the remodeling of heterochromatin configuration during early preimplantation development in Mice1. *Biol Reprod* 89: 1–10

Peters AHFM, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schöfer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A *et al* (2001) Loss of the Suv39h histone Methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* 107: 323–337

Policarpi C, Crepaldi L, Brookes E, Nitarska J, French SM, Coatti A, Riccio A (2017) Enhancer SINEs link pol III to pol II transcription in neurons. *Cell Rep* 21: 2879–2894

Probst AV, Okamoto I, Casanova M, El Marjou F, Le Baccon P, Almouzni G (2010) A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. *Dev Cell* 19: 625–638

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44: W160–W165

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140

Rudert F, Bronner S, Garnier JM, Dollé P (1995) Transcripts from opposite strands of gamma satellite DNA are differentially expressed during mouse development. *Mamm Genome* 6: 76–83

Saksouk N, Simboeck E, Dejardin J (2015) Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* 8: 3

Santenard A, Ziegler-Birling C, Koch M, Tora L, Bannister AJ, Torres-Padilla M-E (2010) Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat Cell Biol* 12: 853–862

Sarkar D (2008) *Lattice: multivariate data visualization with R*. New York, NY: Springer

Seczynska M, Bloor S, Cuesta SM, Lehner PJ (2022) Genome surveillance by HUSH-mediated silencing of intronless mobile elements. *Nature* 601: 440–445

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG *et al* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343: 84–87

Shestakova EA, Mansuroglu Z, Mokrani H, Ghinea N, Bonnefoy E (2004) Transcription factor YY1 associates with pericentromeric γ-satellite DNA in cycling but not in quiescent (G0) cells. *Nucleic Acids Res* 32: 4390–4399

Shirai A, Kawaguchi T, Shimojo H, Muramatsu D, Ishida-Yonetani M, Nishimura Y, Kimura H, Nakayama J, Shinkai Y (2017) Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. *eLife* 6: e25317

Smit A, Hubley R, Green P (2013) RepeatMasker Open-4.0.

Stauffer F, Weiss A, Scheufler C, Möbitz H, Ragot C, Beyer KS, Calkins K, Guthy D, Kiffe M, Van Eerdenbrugh B *et al* (2019) New potent DOT1L inhibitors for in vivo evaluation in mouse. *ACS Med Chem Lett* 10: 1655–1660

Steger DJ, Lefterova MI, Ying L, Stonestrom AJ, Schupp M, Zhuo D, Vakoc AL, Kim JE, Chen J, Lazar MA *et al* (2008) DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol Cell Biol* 28: 2825–2839

Stopka T, Skoultchi AI (2003) The ISWI ATPase Snf2h is required for early mouse development. *Proc Natl Acad Sci USA* 100: 14097–14102

Stopka T, Blafkova J, Zakova D, Fuchs O, Cmejla R, Necas E, Jelinek J, Zivny J (2000) Cloning and expression of murine hematopoietic specific chromatin remodeling gene SMARCA5. *Exp Hematol* 28: 119

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102: 15545–15550

Tashiro K, Teissier A, Kobayashi N, Nakanishi A, Sasaki T, Yan K, Tarabykin V, Vigier L, Sumiyama K, Hirakawa M *et al* (2011) A mammalian conserved element derived from SINE displays enhancer properties recapitulating Satb2 expression in early-born Callosal projection neurons. *PLoS ONE* 6: e28497

Ting DT, Lipson D, Paul S, Brannigan BW, Akhavanfard S, Coffman EJ, Contino G, Deshpande V, Iafrate AJ, Letovsky S *et al* (2011) Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* 331: 593–596

Valgardsdottir R, Chiodi I, Giordano M, Rossi A, Bazzini S, Ghigna C, Riva S, Biamonti G (2008) Transcription of satellite III non-coding RNAs is a general stress response in human cells. *Nucleic Acids Res* 36: 423–434

Vargova J, Vargova K, Skoultchi AI, Stopka T (2009) Nuclear localization of ISWI ATPase Smarca5 (Snf2h) in mouse. *Front Biosci* 1: 553–559

Velazquez Camacho O, Galan C, Swist-Rosowska K, Ching R, Gamalinda M, Karabiber F, De La Rosa-Velazquez I, Engist B, Koschorz B, Shukeir N *et al* (2017) Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. *eLife* 6: e25293

Veloso A, Kirkconnell KS, Magnuson B, Biewen B, Paulsen MT, Wilson TE, Ljungman M (2014) Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res* 24: 896–905

Wickham H (2016) *ggplot2: elegant graphics for data analysis*. New York, NY: Springer-Verlag

Wood K, Tellier M, Murphy S (2018) DOT1L and H3K79 methylation in transcription and genomic stability. *Biomolecules* 8: 11

Wu A, Zhi J, Tian T, Cihan A, Cevher MA, Liu Z, David Y, Muir TW, Roeder RG, Yu M (2021) DOT1L complex regulates transcriptional initiation in human erythroleukemic cells. *Proc Natl Acad Sci USA* 118: e2106148118

Yi Q, Chen Q, Liang C, Yan H, Zhang Z, Xiang X, Zhang M, Qi F, Zhou L, Wang F (2018) HP1 links centromeric heterochromatin to centromere cohesion in mammals. *EMBO Rep* 19: e45484

Yu W, Chory EJ, Wernimont AK, Tempel W, Scopton A, Federation A, Marineau JJ, Qi J, Barsyte-Lovejoy D, Yi J *et al* (2012) Catalytic site remodelling of the DOT1L methyltransferase by selective inhibitors. *Nat Commun* 3: 1288

Yu G, Wang L-G, He Q-Y (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31: 2382–2383

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W *et al* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9: R137

Zikmund T, Paszekova H, Kokavec J, Kerbs P, Thakur S, Turkova T, Tauchmanova P, Greif PA, Stopka T (2020) Loss of ISWI ATPase SMARCA5 (SNF2H) in acute myeloid leukemia cells inhibits proliferation and chromatid cohesion. *Int J Mol Sci* 21: E2073

Zofall M, Grewal SIS (2006) Swi6/HP1 recruits a JmjC domain protein to facilitate transcription of heterochromatic repeats. *Mol Cell* 22: 681–692