Distinct contributions of DNA methylation and histone acetylation to the genomic occupancy of transcription factors

Martin Cusack,[1] Hamish W. King,[2] Paolo Spingardi,[1] Benedikt M. Kessler,[3] Robert J. Klose,[2] and Skirmantas Kriaucionis[1]

[1]Ludwig Institute for Cancer Research, University of Oxford, Oxford, OX3 7DQ, United Kingdom; [2]Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, United Kingdom; [3]Target Discovery Institute, University of Oxford, Oxford, OX3 7FZ, United Kingdom

Epigenetic modifications on chromatin play important roles in regulating gene expression. Although chromatin states are often governed by multilayered structure, how individual pathways contribute to gene expression remains poorly understood. For example, DNA methylation is known to regulate transcription factor binding but also to recruit methyl-CpG binding proteins that affect chromatin structure through the activity of histone deacetylase complexes (HDACs). Both of these mechanisms can potentially affect gene expression, but the importance of each, and whether these activities are integrated to achieve appropriate gene regulation, remains largely unknown. To address this important question, we measured gene expression, chromatin accessibility, and transcription factor occupancy in wild-type or DNA methylation-deficient mouse embryonic stem cells following HDAC inhibition. We observe widespread increases in chromatin accessibility at retrotransposons when HDACs are inhibited, and this is magnified when cells also lack DNA methylation. A subset of these elements has elevated binding of the YY1 and GABPA transcription factors and increased expression. The pronounced additive effect of HDAC inhibition in DNA methylation–deficient cells demonstrates that DNA methylation and histone deacetylation act largely independently to suppress transcription factor binding and gene expression.

[Supplemental material is available for this article.]

The most abundant DNA modification in mammals is 5-methylcytosine (5mC) which is found predominantly within CpG dinucleotides (Lister et al. 2009; Stadler et al. 2011). DNA methylation is deposited and maintained through the concerted activity of three essential DNA methyltransferases (DNMT1, DNMT3A, DNMT3B) (Goll and Bestor 2005) and plays important roles in transcriptional repression. For example, DNA methylation is known to regulate imprinted genes (Kaneda et al. 2004), some tissue-restricted genes (De Smet et al. 1999; Maatouk et al. 2006; Borgel et al. 2010), evolutionarily young retrotransposons (Davis et al. 1989; Walsh et al. 1998; Castro-Diaz et al. 2014), genes on the inactive X Chromosome (Beard et al. 1995; Hansen et al. 1996), and certain tumor suppressor genes in cancer cells (Toyota et al. 1999; Rhee et al. 2002; Robert et al. 2003).

Although 60%–80% of CpGs in the genome are methylated, DNA methylation is absent or reduced in regions bound by transcription factors such as in CpG islands, gene promoters, and distal regulatory elements (Stadler et al. 2011; Hon et al. 2013; Ziller et al. 2013). This observation led to the proposal that DNA methylation may limit transcription factor occupancy. In support of this model, in vitro experiments have revealed that 5mC can affect the affinity of most DNA-binding proteins that contain CpG dinucleotides in their binding motif (Hu et al. 2013; Spruijt et al. 2013; Kribelbauer et al. 2017; Yin et al. 2017). Furthermore, certain transcription factors (TFs), including NRF1 and CTCF, have been shown to have altered binding profiles in cells where DNA methylation is reduced or absent (Domcke et al. 2015; Maurano et al. 2015; Yin et al. 2017).

In addition to directly altering the binding of transcription factors, methylated CpG is recognized by methyl binding domain-containing (MBD) and zinc-finger methyl-cytosine binding proteins (Hendrich and Bird 1998; Lopes et al. 2008; Quenneville et al. 2011). MBD proteins interact with histone deacetylase (HDAC) enzymes and promote the formation of transcriptionally repressive chromatin environments (Nan et al. 1998; Zhang et al. 1999; Kokura et al. 2001; Yoon et al. 2003) via the removal of acetyl groups from lysine residues in histones.

Despite the active recruitment of HDACs to methylated CpGs, the extent to which DNA methylation relies on HDAC activity to affect gene expression is still poorly understood. HDAC inhibition, or disruption of the interaction between MBD proteins and the HDAC-containing complexes, has been shown to alleviate transcriptional silencing mediated by DNA methylation in reporter gene assays (Jones et al. 1998; Nan et al. 1998; Ng et al. 1999; Lyst et al. 2013). Nevertheless, at other endogenous genes, inhibiting HDACs did not recapitulate the effects on gene expression that manifest when DNA methylation is removed (Cameron et al. 1999; Coffee et al. 1999; Brunmeir et al. 2010; Reichmann et al. 2012). Therefore, although it is clear that DNA methylation is capable of altering TF-DNA interactions and recruiting HDACs, how these distinct mechanisms contribute to DNA methylation-dependent repression at the genome scale remains unknown.

Failure to inactivate transcription and mobility of retrotransposons has been linked to genome instability in cancer (Carreira et al. 2014; Tubio et al. 2014; Scott and Devine 2017; Schauer et al. 2018) and the manifestation of neurological diseases (Coufal et al. 2011). DNA methylation is an important contributor to silencing of retrotransposons, where it acts in concert with histone deacetylation and histone (H3K9) methylation (Bourc'his and Bestor 2004; Karimi et al. 2011; Rowe et al. 2013; Sharif et al. 2016). Whether chromatin mechanisms are redundant or collaborative while acting on different classes of retrotransposons is incompletely understood.

In this study, we aimed to determine individual and collaborative roles of DNA methylation and HDAC activity. To this purpose, we disrupted both processes, one at a time or in combination, in mouse embryonic stem cells and measured the impact on chromatin accessibility, transcription factor occupancy, and gene regulation.

Results

Disruption of HDAC activity and DNA methylation in mouse embryonic stem cells

To determine the contribution of DNA methylation and the activity of HDAC enzymes on transcriptional regulation, we treated wild-type (J1) or *Dnmt3a/Dnmt3b/Dnmt1* triple knockout (DNMT.TKO) mouse embryonic stem cells (Tsumura et al. 2006) with the HDAC inhibitor trichostatin A (TSA) (Fig. 1A). The chosen dose of TSA had minimal effects on the morphology or cell cycle of mESCs

(Supplemental Fig. S1A) but yielded a robust increase in global histone acetylation as detected by immunoblotting (Fig. 1B). There were no significant changes in global 5mC levels between TSA- and DMSO (solvent)-treated cells (Supplemental Fig. S1B).

To examine histone acetylation in more detail, we performed native calibrated ChIP-seq for acetylated histone H3 (H3ac). TSA treatment led to a global increase in H3ac in wild-type and DNMT.TKO cells both at previously hyperacetylated sites (ChIP-seq peaks) and the remaining genomic space (Fig. 1C–E; Supplemental Fig. S1C). The increase in histone acetylation at repetitive regions and unannotated genomic loci following HDAC inhibition was greater in DNMT.TKO compared to wild-type cells (Fig. 1E). This observation coincides with high DNA methylation levels in the aforementioned regions in WT cells, whereas hyperacetylated sites contain low median DNA methylation (Supplemental Fig. S1D). This additive effect of DNA methylation loss and HDAC inhibition on histone acetylation in these regions could be explained by enhanced recruitment of histone acetyltransferases (HATs), potentiated by DNA-binding proteins with specificity to unmodified CpGs (e.g., CxxC-domain proteins). Alternatively, the lack of DNA methylation may disrupt the recruitment of HDACs by MBD-domain proteins, resulting in a lower capacity for compensatory mechanisms to deacetylate histones or prevent histone acetylation.

In summary, HDAC inhibition using TSA induces histone acetylation broadly across the genome with no detectable change in DNA methylation, enabling us to evaluate the extent and the mode by which these two mechanisms contribute to chromatin function.
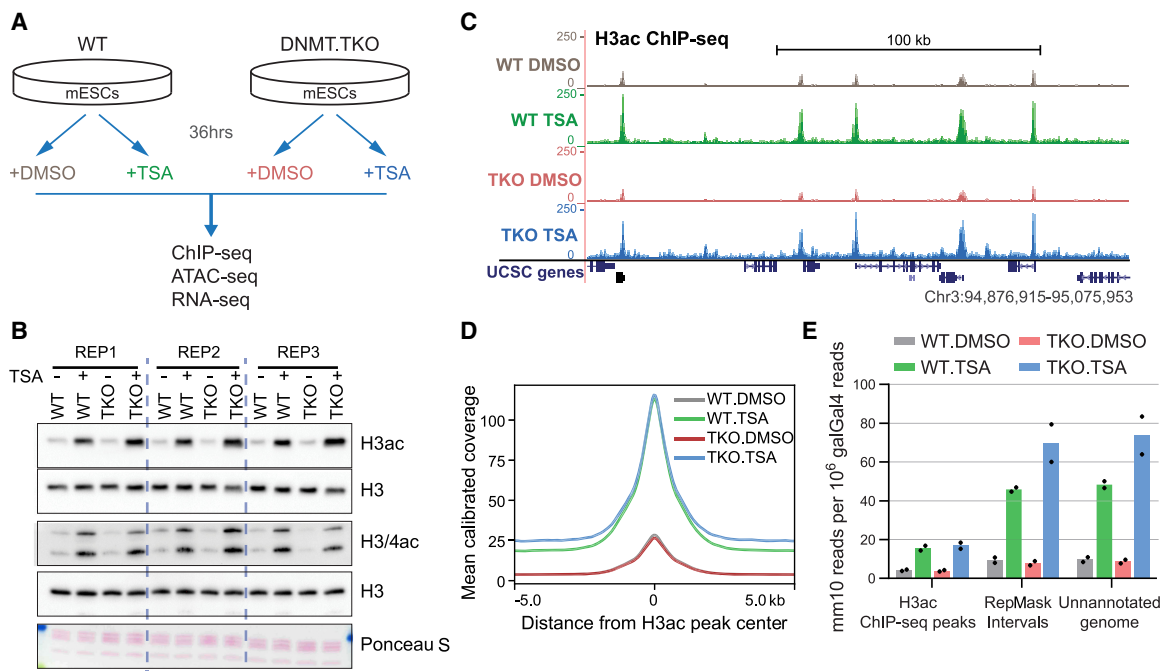

Figure 1. Disruption of HDAC activity and DNA methylation in mouse embryonic stem cells. (*A*) Schematic of the experimental approach used in this study. (*B*) Immunoblot analysis of global histone H3 and H4 acetylation levels after 36-h TSA treatment in mESCs. Samples were derived from three biological replicate experiments. (*C*) Representative snapshot of genomic region showing H3ac ChIP-seq read coverage calibrated to the spike-in galGal4 genome. (*D*) Metaplot showing the average calibrated H3ac ChIP-seq signal from DMSO- or TSA-treated wild-type and DNMT.TKO cells in 10-kb regions surrounding the center of ChIP-seq peaks (*n* = 28608). (*E*) Distribution of H3ac ChIP-seq reads within genomic intervals separated into three mutually exclusive categories. For every sample, the number of mm10 reads overlapping each category was divided by the total number of reads mapping to the galGal4 genome. Data are represented as the mean; points indicate the values for two biological replicates.

DNA methylation and HDAC activity have distinct contributions to the chromatin accessibility landscape

To compare the impact of DNA methylation and histone acetylation on chromatin function, we profiled chromatin accessibility in wild-type and DNMT.TKO cells in the absence or presence of TSA using ATAC-seq (Fig. 1A).

Employing these data, we first defined a set of 83,395 transposase hypersensitive sites (THSs) that were called as significantly accessible in at least one experimental condition. Sample-to-sample correlations and a principle component analysis both clustered biological replicate samples together, demonstrating that treatment and genotype induced reproducible changes in accessibility (Supplemental Fig. S2A–C). We identified THSs that display differential accessibility by examining the effect of DNA methylation loss, or that of HDAC inhibition in either wild-type or DNA methylation-deficient cells (Fig. 2A). In all comparisons, regions that gained accessibility had a lower CpG and GC content than unchanged ones, which was related to the reduced presence of differential THSs at promoters (Supplemental Fig. S2D–G).

We first focused on comparing DNMT.TKO to WT DMSO-treated cells in order to characterize the chromatin accessibility changes that occur following the loss of 5mC. We identified 6427 THS regions that significantly gained accessibility and 7135 that significantly lost accessibility in the absence of DNA methylation (Fig. 2A, left). Significant increases in ATAC-seq signal occurred preferentially at loci that showed low accessibility in wild-type cells, whereas significant decreases were more broadly detected at regions with intermediate levels of wild-type accessibility, a finding that was consistent when using more stringent FDR thresholds for calling significant differences (Supplemental Fig. S2H). Differential THS regions account for ~16% of all sites that were detected as being accessible in wild-type or DNMT.TKO mES cells when using a FDR threshold of 0.01, demonstrating that DNA methylation has an important role in maintaining accessibility patterns. However, complete loss of DNA methylation in mouse ES cells does not produce widespread alterations, in agreement with similar results obtained using DNase-seq (Domcke et al. 2015).

To assess the fraction of differential accessibility events that occur as a direct consequence of a loss in DNA methylation, we made use of public whole-genome bisulfite sequencing (WGBS) data generated from E14 mESCs grown under the same growth conditions (Habibi et al. 2013). Similar to previous observations (Domcke et al. 2015), we found that gains in accessibility in DNMT.TKO cells frequently occur at regions that are methylated in WT cells (Supplemental Fig. S2I, top right quadrant). Specifically, 56% of DNMT.TKO loci that significantly gain in accessibility have a mean methylation value above 60% in WT cells (Supplemental Fig. S2I). These represent loci at which DNA methylation is likely to play a role in restraining the formation of a Tn5 accessible state, potentially through interference with transcription factor binding. On the other hand, the relationship between accessibility reduction in DNMT.TKO cells and DNA methylation is less straightforward (Supplemental Fig. S2I). Only 2.8% of the regions displaying significantly reduced accessibility in DNMT.TKO cells have high levels of methylation in wild-type cells (≥60% methylation), making it more difficult to attribute these differences in chromatin accessibility to changes in methylation.

Screening a collection of 264 known vertebrate transcription factor motifs (HOMER database) (Heinz et al. 2010) for enrichment at DNMT.TKO-specific THSs, we identified the NRF1 motif as being the most enriched, consistent with the findings of Domcke et al. (2015) (Supplemental Fig. S2J; Supplemental Table S5). Similar to their results, sequence motifs recognized by the NFY complex, ETS and basic Helix-Loop-Helix (bHLH) transcription factors were also enriched in DNMT.TKO THSs. In vitro affinity measurements and locus-specific ChIP evidence suggest that the binding of these transcription factors to their cognate recognition sites is sensitive to DNA methylation (Hu et al. 2017; Yin et al. 2017), supporting the idea that accessibility gains occur in the absence of DNA methylation due to unimpeded binding of these proteins.

We next examined the effects of HDAC inhibition on the chromatin accessibility landscape of wild-type and DNMT.TKO mES cells. Few loci displayed significant accessibility gains in TSA- versus DMSO-treated cells, relative to the numbers identified when comparing DNMT.TKO to WT. In fact, we observed that the vast majority of regions display reduced accessibility in cells upon HDAC inhibition, in both WT and DNMT.TKO cells (Fig. 2A; Supplemental Fig. S2H). The effect was more pronounced in cells depleted of DNA methylation, in which 38% of all THSs showed a significant reduction in ATAC-seq signal when treated with TSA. Congruently, THSs in TSA-treated cells are smaller, as seen when aggregating the signal across all THSs or at individual sites in the genome (Fig. 2B,C). We further noted that genome accessibility changes at THSs were similar between TSA-treated WT and DNMT.TKO cell lines (Spearman's correlation coefficient = 0.649) (Fig. 2D). In contrast, there was a considerable discrepancy in the changes occurring with either DNA methylation loss or TSA treatment (Spearman's correlation coefficient = 0.187) (Fig. 2E). Consequently, the main difference in the response to HDAC inhibition in either the presence or absence of DNA methylation is the magnitude of the changes.

We then examined the relationship between TSA-induced changes in H3 acetylation and genome accessibility (Supplemental Fig. S3A). THSs which lose or display no change after TSA treatment had similar gains in H3 acetylation, whereas THSs which gained accessibility demonstrated elevated acetylation (Supplemental Fig. S3A, bottom panel). Discordance between loss of accessibility and H3 acetylation prompted us to examine the possibility that the observed loss of accessibility might be an indirect effect and subject to a different interpretation. Reduced signal at THS intervals in TSA-treated samples indicates that DNA fragments originating from these regions are underrepresented as a fraction of the entire ATAC-seq library, relative to their fraction in untreated samples. Concurrently, a larger proportion of fragments mapping to the inaccessible compartment of the genome (i.e., the background) are sequenced. We envision two possible explanations for this observed reduction of ATAC-seq signal at THS intervals.

The first possibility is that, upon TSA treatment, chromatin accessibility is reduced at THSs, meaning that fewer reads would be obtained from within these intervals. As a consequence of sequencing a fixed number of DNA fragments, we would thus observe a relative increase in accessibility signal from the genomic compartment outside THSs (the "inaccessible" compartment). According to this hypothesis, the additional reads sequenced from the inaccessible compartment would be distributed evenly throughout the remaining genomic space. The second possibility is that TSA treatment results in elevated accessibility in genomic regions outside of THSs. The relative reduction in the ATAC-seq signal at THSs then becomes a secondary consequence of elevated accessibility outside THSs. This scenario would be supported if we were to identify an enrichment in accessibility at particular subdivisions of the inaccessible compartment.
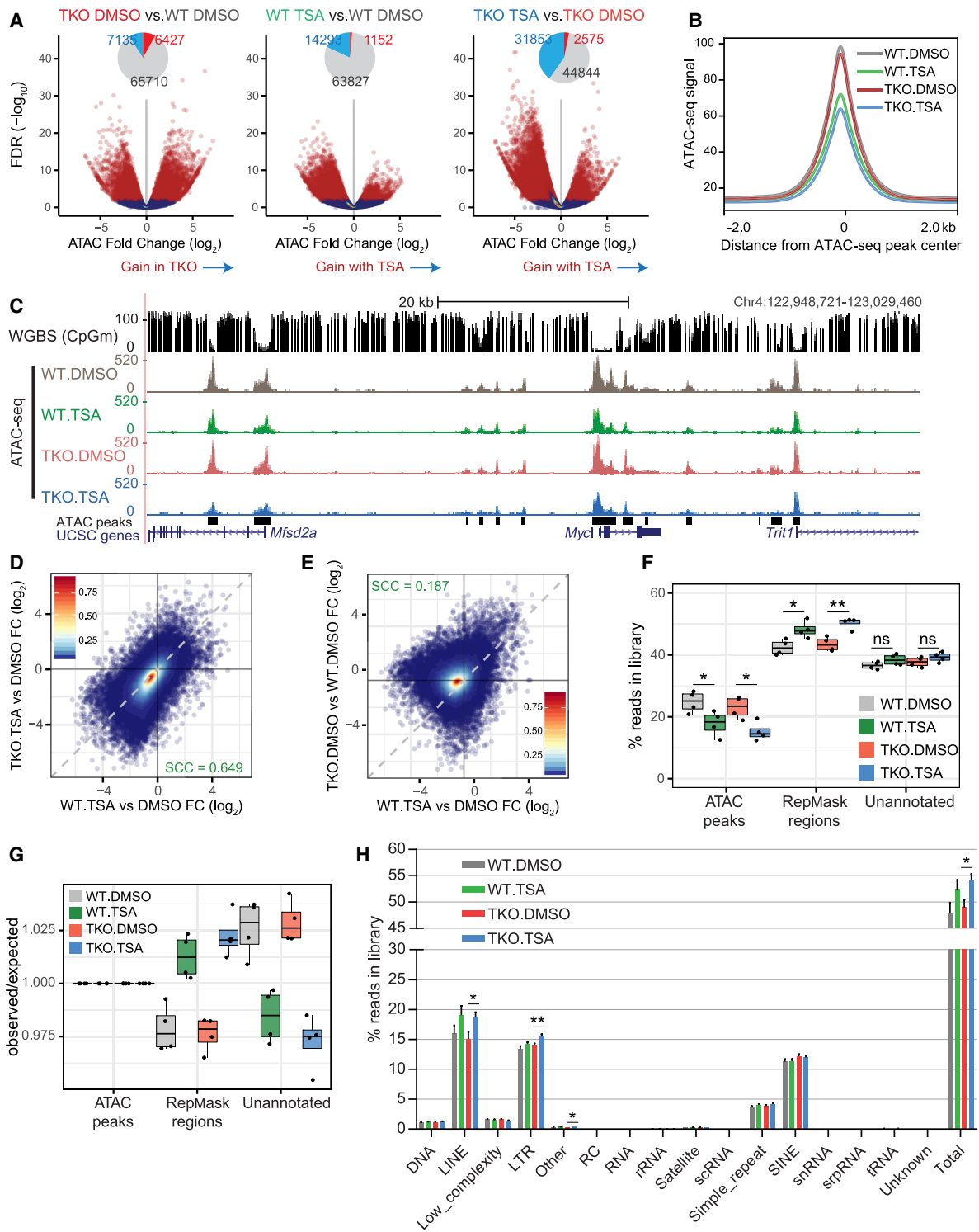
Figure 2. DNA methylation and HDAC activity have distinct contributions to the chromatin accessibility landscape. (*A*) Volcano plots representing the false discovery rate (FDR) and fold change values obtained through pairwise differential analyses of ATAC-seq signal at 83,395 THSs. Regions with significantly differential accessibility are shown in red on the scatterplots and their numbers are summarized in the form of pie charts (light blue = significant decrease; red = significant increase). (*B*) Metaplot showing the average ATAC-seq signal from DMSO- or TSA-treated wild-type and DNMT.TKO cells in 4-kb regions surrounding the center of THSs ($n = 83,395$). (*C*) Representative UCSC Genome Browser snapshot showing CpG methylation levels and ATAC-seq read coverage. (*D*) Scatterplot comparing the fold change in ATAC-seq signal following TSA treatment in DNMT.TKO versus wild-type cells. (*E*) Scatterplot comparing the fold change in ATAC-seq signal after loss of DNA methylation versus the change seen following TSA treatment in wild-type cells at 83,395 THSs. In *D* and *E*, the dashed line has a slope of 1 and intercept of 0. Colors indicate density of points per graph area. SCC = Spearman's correlation coefficient. (*F*) Box plots summarizing the percentage of reads from each ATAC-seq library that map to intervals split into three mutually exclusive categories. Two-tailed Student's *t*-tests; (*) *P*-value < 0.05, (**) *P*-value < 0.01, (ns) nonsignificant (*P*-value > 0.05). (*G*) Box plots summarizing the distribution of reads from each ATAC-seq library that map to intervals split into three mutually exclusive categories, relative to the distribution expected by chance, that is, if non-THS reads were shuffled randomly within the genomic space outside of THSs. Values shown in *F* are divided by those in Supplemental Figure S3B. (*H*) Distribution of ATAC-seq reads across different classes of repetitive elements as a percentage of the total library size. Data are represented as mean + SD. Two-tailed Student's *t*-tests; (*) *P*-value < 0.05, (**) *P*-value < 0.01, nonsignificant differences are not indicated.

To explore the above-mentioned interpretations, we calculated the fraction of reads in the ATAC-seq libraries that overlapped with three categories of genomic regions: transposase hyperaccessible regions (THSs, 83,395 intervals covering 55 Mb, 2.03% of the mm10 genome); genomic intervals consisting of repetitive elements (RepMask intervals, excluding sequences that overlap THSs, 1.19 Gb, 43.39% of mm10 genome); and the remaining genomic space (unannotated regions, 54.58%) (Fig. 2F). This analysis shows that gains in accessibility in TSA-treated samples occur preferentially at repetitive elements compared to unannotated regions (Fig. 2F), supporting a contribution from the second scenario outlined above. The increase in coverage observed at repeats is larger than that obtained if the same number of non-THS sequencing reads are randomly shuffled into the combined "RepMask" and unannotated genomic compartments (Fig. 2F; Supplemental Fig. S3B,C). Furthermore, the distinct accessibility changes observed at repeat regions are unlikely to be solely a consequence of differences in the induction of histone acetylation, as the H3ac levels in both compartments are equivalent following HDAC inhibition (Supplemental Fig. S3D). These results argue against the first proposed mechanism being responsible for the reduction in ATAC-seq signal at THSs in TSA-treated samples. Rather, our findings suggest that genome-wide hyperacetylation following HDAC inhibition encourages gains in accessibility at repeat elements. Although gains in accessibility occur preferentially at repetitive elements in TSA-treated cells, any additional increase in accessibility globally cannot be excluded using our experimental setup.

We next subdivided the repeat regions further by examining the changes in accessibility occurring at various classes of repetitive elements (Fig. 2H). We found that the fraction of reads mapping to LINE and LTR elements increased upon TSA treatment, although this was significant in DNMT.TKO but not in WT cells. Differential analysis of the ATAC-seq coverage between samples revealed that there were significant increases in the number of ATAC-seq reads mapping to 33 of the 132 LINE-1 (or L1) element subfamilies in TSA- versus DMSO-treated wild-type cells (Supplemental Fig. S3E; Supplemental Table S3). In DNMT.TKO cells, significant gains were detected for an additional five subfamilies (Supplemental Fig. S3F). The most significant gains in accessibility in TSA-treated cells were observed at the youngest LINE-1 subfamilies (Supplemental Fig. S3G), suggesting that the presence of intact transcription factor binding sites may be important in permitting accessibility. Among the LTR repeats, accessibility gains in DNMT.TKO cells predominantly occurred at ERVK-type elements (Supplemental Fig. S3H). HDAC inhibition potentiated these gains while also increasing accessibility at some ERVL and ERVL-MaLR-type repeats. Together, our ATAC-seq analysis indicates that a subset of normally methylated loci increases in accessibility in the absence of DNA methyltransferases, likely representing novel transcription factor binding sites. A distinct effect was observed upon HDAC inhibition, where the resulting genome-wide hyperacetylation led to an increase in accessibility outside of previously hyperaccessible sites (THSs), particularly at a subset of LINE elements and LTR retrotransposons.

DNA methylation and HDAC activity can modulate transcription factor occupancy

NRF1 and NFY transcription factor motifs were enriched in THSs that appeared in DNMT.TKO cells, whereas HDAC inhibition resulted in the appearance of THSs containing DNA sequences recognized by the AP-2 complex and TRP53 (Supplemental Figs. S2J, S3I). Elevated chromatin accessibility may be a result of abnormal TF binding (Domcke et al. 2015; Maurano et al. 2015); however, it is not known whether our chromatin perturbations cause more widespread TF binding without inducing transposase hypersensitivity. Thus, we decided to directly measure how DNA methylation and histone acetylation affect the genomic distribution of selected transcription factors. We performed ChIP-seq in WT and DNMT.TKO mESCs in the presence or absence of TSA, to map the genome-wide occupancy of five transcription factors (GABPA, MAX, NRF1, SP1, and YY1) that possess distinct DNA-binding domains and are known to bind cognate sequences containing CpG dinucleotides (Fig. 3A).

We identified 4735, 19,236, 3102, 6666, and 6055 regions as being reproducibly occupied by GABPA, MAX, NRF1, SP1, and YY1 in at least one experimental condition, respectively (Supplemental Table S2). Biological replicates clustered together (Supplemental Fig. S4A) and the DNA motifs known to be bound by these transcription factors were the most highly enriched within the identified ChIP-seq peaks relative to the entire collection of motifs in the HOMER database, providing assurance of the specificity of these ChIP assays (Supplemental Fig. S4B). Moreover GABPA, NRF1, SP1, and YY1 protein abundance is not substantially affected in our examined conditions (Supplemental Fig. S4C). A small reduction of SP1 and NRF1 is observed after treatment with TSA; however, this would not be able to explain the predominant gains in occupancy of these TFs (Fig. 3B).

For all five transcription factors, loss of DNA methylation or HDAC inhibition resulted in increased transcription factor occupancy (Fig. 3B). However, the relative impact of these two alterations varied depending on the individual protein. In accordance with our ATAC-seq analysis (Supplemental Fig. S2J), NRF1 gained novel binding sites in DNMT.TKO compared to wild-type cells (Fig. 3B). As these regions are methylated in wild-type cells (Fig. 4A), this suggests that DNA methylation restricts NRF1 binding at certain genomic loci, in line with previous reports (Kumari and Usdin 2001; Domcke et al. 2015). On the other hand, HDAC inhibition did not have a substantial effect on the genome-wide occupancy of NRF1, with relatively few sites showing differential binding upon TSA treatment in wild-type or DNMT.TKO cells (Fig. 3B; Supplemental Fig. S4D).

A small number of differential binding sites were identified for SP1 when comparing TSA-treated or DNMT.TKO cells to wild-type cells, suggesting that this transcription factor's occupancy is mostly unaltered by global loss of DNA methylation or genome-wide hyperacetylation alone (Fig. 3B).

Occupancy of GABPA, MAX, and YY1 can be modulated by both DNA methylation and histone acetylation (Fig. 3B). At some binding sites, either DNA methylation or HDAC activity alone seem to be sufficient to restrict binding of these factors (Fig. 3C; Supplemental Fig. S5A,D vs. Fig. 3D; Supplemental Fig. S5B,E), whereas at other loci the combination of both appears to be responsible for impeding occupancy (Fig. 3E; Supplemental Fig. S5C,F). Binding sites newly acquired in DNMT.TKO cells predominantly harbor the respective TF binding motifs, and these are methylated in WT cells (Fig. 4A,B). Furthermore, treatment with TSA resulted in elevated histone acetylation in the novel peaks, reinforcing the model that occupancy gains are a direct result of chromatin perturbations (Supplemental Fig. S4E).

Although all five transcription factors predominantly occupy gene promoter regions, the majority of ChIP-seq peaks that showed significant increases in occupancy upon DNA methylation loss or HDAC inhibition were situated at promoter distal
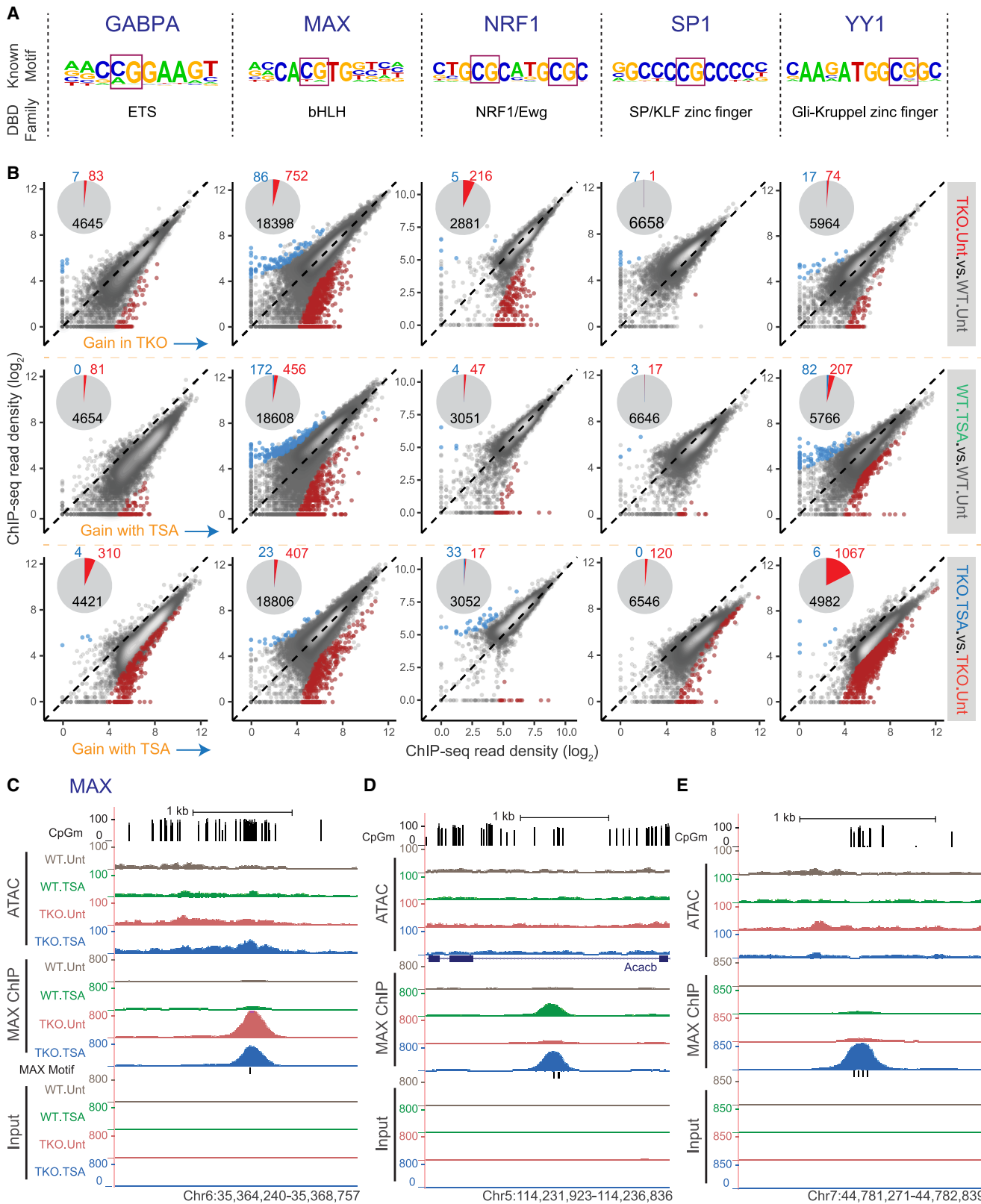
Figure 3. DNA methylation and HDAC activity can modulate transcription factor occupancy. (*A*) For each transcription factor, the position weight matrix (PWM) of its known motif and its DNA-binding domain (DBD) type are shown. (*B*) Pairwise comparisons of normalized GABPA, MAX, NRF1, SP1, or YY1 ChIP-seq signal for all identified occupancy peaks. Regions with significantly differential occupancy are colored on the scatterplots, and their numbers are summarized in the form of pie charts (light blue = significant decrease; red = significant increase). ChIP-seq signal from three biological replicate samples was averaged. (*C–E*) Representative UCSC Genome Browser snapshots showing CpG methylation levels, ATAC-seq, MAX ChIP-seq, and Input ChIP-seq read coverage.
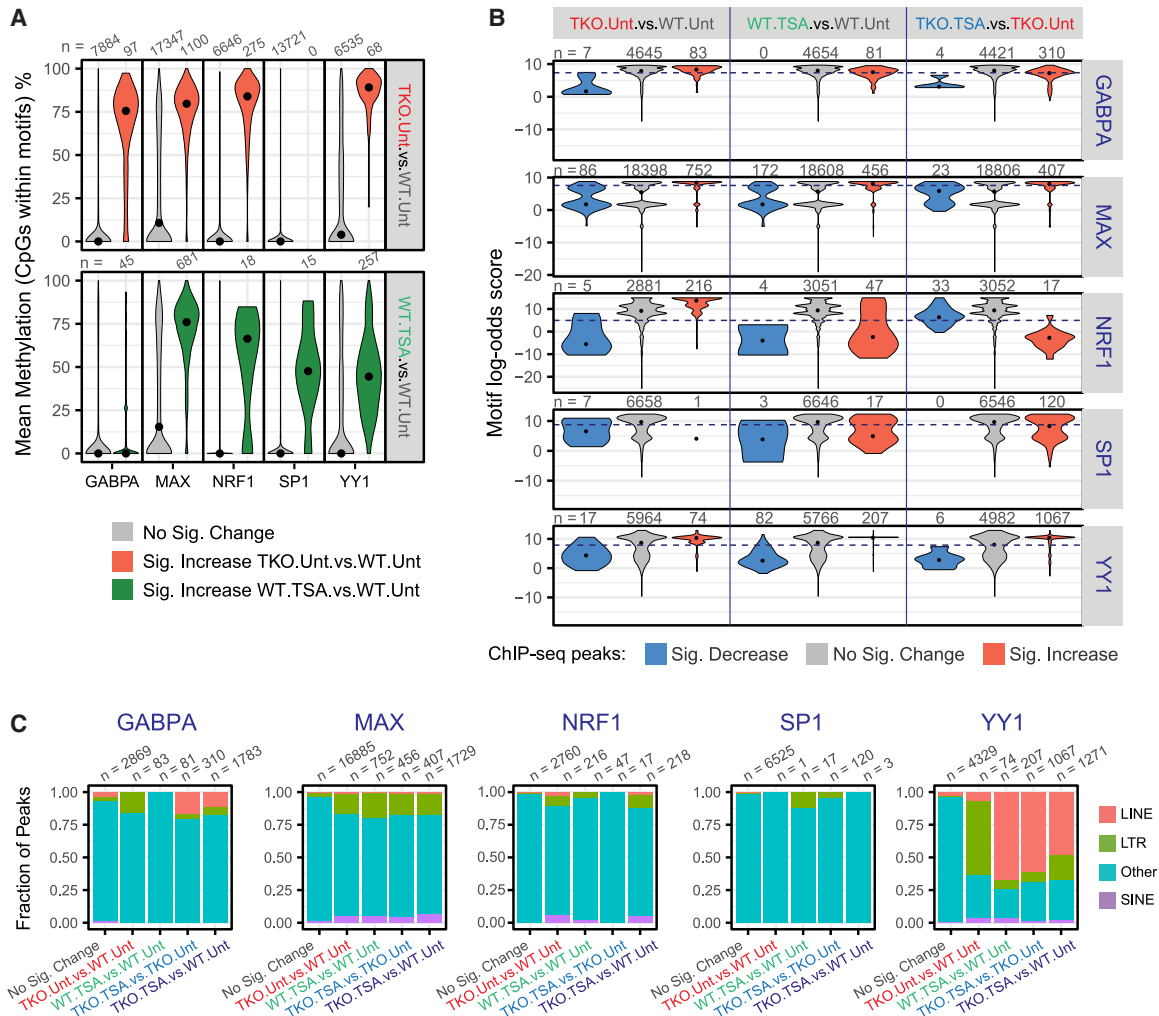
Figure 4. Characteristics of transcription factor (TF) binding sites. (*A*) Distribution of CpG methylation levels within the TF motifs that underlie ChIP-seq peaks. For each TF, we isolated all sequences that were located within peak regions that matched the relevant position weight matrix and determined the methylation status of the CpG nucleotides. CpG sites are grouped according to their differential occupancy. Black dots indicate the median. (*B*) Distribution of maximum log-odds scores found at different subsets of ChIP-seq peaks. Within each interval of the GABPA, MAX, NRF1, SP1, and YY1 peak-sets, the sequence most similar to the respective PWM was identified and its log-odds score, indicative of the deviation from the consensus sequence, was plotted. Black dots indicate the median. The dashed lines indicate the log-odds score threshold above which a sequence is said to match the PWM. (*C*) Fraction of ChIP-seq peaks that overlap an annotated transposable element. ChIP-seq peaks are classified based on their differential occupancy.

genomic sites (Supplemental Fig. S6A). The exception to this is GABPA, for which TSA treatment resulted in occupancy of loci proximal to TSSs. This is likely related to the observation that TSA treatment relocates some of GABPA exclusively to nonmethylated loci (Fig. 4A), often found within CpG islands near TSSs.

We further examined whether novel intergenic ChIP-seq peaks overlapped with transposable elements (LTRs, LINEs, and SINEs). Most strikingly, over half of DNMT.TKO or TSA-specific YY1 ChIP-seq peaks overlapped with transposable elements (Fig. 4C). Particularly, 652 out of 1067 YY1 binding sites acquired in the absence of both DNA methylation and HDAC activity locate to LINE retrotransposons. Multiple LINE subfamilies exist in the mouse genome that are classified by the sequence of the monomeric units that are tandemly repeated to make up their 5′ UTR promoters (Goodier et al. 2001; Sookdeo et al. 2013; Zhou and Smith 2019). Only full-length members of the youngest L1 subfamilies, containing A-, T_F-, or G_F-type promoters are capable of

retrotransposition, with the T_F subtype being the most active in somatic cells and the germline (Akagi et al. 2008; Richardson et al. 2017; Schauer et al. 2018). Of the LINE-1 elements bound by YY1 in TKO.TSA samples, 537 belonged to the L1Md_T subfamily (82%) and 54 to the L1Md_Gf type (8.3%) (Supplemental Fig. S6B). Both of these L1 subtypes are known to have YY1 binding motifs in their promoter sequences that are important for retrotransposition (DeBerardinis and Kazazian 1999; Goodier et al. 2001; Athanikar et al. 2004; Sanchez-Luque et al. 2019). Indeed, 525 of the L1Md_T and 25 of the L1Md_Gf elements that showed increased YY1 occupancy in the absence of DNA methylation and HDAC activity were full-length members of their families (>5000 bp and >4000 bp, respectively) and most contained at least one YY1-containing monomeric units within their 5′UTR regions (Supplemental Figs. S6C, S5E). LINE elements and particularly the L1Md_T and Gf subtypes also showed significant gains in accessibility in TSA-treated cells (Fig. 2G; Supplemental Table S3),

consistent with a model in which transcription factor binding in the absence of HDAC activity and DNA methylation is contributing to chromatin opening.

A high fraction of the YY1 binding sites acquired in DNMT.TKO cells overlapped with LTR retrotransposons, of which 34/42 were ERVK-type elements (Fig. 4C). YY1 occupancy increases only modestly in the absence of DNA methylation or following HDAC inhibition (Supplemental Fig. S6D, left panel; Supplemental Fig. S5F). This differs from YY1 occupancy at LINE elements, which occurs after HDAC inhibition but not as a consequence of DNA methylation loss alone (Supplemental Fig. S6D, right panel). In both cases, YY1 occupancy is highest after HDAC inhibition in DNA methylation-deficient cells (Supplemental Fig. S6D).

Altogether, our analysis of ChIP-seq data indicates that both DNA methylation and HDAC activity play distinct roles in limiting occupancy of each of the examined TFs to a fraction of their binding sites. Furthermore, we did not detect a significant loss in transcription factor occupancy at pre-existing peaks following HDAC inhibition, whereas novel binding events were observed throughout the genome. These findings support our previous interpretation of the ATAC-seq data: that chromatin accessibility increases preferentially outside of established THSs and particularly at LINE and LTR retrotransposons.

Transcription factor occupancy can promote chromatin accessibility in mESCs with perturbed DNA methylation or HDAC activity

We next asked whether the gains in transcription factor occupancy in unmethylated or hyperacetylated cells were directly associated with increased chromatin accessibility. For each transcription factor, we evaluated the changes in ATAC-seq signal at intervals that were occupied in one of our conditions. Generally, increased GABPA, MAX, NRF1, and YY1 binding in DNMT.TKO cells was associated with significant gains in accessibility (Fig. 5A, top row). Novel binding events mostly occurred at sites that were previously inaccessible (Fig. 5B, top), and a few coincided with an increase in

ATAC-seq signal that was sufficient to be detected as a THS. Elevated transcription factor occupancy following HDAC treatment was also associated with significant gains in accessibility, although for GABPA this was only found to be the case after TSA treatment in
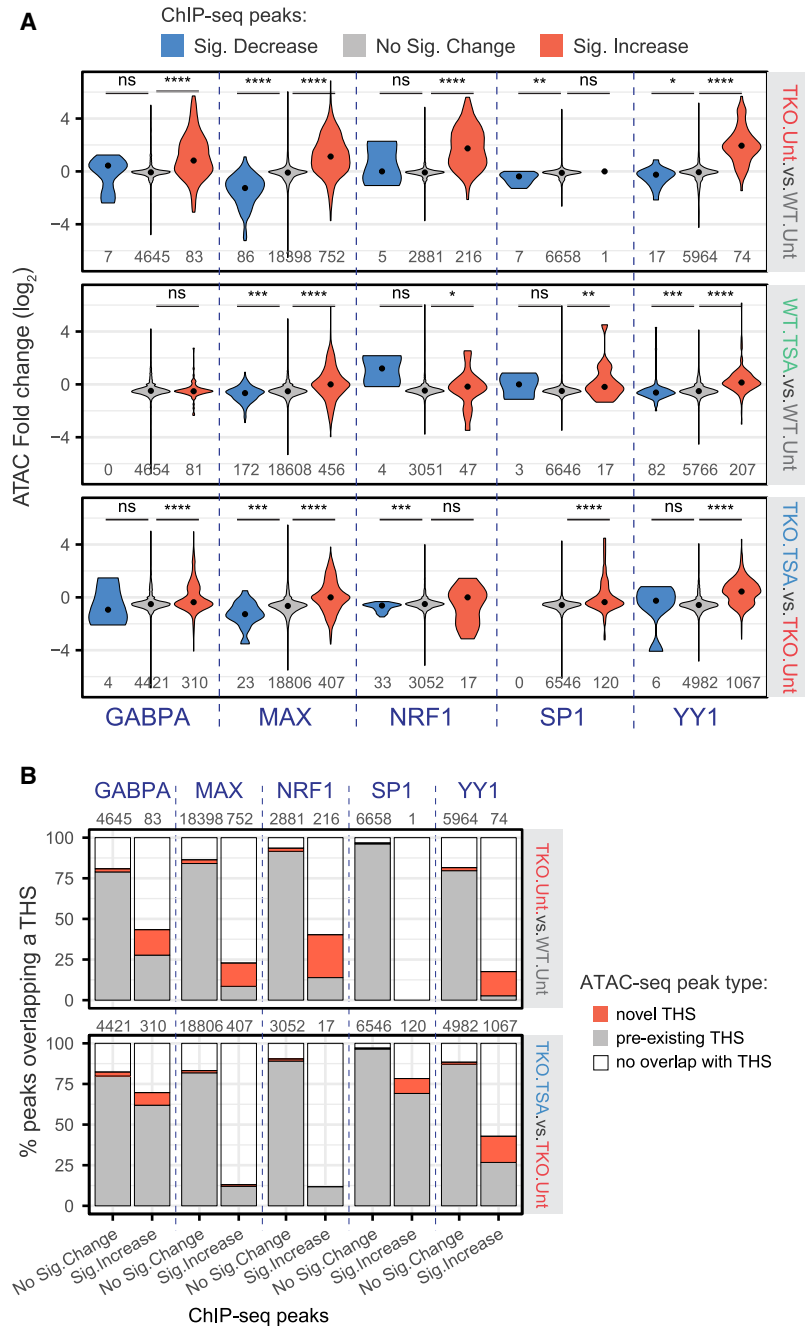

Figure 5. Transcription factor occupancy can promote chromatin accessibility in mESCs with perturbed DNA methylation or HDAC activity. (*A*) Quantitation of ATAC-seq fold change at transcription factor ChIP-seq peaks grouped according to their differential occupancy. Black dots indicate the median ATAC-seq fold change. One-tailed Mann–Whitney *U* tests; (*) *P*-value <0.05, (**) *P*-value <0.01, (***) *P*-value < 0.001, (****) *P*-value < 10^{-10}, (ns) nonsignificant (*P*-value > 0.05). (*B*) Percentage of ChIP-seq peaks that overlap with a Tn5 hypersensitive site (THS). In the *top* panel, a "pre-existing" THS refers to an ATAC-seq peak identified in samples generated from untreated WT cells, whereas a novel THS is identified in DNMT.TKO but not WT cells. In the *bottom* panel, a "pre-existing" THS was identified in samples generated from untreated DNMT.TKO cells, whereas a novel THS is identified in TSA-treated DNMT.TKO cells only.

DNMT.TKO but not WT cells (Fig. 5A, bottom two rows). Whereas HDAC inhibition led to increased occupancy of MAX, NRF1, and YY1 at predominantly inaccessible loci (i.e., that do not overlap any THS), the majority of sites to which GABPA and SP1 were

relocated were previously accessible (Fig. 5B, bottom). This finding is in agreement with our previous observations that all GAPBA and half of SP1 TSA-specific binding sites occur at promoter regions with low levels of DNA methylation (Fig. 4A; Supplemental Fig. S6A). Novel TF binding events that appear following HDAC inhibition are less commonly associated with increased accessibility when compared to DNMT.TKO-specific binding events, whereas the amplitude of change is also smaller (Fig. 5A). Although this observation might reflect a different capacity of HDAC inhibition to induce THSs, compared to DNA demethylation, we cannot exclude the possibility that the lower signal is a consequence of the depletion in ATAC-seq reads from THSs after treatment with TSA.

DNA methylation loss and HDAC inhibition affect the expression of specific genes and retrotransposons

Having established that both DNA methylation loss and HDAC inactivation can lead to changes in chromatin accessibility and transcription factor occupancy, we next examined the consequences for gene expression. To this aim, we performed RNA-seq using wild-type and DNMT.TKO cells following treatment with TSA or DMSO. Loss of DNA methylation was associated with the differential expression of hundreds of genes (5.8% of total) whereas HDAC inhibition had a more profound effect, with 20% and 27% of genes being significantly deregulated in TSA-treated wild-type and DNMT.TKO cells, respectively (Supplemental Fig. S7A; Supplemental Table S4).

We first explored the relationship between changes in chromatin accessibility and gene expression. We assigned each THS to its nearest transcription start site and examined the gene expression differences. We observed weak correlations between changes in chromatin accessibility and expression of the neighboring gene (Spearman's correlation coefficients for TKO.DMSO vs. WT.DMSO, WT.TSA vs. WT.DMSO, or TKO.TSA vs. TKO.DMSO are 0.15, 0.11, and 0.10, respectively) (Fig. 6A), indicating that most chromatin accessibility alterations are not associated with transcriptional changes at the closest gene. Nevertheless, more differentially accessible THSs show matching changes in neighboring gene expression than expected by chance (Fig. 6A). Many of the differential THSs are promoter distal and may represent loci with no regulatory function or that regulate transcriptional elements other than the closest gene. Limiting the list of THSs to those that are promoter proximal strengthens the correlation with gene expression when comparing WT and DNMT.TKO cells but not TSA and control treatments (Supplemental Fig. S7B).

At 28 promoters, significant changes in accessibility did associate with significantly increased gene expression in DNMT.TKO cells (Fig. 6B,C; Supplemental Fig. S7B). These 28 genes are enriched for genes whose expression is normally restricted to testes (Supplemental Table S5), fitting with a known role for DNA methylation in controlling the expression of germ cell-specific genes (De Smet et al. 1999; Maatouk et al. 2006; Fouse et al. 2008; Borgel et al. 2010; Auclair et al. 2014). In conclusion, DNA methylation can restrict gene expression by maintaining an inaccessible chromatin state; however, the majority of gained THSs both in demethylated or hyperacetylated genomes are unable to affect the mRNA abundance of the nearest genes.

Considering our observations that certain subtypes of LINE-1 and LTR retrotransposons were selectively accessible and bound by transcription factors following DNA methylation loss or HDAC inhibition, we turned our attention to transcription originating from these repetitive elements. In the case of LINE-1 and ERVL-type LTR

repeats, we noticed that increases in accessibility that are observed following HDAC inhibition but not after loss of DNA methylation were associated with significantly elevated levels of RNA originating from various subtypes (Fig. 6D; Supplemental Fig. S8A,C; Supplemental Table S3). These included the young L1Md_T, Gf, and A subtypes for which there was also enrichment for YY1 or GABPA ChIP-seq reads (Fig. 6D), as well as MT2-Mm and MERVL-int ERVL elements. In the case of ERVK-type repeats, loss of DNA methylation rather than HDAC inhibition accounts for the most pronounced increases in accessibility, although this is restricted to a subset of repeat types, including IAP elements (Supplemental Fig. S8B,C). However, the expression from ERVK-type elements is most strongly induced when both DNA methylation and HDAC activity are abolished (comparing TKO.TSA cells to WT.DMSO). This observation agrees with previous reports documenting a synergistic activation of ERVK-type repeats in response to a combined removal of DNA methylation and HDAC activity (Cameron et al. 1999; Coffee et al. 1999; Walter et al. 2016; Brocks et al. 2017).

In summary, elevated transcription factor binding and genome accessibility in the absence of DNA methylation or after HDAC inhibition promote activation of a small subset of genes, while up-regulating expression of the youngest and potentially mobile LINE-1 and LTR repeats.

Discussion

In this study, we demonstrate that transcription factors have different abilities to sense chromatin modifications. Even when examining a small number of TFs, we were able to capture distinct scenarios in which TF occupancy is affected by neither DNA modifications nor histone acetylation (SP1), exclusively by DNA methylation (NRF1), or by both (GABPA, MAX, SP1, and YY1) (Fig. 7).

The total number of new binding events varies between transcription factors and does not correlate with the number of all potential binding sites (i.e., sequences that match its cognate motif perfectly). For example, in DNA methylation-deficient cells, NRF1 gains occupancy at 216 loci (increase by 0.6% of all potential binding sites), whereas MAX gains occupancy at 752 (0.26%), YY1 – 74 (0.02%), and GABPA – 83 (0.02%) loci. Changes of a similar magnitude have been reported independently for another two proteins, MYCN and CTCF, in DNMT.TKO mES cells (Stadler et al. 2011; Yin et al. 2017). Comprehensive in vitro experiments indicate that there are >100 TFs whose affinity to DNA is diminished by cytosine methylation (Hu et al. 2013; Yin et al. 2017). Therefore, although the contribution of DNA methylation in restricting the occupancy of individual transcription factors appears limited, the cumulative effect of DNA methylation loss on the gene regulatory network is likely to be more substantial.

Unlike the effect of DNA methylation depletion, HDAC inhibition did not result in the emergence of many discrete Tn5 hypersensitive loci. Rather, increased chromatin accessibility was observed throughout the genome and particularly at a subset of abundant LINE and LTR transposable elements that led to a significant skew in the composition of ATAC-seq libraries. Despite genome-wide hyperacetylation, aberrant binding of the examined transcription factors occurred at a scale similar to that seen following the loss of DNA methylation. Perturbations to the DNA methylation machinery or HDAC activity had distinct impacts on MAX, GABPA, and YY1 localization both in terms of the number of aberrantly occupied loci as well as the genomic location of these alterations, indicating that DNA methylation does not depend on
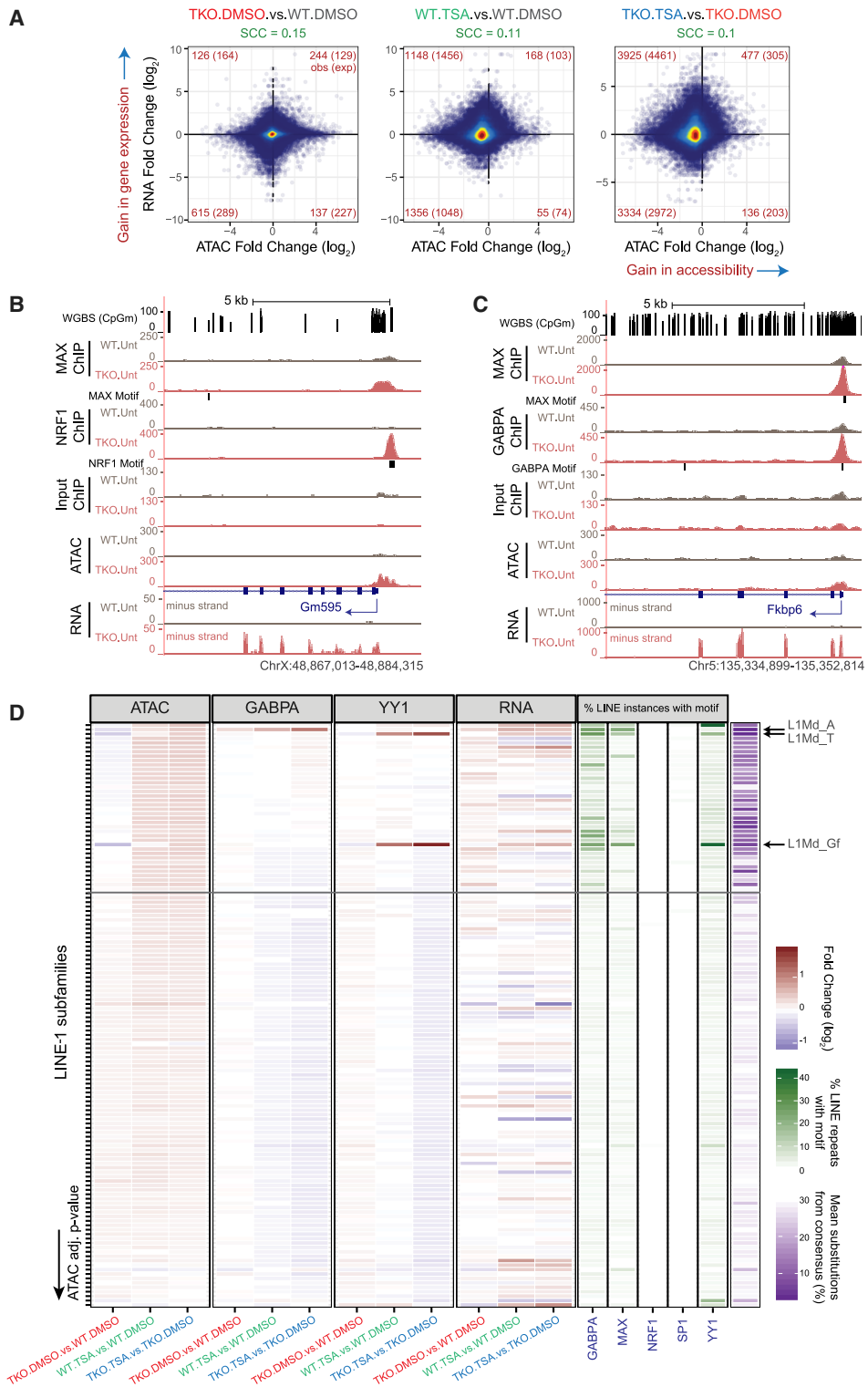
Figure 6. DNA methylation loss and HDAC inhibition affect the expression of specific genes and retrotransposons. (*A*) Changes in accessibility at every THS region were compared to changes in expression at their closest gene. The analysis was performed on 60052 THS:gene pairs involving 15,298 genes. The numbers of THSs associated with significant changes in both accessibility and gene expression are indicated in red in each quadrant. In parentheses are indicated the expected numbers of sites showing both significant changes in accessibility and gene expression based on the total number of significant differential events. SCC = Spearman's correlation coefficient. (*B,C*) Representative UCSC Genome Browser snapshots showing CpG methylation levels, ChIP-seq, ATAC-seq, and strand-specific RNA-seq read coverage. The position of genes and that of sequences that match the TF motifs are shown. (*D*) For each LINE-1 subtype (*N* = 132), we plotted the fold change in ATAC-seq, ChIP-seq, or RNA-seq signal along with scores relating to their sequence conservation (purple) or the presence of selected TF binding motifs (green). LINE-1 subtypes were sorted based on ATAC-seq adjusted *P*-values when comparing TSA- to DMSO- treated DNMT.TKO cells. See also Supplemental Table S3.
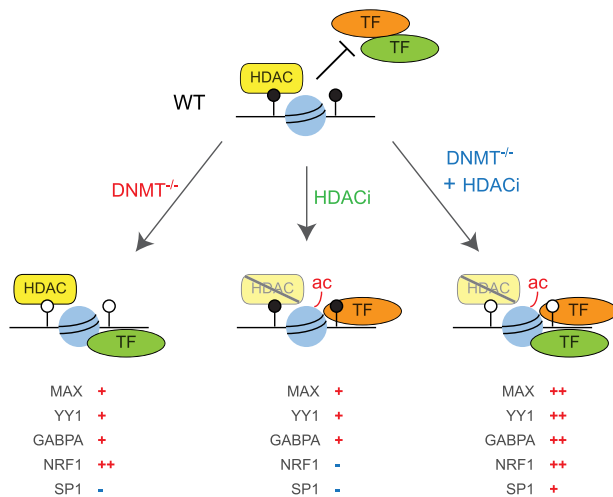
Figure 7. Summary model illustrating impact of DNA methylation and HDAC inhibition on transcription factor occupancy.

HDAC activity to restrict TF occupancy in mouse ES cells. One of the models that is recurrently proposed to explain the transcriptional silencing activity of DNA methylation suggests that methylated CpG sequences recruit MBD-domain proteins whose associated HDAC enzymes promote transcriptional repression (Eden et al. 1998; Nan et al. 1998; Kokura et al. 2001; Lyst et al. 2013). Except for a small number of cases, the effect of HDAC inhibition on chromatin accessibility, TF occupancy and gene expression does not recapitulate that of DNA methylation loss, in line with similar observations in mESCs (Brunmeir et al. 2010; Reichmann et al. 2012) and other cell lines (Lorincz et al. 2000; Brocks et al. 2017), arguing that, at large, DNA methylation functions independently of HDACs. A more direct role for cytosine methylation in lowering the affinity between transcription factors and their cognate DNA sequences is, however, consistent with our data, as well as being supported by various biochemical and structural experiments (Prendergast and Ziff 1991; Spruijt et al. 2013; Dantas Machado et al. 2015; Stephens and Poon 2016; Yin et al. 2017).

On the other hand, the manner by which histone deacetylation hinders transcription factor binding is unclear and remains to be elucidated. One potential model might be that TFs are unable to outcompete compact chromatin, which hypoacetylated nucleosomes are known to produce in vitro (Hong et al. 1993; Lee et al. 1993; Anderson et al. 2001; Shogren-Knaak et al. 2006; Wang and Hayes 2008).

Although there are only a few genomic loci for which HDAC inhibition or DNA methylation loss are directly responsible for altered TF binding and transcriptional activity, these elements are of biological and potentially clinical significance. LINEs and ERVLs, for example, are reported to be essential for early embryonic development following fertilization (Jachowicz et al. 2017; Percharde et al. 2018; Kruse et al. 2019), a period during which histone acetylation is elevated and DNA methylation is low (Macfarlan et al. 2012; Ishiuchi et al. 2015; Eckersley-Maslin et al. 2016). Moreover, LINE mobility is observed in ES cells and is important in genomic diversification (Akagi et al. 2008; MacLennan et al. 2017; Richardson et al. 2017). In human somatic tissues, aberrant L1 retrotransposition has been linked to cancer (Carreira et al. 2014; Tubio et al. 2014; Scott and Devine 2017; Schauer et al.

2018). Inhibitors of HDAC and DNMT enzymes are being used for the treatment of acute myelogenous leukemia, and their use against solid tumors is being considered in combination with immune checkpoint inhibitors (Navada et al. 2014; West and Johnstone 2014; Daskalakis et al. 2018).

In summary, we demonstrate that deacetylation of histones and DNA methylation are two important activities in chromatin, which collaborate to limit TF occupancy and contribute to silencing of retrotransposon expression.

Methods

Cell culture and treatment

$Dnmt1^{-/-}$, $Dnmt3a^{-/-}$, and $Dnmt3b^{-/-}$ TKO (DNMT.TKO, [Tsumura et al. 2006]) and their parental wild-type J1 male mouse embryonic stem cells (strain 129S4/SvJae) were grown on 0.1% gelatin-coated dishes in DMEM (Lonza) complemented with 10% (v/v) fetal bovine serum (Biosera, FB-1001G/500, lot #11484), 2 mM L-glutamine (Thermo Fisher Scientific), 1% (v/v) nonessential amino acids (Thermo Fisher Scientific), 1% (v/v) penicillin-streptomycin (Thermo Fisher Scientific), 10 ng/mL leukaemia inhibitory factor (LIF, produced in-house following a protocol by Tomala et al. 2010), and 0.1 mM beta-mercaptoethanol (Sigma-Aldrich). Chicken DF-1 (Doug Foster strain 1) fibroblast cells were grown in DMEM (Lonza) supplemented with 10% (v/v) fetal bovine serum (Biosera, FB-1001/500, lot #014BS386), 1% (v/v) penicillin-streptomycin (Thermo Fisher Scientific), 2 mM L-glutamine (Thermo Fisher Scientific), and 0.15% sodium bicarbonate (Thermo Fisher Scientific). All cells were incubated in a humidified incubator set at 37°C and 5% atmospheric CO_2.

HDAC inhibition in mESCs was achieved using Trichostatin A (TSA, Sigma-Aldrich, T1952). Cells were seeded 4–6 h prior to treatment; once attached, their growth medium was replaced with medium containing 5 ng/μL TSA. TSA-containing media was refreshed after 24 h, and cells were harvested for downstream applications after 36 h. Control DMSO-treated cells were seeded and treated in a similar manner; however, medium contained 0.0033% DMSO.

Calibrated native chromatin immunoprecipitation with exogenous spike-in

Chicken DF-1 and mouse ES cells were harvested by trypsinization. For each experimental condition, 5×10^6 chicken DF-1 cells were added to 5×10^7 mESCs. Cell mixtures were pelleted and resuspended in 1 ml ice-cold RSB Buffer (10 mM NaCl, 10 mM Tris [pH 8.0], 3 mM $MgCl_2$). Chromatin was fragmented by adding 200 units of Micrococcal Nuclease (MNase, Fermentas) and incubating at 37°C for 5 min. The reaction was stopped by the addition of 8 μL of 0.5 M EDTA and spun at 2348g for 5 min at 4°C. The supernatant containing soluble nucleosomes was collected (S1 fraction). Nucleosomes in the remaining insoluble fraction were extracted by rotating in 300 μL Nucleosome Release Buffer (10 mM Tris [pH 7.5], 10 mM NaCl, 0.2 mM EDTA, with 1× cOmplete EDTA-free Protease Inhibitor Cocktail [Roche] and 10 mM sodium butyrate) for 1 h at 4°C followed by passing through a 25G needle five times. The supernatant was collected as described above and pooled with the S1 chromatin fraction. DNA fragment size was checked by agarose gel electrophoresis, indicating that the majority of the fragments are mononucleosomal DNA (150–200 bp) with a small fraction of di- and tri nucleosomal.

Immunoprecipitations were set up with 100 μL solubilized chromatin (roughly equivalent to 4×10^6 cells) and 3 μg of an anti-acetylated histone H3 antibody (Millipore-Sigma 06-599) in a total

of 1 mL Native ChIP Incubation Buffer (10 mM Tris [pH 7.5], 70 mM NaCl, 2 mM MgCl$_2$, 2 mM EDTA, and 0.1% Triton X-100, 1× cOmplete EDTA-free Protease Inhibitor Cocktail, and 10 mM sodium butyrate) and rotated overnight at 4°C. Immune complexes were captured using 20 µL Protein A agarose beads (Roche, previously saturated with 1 mg/mL BSA and 1 mg/mL yeast tRNA) for 1 h at 4°C with rotation, washed four times with Native ChIP Wash Buffer (20 mM Tris [pH 7.5], 2 mM EDTA, 125 mM NaCl, 0.1% Triton X-100) and once in TE buffer (10 mM Tris HCl [pH 8.0], 1 mM EDTA), and subsequently eluted in 100 µL elution buffer (1% SDS, 0.1 M NaHCO$_3$) by vigorous shaking for 30 min at room temperature. DNA was purified using the ZymoResearch ChIP DNA clean and concentrator kit (Zymo Research) following the manufacturer's instructions. For input samples, DNA was purified from 100 µL of solubilized chromatin using the same DNA purification procedure.

Material was quantified using Qubit (Invitrogen) and the size profile analyzed on the 2200 or 4200 TapeStation (Agilent, dsDNA HS Assay). Automated library preparation was performed on 5 ng input material using the Apollo prep system (Wafergen, PrepX ILMN 32i, 96 sample kit) and standard Illumina multiplexing adapters following the manufacturer's protocol up to pre-PCR amplification. Libraries were PCR-amplified (18 cycles) on a Tetrad (Bio-Rad) using the NEBNext High-Fidelity 2X PCR Master Mix (NEB) and in-house single indexing primers (Lamble et al. 2013). Equivalent amounts of individual libraries were pooled. Paired-end sequencing was performed using a HiSeq 4000 75-bp platform (Illumina, HiSeq 3000/4000 PE Cluster kit and 150 cycle SBS kit), generating a raw read count of >30 million reads per sample (Supplemental Table S1). Data analysis is provided in Supplemental Material.

ATAC-seq with exogenous spike-ins

Assay for Transposase Accessible Chromatin (ATAC)-seq experiments were carried out by adapting the protocol published by Buenrostro et al. (2013). Details of the sample preparation and data analysis are provided in Supplemental Material.

Chromatin immunoprecipitation for transcription factors

We describe the protocol for chromatin immunoprecipitation in detail in the Supplemental Material. In brief, mouse ES cells were crosslinked in 1% methanol-free formaldehyde (Thermo Fisher Scientific) alone or sequentially with 2 mM disuccinimidyl glutarate (DSG) first, then 1% formaldehyde (YY1 only). Immunoprecipitation was carried out on sonicated chromatin using antibodies listed in Supplemental Table S6. Following stringent washing, reverse crosslinking and DNA purification, libraries were prepared and sequenced using the same procedure as described for the native chromatin immunoprecipitation with exogenous spike-in experiments.

RNA-seq sample preparation and sequencing

Total RNA was isolated from mES cells using the TRI Reagent (Sigma-Aldrich) following the protocol provided by the manufacturer. Contaminating genomic DNA was removed by incubating 10 µg RNA with 10 U DNase I (Thermo Fisher Scientific) in Reaction Buffer for 30 min at 37°C. The reaction was stopped by adding 10 µL of 50 mM EDTA after which RNA was extracted using a phenol-choloroform purification and ethanol precipitation. Material was quantified using RiboGreen (Invitrogen) on the FLUOstar OPTIMA plate reader (BMG Labtech) and the size profile and integrity analyzed on the 2200 (Agilent, RNA ScreenTape). Input material was normalized to 1 µg prior to library preparation.

Total RNA was depleted of ribosomal RNA using a Ribo-Zero rRNA Removal kit (Epicentre/Illumina, Human/Mouse/Rat), and library preparation was completed using the NEBNext Ultra Directional RNA Library Prep kit for Illumina (New England Biolabs), both following the manufacturer's instructions. Libraries were amplified for 12 cycles. Equivalent amounts of individual libraries were pooled together. Paired-end sequencing was performed using a HiSeq 2000 51-bp platform (v3 SBS chemistry), generating a raw read count of >40 million reads per sample (Supplemental Table S1). Data analysis is provided in Supplemental Material.

Analysis of publicly available WGBS data

Whole-genome bisulfite sequencing data from E14 mouse ESCs (GSM1027571, Habibi et al. 2013) were analyzed using Bismark (v 0.12.5). Only cytosines with a read coverage of five or more were taken into account in this study.

Data access

ChIP-seq, ATAC-seq, and RNA-seq sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE131366.

Competing interest statement

The authors declare no competing interests.

References

Akagi K, Li J, Stephens RM, Volfovsky N, Symer DE. 2008. Extensive variation between inbred mouse strains due to endogenous L1 retrotransposition. *Genome Res* **18:** 869–880. doi:10.1101/gr.075770.107

Anderson JD, Lowary PT, Widom J. 2001. Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Biol* **307:** 977–985. doi:10.1006/jmbi.2001.4528

Athanikar JN, Badge RM, Moran JV. 2004. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res* **32:** 3846–3855. doi:10.1093/nar/gkh698

Auclair G, Guibert S, Bender A, Weber M. 2014. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol* **15:** 545. doi:10.1186/s13059-014-0545-5

Beard C, Li E, Jaenisch R. 1995. Loss of methylation activates Xist in somatic but not in embryonic cells. *Genes Dev* **9:** 2325–2334. doi:10.1101/gad.9.19.2325

Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, Forné T, Weber M. 2010. Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* **42:** 1093–1100. doi:10.1038/ng.708

Bourc'his D, Bestor TH. 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431:** 96–99. doi:10.1038/nature02886

Brocks D, Schmidt CR, Daskalakis M, Jang HS, Shah NM, Li D, Li J, Zhang B, Hou Y, Laudato S, et al. 2017. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat Genet* **49:** 1052–1060. doi:10.1038/ng.3889

Brunmeir R, Lagger S, Simboeck E, Sawicka A, Egger G, Hagelkruys A, Zhang Y, Matthias P, Miller WJ, Seiser C. 2010. Epigenetic regulation of a murine retrotransposon by a dual histone modification mark. *PLoS Genet* **6:** e1000927. doi:10.1371/journal.pgen.1000927

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Meth* **10:** 1213–1218. doi:10.1038/nmeth.2688

Cameron EE, Bachman KE, Myöhänen S, Herman JG, Baylin SB. 1999. Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat Genet* **21:** 103–107. doi:10.1038/5047

Carreira PE, Richardson SR, Faulkner GJ. 2014. L1 retrotransposons, cancer stem cells and oncogenesis. *FEBS J* **281:** 63–73. doi:10.1111/febs.12601

Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, Duc J, Jang SM, Turelli P, Trono D. 2014. Evolutionarily dynamic L1 regulation in embryonic stem cells. *Genes Dev* **28:** 1397–1409. doi:10.1101/gad.241661.114

Coffee B, Zhang F, Warren ST, Reines D. 1999. Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. *Nat Genet* **22:** 98–101. doi:10.1038/8807

Coufal NG, Garcia-Perez JL, Peng GE, Marchetto MCN, Muotri AR, Mu Y, Carson CT, Macia A, Moran JV, Gage FH. 2011. Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *PNAS* **108:** 20382–20387. doi:10.1073/pnas.1100273108

Dantas Machado AC, Zhou T, Rao S, Goel P, Rastogi C, Lazarovici A, Bussemaker HJ, Rohs R. 2015. Evolving insights on how cytosine methylation affects protein-DNA binding. *Brief Funct Genomics* **14:** 61–73. doi:10.1093/bfgp/elu040

Daskalakis M, Brocks D, Sheng Y-H, Islam MS, Ressnerova A, Assenov Y, Milde T, Oehme I, Witt O, Goyal A, et al. 2018. Reactivation of endogenous retroviral elements via treatment with DNMT- and HDAC-inhibitors. *Cell Cycle* **17:** 811–822. doi:10.1080/15384101.2018.1442623

Davis CM, Constantinides PG, van der Riet F, van Schalkwyk L, Gevers W, Parker MI. 1989. Activation and demethylation of the intracisternal A particle genes by 5-azacytidine. *Cell Differ Dev* **27:** 83–93. doi:10.1016/0922-3371(89)90738-7

DeBerardinis RJ, Kazazian HH. 1999. Analysis of the promoter from an expanding mouse retrotransposon subfamily. *Genomics* **56:** 317–323. doi:10.1006/geno.1998.5729

De Smet C, Lurquin C, Lethé B, Martelange V, Boon T. 1999. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol Cell Biol* **19:** 7327–7335. doi:10.1128/MCB.19.11.7327

Domcke S, Bardet AF, Ginno PA, Hartl D, Burger L, Schübeler D. 2015. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528:** 575. doi:10.1038/nature16462

Eckersley-Maslin MA, Svensson V, Krueger C, Stubbs TM, Giehr P, Krueger F, Miragaia RJ, Kyriakopoulos C, Berrens RV, Milagre I, et al. 2016. MERVL/zscan4 network activation results in transient genome-wide DNA demethylation of mESCs. *Cell Rep* **17:** 179–192. doi:10.1016/j.celrep.2016.08.087

Eden S, Hashimshony T, Keshet I, Cedar H, Thorne AW. 1998. DNA methylation models histone acetylation. *Nature* **394:** 842. doi:10.1038/29680

Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A, Van Neste L, Jaenisch R, Fan G. 2008. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, Polycomb binding and histone H3 lys4/lys27 trimethylation. *Cell Stem Cell* **2:** 160–169. doi:10.1016/j.stem.2007.12.011

Goll MG, Bestor TH. 2005. Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* **74:** 481–514. doi:10.1146/annurev.biochem.74.010904.153721

Goodier JL, Ostertag EM, Du K, Kazazian HH. 2001. A novel active L1 retrotransposon subfamily in the mouse. *Genome Res* **11:** 1677–1685. doi:10.1101/gr.198301

Habibi E, Brinkman AB, Arand J, Kroeze LI, Kerstens HHD, Matarese F, Lepikhov K, Gut M, Brun-Heath I, Hubner NC, et al. 2013. Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13:** 360–369. doi:10.1016/j.stem.2013.06.002

Hansen RS, Canfield TK, Fjeld AD, Gartler SM. 1996. Role of late replication timing in the silencing of X-linked genes. *Hum Mol Genet* **5:** 1345–1353. doi:10.1093/hmg/5.9.1345

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. doi:10.1016/j.molcel.2010.05.004

Hendrich B, Bird A. 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18:** 6538–6547. doi:10.1128/MCB.18.11.6538

Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, Ren B. 2013. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet* **45:** 1198–1206. doi:10.1038/ng.2746

Hong L, Schroth GP, Matthews HR, Yau P, Bradbury EM. 1993. Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *J Biol Chem* **268:** 305–314.

Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, Shin J, Cox E, Rho HS, Woodard C, et al. 2013. DNA methylation presents distinct binding sites for human transcription factors. *eLife* **2:** e00726. doi:10.7554/eLife.00726

Hu T, Zhu X, Pi W, Yu M, Shi H, Tuan D. 2017. Hypermethylated LTR retrotransposon exhibits enhancer activity. *Epigenetics* **12:** 226–237. doi:10.1080/15592294.2017.1289300

Ishiuchi T, Enriquez-Gasca R, Mizutani E, Bošković A, Ziegler-Birling C, Rodriguez-Terrones D, Wakayama T, Vaquerizas JM, Torres-Padilla M-E. 2015. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat Struct Mol Biol* **22:** 662–671. doi:10.1038/nsmb.3066

Jachowicz JW, Bing X, Pontabry J, Bošković A, Rando OJ, Torres-Padilla M-E. 2017. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat Genet* **49:** 1502–1510. doi:10.1038/ng.3945

Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19:** 187–191. doi:10.1038/561

Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H. 2004. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429:** 900–903. doi:10.1038/nature02633

Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M, et al. 2011. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements and chimaeric transcripts in mouse ES cells. *Cell Stem Cell* **8:** 676–687. doi:10.1016/j.stem.2011.04.004

Kokura K, Kaul SC, Wadhwa R, Nomura T, Khan MM, Shinagawa T, Yasukawa T, Colmenares C, Ishii S. 2001. The Ski protein family is required for MeCP2-mediated transcriptional repression. *J Biol Chem* **276:** 34115–34121. doi:10.1074/jbc.M105747200

Kribelbauer JF, Laptenko O, Chen S, Martini GD, Freed-Pastor WA, Prives C, Mann RS, Bussemaker HJ. 2017. Quantitative analysis of the DNA methylation sensitivity of transcription factor complexes. *Cell Rep* **19:** 2383–2395. doi:10.1016/j.celrep.2017.05.069

Kruse K, Diaz N, Enriquez-Gasca R, Gaume X, Torres-Padilla M-E, Vaquerizas JM. 2019. Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. *bioRxiv* doi:10.1101/523712

Kumari D, Usdin K. 2001. Interaction of the transcription factors USF1, USF2, and α-Pal/Nrf-1 with the FMR1 promoter. Implications for fragile X mental retardation syndrome. *J Biol Chem* **276:** 4357–4364. doi:10.1074/jbc.M009629200

Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-Fahmawi B, Piazza P. 2013. Improved workflows for high throughput library preparation using the transposome-based Nextera system. *BMC Biotechnol* **13:** 104. doi:10.1186/1472-6750-13-104

Lee DY, Hayes JJ, Pruss D, Wolffe AP. 1993. A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* **72:** 73–84. doi:10.1016/0092-8674(93)90051-q

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462:** 315–322. doi:10.1038/nature08514

Lopes EC, Valls E, Figueroa ME, Mazur A, Meng F-G, Chiosis G, Laird PW, Schreiber-Agus N, Greally JM, Prokhortchouk E, et al. 2008. Kaiso contributes to DNA methylation-dependent silencing of tumor suppressor genes in colon cancer cell lines. *Cancer Res* **68:** 7258–7263. doi:10.1158/0008-5472.CAN-08-0344

Lorincz MC, Schübeler D, Goeke SC, Walters M, Groudine M, Martin DIK. 2000. Dynamic analysis of proviral induction and de novo methylation: implications for a histone deacetylase-independent, methylation density-dependent mechanism of transcriptional repression. *Mol Cell Biol* **20:** 842–850. doi:10.1128/MCB.20.3.842-850.2000

Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F, et al. 2013. Rett syndrome mutations

abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat Neurosci* **16:** 898–902. doi:10.1038/nn.3434

Maatouk DM, Kellam LD, Mann MRW, Lei H, Li E, Bartolomei MS, Resnick JL. 2006. DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133:** 3411–3418. doi:10.1242/dev.02500

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. 2012. ES cell potency fluctuates with endogenous retrovirus activity. *Nature* **487:** 57–63. doi:10.1038/nature11244

MacLennan M, García-Cañadas M, Reichmann J, Khazina E, Wagner G, Playfoot CJ, Salvador-Palomeque C, Mann AR, Peressini P, Sanchez L, et al. 2017. Mobilization of LINE-1 retrotransposons is restricted by Tex19.1 in mouse embryonic stem cells. *eLife* **6:** e26152. doi:10.7554/eLife.26152

Maurano MT, Wang H, John S, Shafer A, Canfield T, Lee K, Stamatoyannopoulos JA. 2015. Role of DNA methylation in modulating transcription factor occupancy. *Cell Rep* **12:** 1184–1195. doi:10.1016/j.celrep.2015.07.024

Nan X, Ng H-H, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393:** 386–389. doi:10.1038/30764

Navada SC, Steinmann J, Lübbert M, Silverman LR. 2014. Clinical development of demethylating agents in hematology. *J Clin Invest* **124:** 40–46. doi:10.1172/JCI69739

Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A. 1999. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* **23:** 58–61. doi:10.1038/12659

Percharde M, Lin C-J, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, Biechele S, Huang B, Shen X, Ramalho-Santos M. 2018. A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell* **174:** 391–405.e19. doi:10.1016/j.cell.2018.05.043

Prendergast GC, Ziff EB. 1991. Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science* **251:** 186–189. doi:10.1126/science.1987636

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, et al. 2011. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol Cell* **44:** 361–372. doi:10.1016/j.molcel.2011.08.032

Reichmann J, Crichton JH, Madej MJ, Taggart M, Gautier P, Garcia-Perez JL, Meehan RR, Adams IR. 2012. Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells. *PLoS Comput Biol* **8:** e1002486. doi:10.1371/journal.pcbi.1002486

Rhee I, Bachman KE, Park BH, Jair K-W, Yen R-WC, Schuebel KE, Cui H, Feinberg AP, Lengauer C, Kinzler KW, et al. 2002. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416:** 552–556. doi:10.1038/416552a

Richardson SR, Gerdes P, Gerhardt DJ, Sanchez-Luque FJ, Bodea G-O, Muñoz-Lopez M, Jesuadian JS, Kempen M-JHC, Carreira PE, Jeddeloh JA, et al. 2017. Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Res* **27:** 1395–1405. doi:10.1101/gr.219022.116

Robert M-F, Morin S, Beaulieu N, Gauthier F, Chute IC, Barsalou A, MacLeod AR. 2003. DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. *Nat Genet* **33:** 61–65. doi:10.1038/ng1068

Rowe HM, Friedli M, Offner S, Verp S, Mesnard D, Marquis J, Aktas T, Trono D. 2013. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* **140:** 519–529. doi:10.1242/dev.087585

Sanchez-Luque FJ, Kempen M-JHC, Gerdes P, Vargas-Landin DB, Richardson SR, Troskie R-L, Jesuadian JS, Cheetham SW, Carreira PE, Salvador-Palomeque C, et al. 2019. LINE-1 evasion of epigenetic repression in humans. *Mol Cell* **75:** 590–604.e12. doi:10.1016/j.molcel.2019.05.024

Schauer SN, Carreira PE, Shukla R, Gerhardt DJ, Gerdes P, Sanchez-Luque FJ, Nicoli P, Kindlova M, Ghisletti S, Santos AD, et al. 2018. L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. *Genome Res* **28:** 639–653. doi:10.1101/gr.226993.117

Scott EC, Devine SE. 2017. The role of somatic L1 retrotransposition in human cancers. *Viruses* **9:** 131. doi:10.3390/v9060131

Sharif J, Endo TA, Nakayama M, Karimi MM, Shimada M, Katsuyama K, Goyal P, Brind'Amour J, Sun M-A, Sun Z, et al. 2016. Activation of endogenous retroviruses in *Dnmt1*$^{-/-}$ ESCs involves disruption of SETDB1-mediated repression by NP95 binding to hemimethylated DNA. *Cell Stem Cell* **19:** 81–94. doi:10.1016/j.stem.2016.03.013

Shogren-Knaak M, Ishii H, Sun J-M, Pazin MJ, Davie JR, Peterson CL. 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311:** 844–847. doi:10.1126/science.1124000

Sookdeo A, Hepp CM, McClure MA, Boissinot S. 2013. Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob DNA* **4:** 3. doi:10.1186/1759-8753-4-3

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, Münzel M, Wagner M, Müller M, Khan F, et al. 2013. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152:** 1146–1159. doi:10.1016/j.cell.2013.02.004

Stadler MB, Murr R, Burger L, Ivánek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, et al. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480:** 490–495. doi:10.1038/nature10716

Stephens DC, Poon GMK. 2016. Differential sensitivity to methylated DNA by ETS-family transcription factors is intrinsically encoded in their DNA-binding domains. *Nucleic Acids Res* **44:** 8671–8681. doi:10.1093/nar/gkw528

Tomala M, Lavrentieva A, Moretti P, Rinas U, Kasper C, Stahl F, Schambach A, Warlich E, Martin U, Cantz T, et al. 2010. Preparation of bioactive soluble human leukemia inhibitory factor from recombinant *Escherichia coli* using thioredoxin as fusion partner. *Protein Expr Purif* **73:** 51–57. doi:10.1016/j.pep.2010.04.002

Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. 1999. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci* **96:** 8681–8686. doi:10.1073/pnas.96.15.8681

Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR, et al. 2006. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11:** 805–814. doi:10.1111/j.1365-2443.2006.00984.x

Tubio JMC, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, Gundem G, Pipinikas CP, Zamora J, Raine K, et al. 2014. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345:** 1251343. doi:10.1126/science.1251343

Walsh CP, Chaillet JR, Bestor TH. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20:** 116–117. doi:10.1038/2413

Walter M, Teissandier A, Pérez-Palacios R, Bourc'his D. 2016. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *eLife* **5:** e11418. doi:10.7554/eLife.11418

Wang X, Hayes JJ. 2008. Acetylation mimics within individual core histone tail domains indicate distinct roles in regulating the stability of higher-order chromatin structure. *Mol Cell Biol* **28:** 227–236. doi:10.1128/MCB.01245-07

West AC, Johnstone RW. 2014. New and emerging HDAC inhibitors for cancer treatment. *J Clin Invest* **124:** 30–39. doi:10.1172/JCI69738

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, et al. 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356:** eaaj2239. doi:10.1126/science.aaj2239

Yoon H-G, Chan DW, Reynolds AB, Qin J, Wong J. 2003. N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Mol Cell* **12:** 723–734. doi:10.1016/j.molcel.2003.08.008

Zhang Y, Ng H-H, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. 1999. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13:** 1924–1935. doi:10.1101/gad.13.15.1924

Zhou M, Smith AD. 2019. Subtype classification and functional annotation of L1Md retrotransposon promoters. *Mob DNA* **10:** 14. doi:10.1186/s13100-019-0156-5

Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, et al. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500:** 477–481. doi:10.1038/nature12433