

Supplementary material

1. Table legends.
2. Supplementary figures.
3. Supplementary materials and methods.
4. Supplementary references.

1. Table legends.

Table S1. Number of raw and mapped reads for all high-throughput sequencing samples.

Table S2. Lists of enriched regions for ATAC-seq and ChIP-seq experiments along with differential analyses results and annotations.

Table S3. Raw read counts and differential analyses results at repetitive elements.

Table S4. Results of differential gene expression analyses.

Table S5. Complete results from motif enrichment and functional annotation analyses.

Table S6. List of antibodies used in ChIP-seq experiments. Nucleoside transition and retention times for Mass Spectrometry analysis.

2. Supplementary figures.

Figure S1

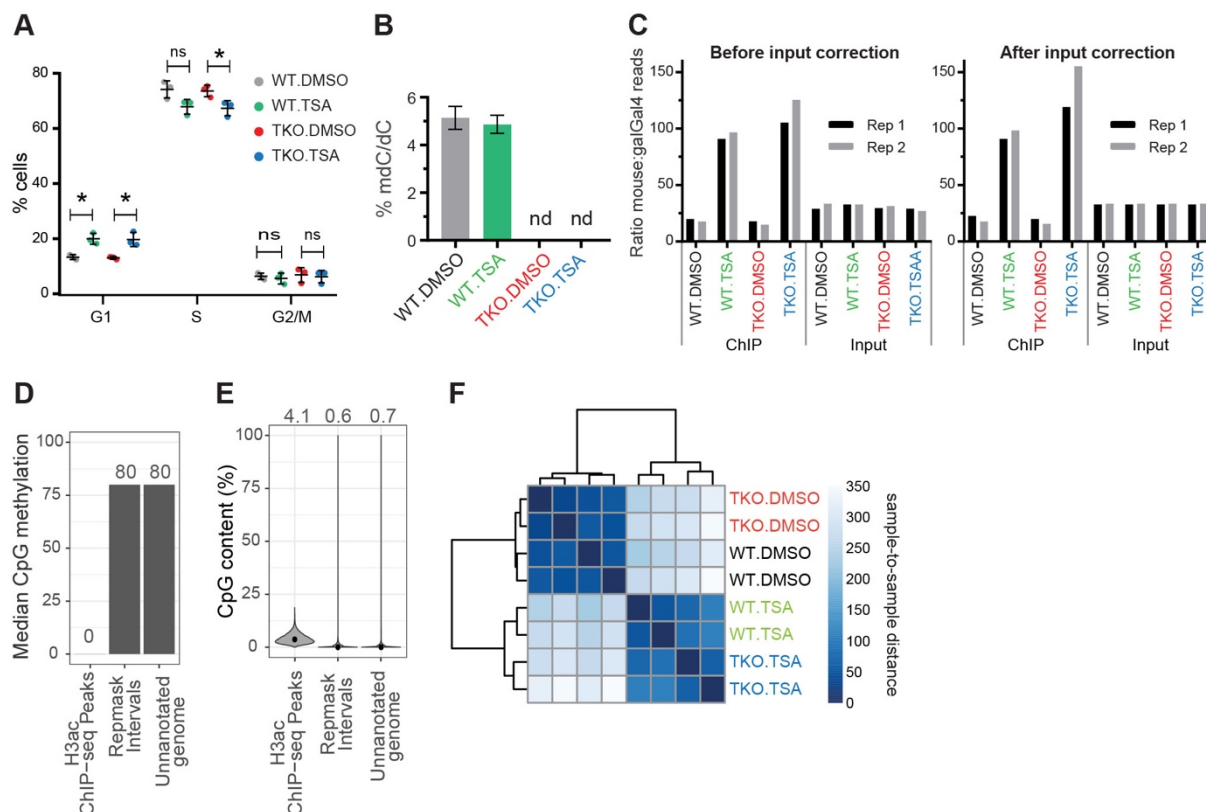
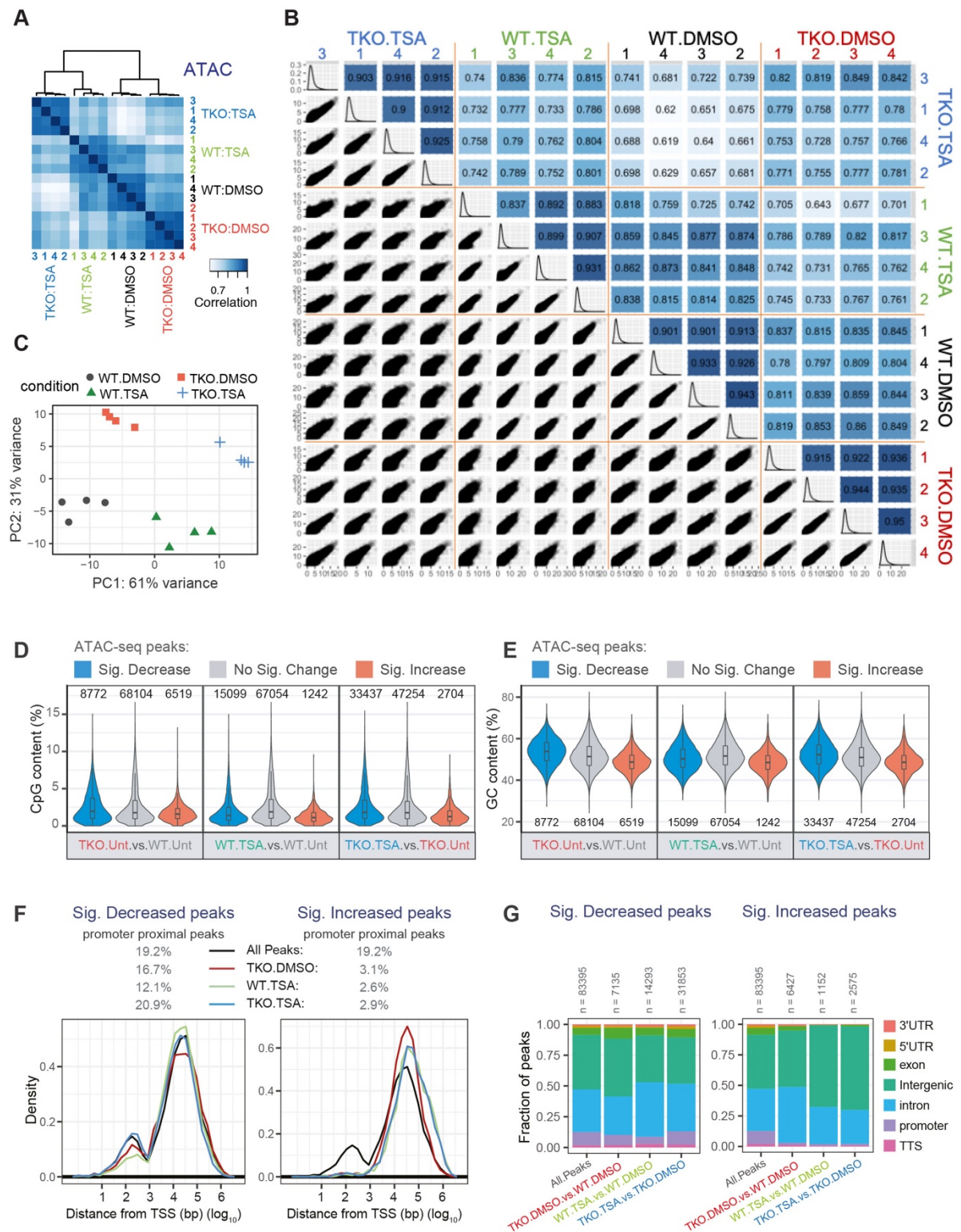
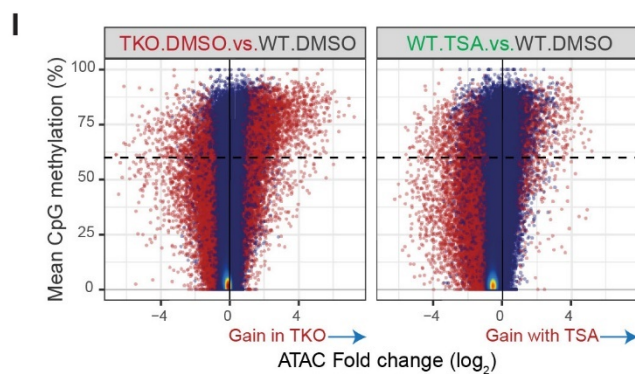
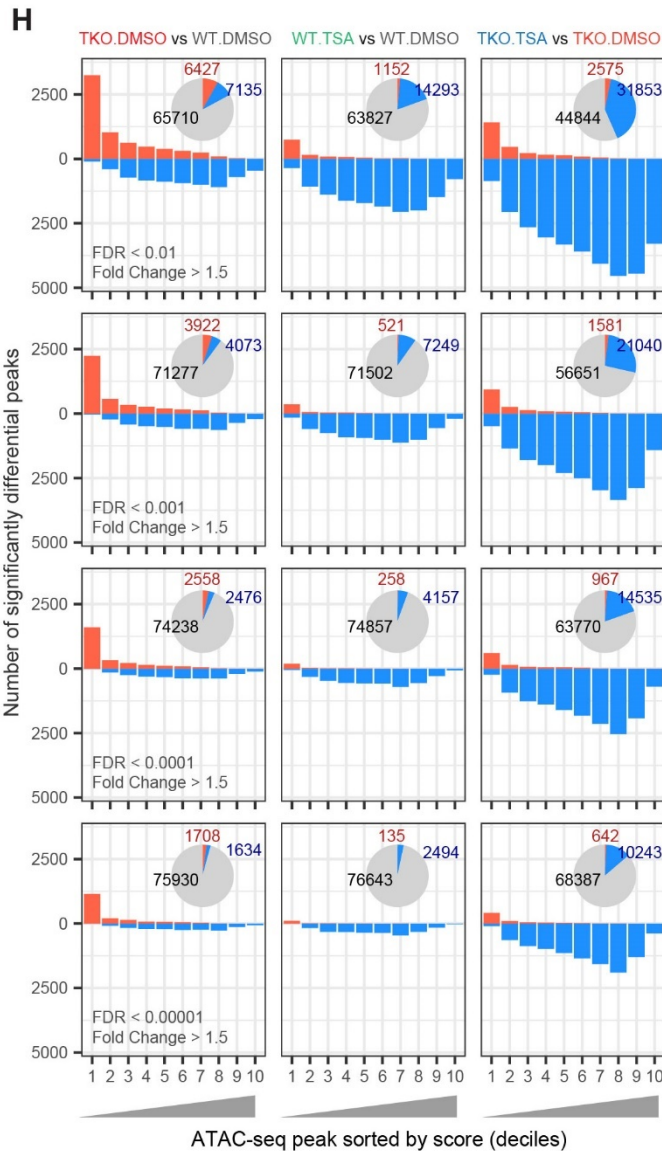


Figure S1. Disruption of HDAC activity and DNA methylation in mESCs. Related to Figure 1.

(A) Cell cycle profile after 36 h of TSA or DMSO control treatment. Interphase (G1), actively replicating (S) and mitotic (G2/M) cells were characterised using BrdU incorporation and 7-AAD staining. Each point represents the mean of two technical duplicates. Mean and SD of three biological replicates are shown, Student t-tests were used to determine significant differences: ns = non-significant (p -value > 0.05). (B) Mass spectroscopy measurements of 5mC in DMSO or TSA-treated wild-type and DNMT.TKO cells shown as a percentage of total cytosine. Data are represented as mean \pm SD. nd: not detected. (C) Number of reads mapping to the mm10 versus galGal4 genomes for each sample. Data are shown before and after adjusting for the variation within input samples (see Materials and Methods, Supplementary Material and Table S1). (D) Median CpG methylation levels for genomic intervals separated into three mutually exclusive categories. Whole genome bisulfite sequencing data from E14 mESCs was obtained from Habibi et al., 2013. Exact values are shown above the bars. (E) Distribution of CpG density at genomic intervals separated into three mutually exclusive categories. Mean values are indicated above the plot. (F) Heatmap of sample-to-sample euclidean distances calculated using calibrated and rlog transformed read count data from every ChIP-seq peak.

Figure S2



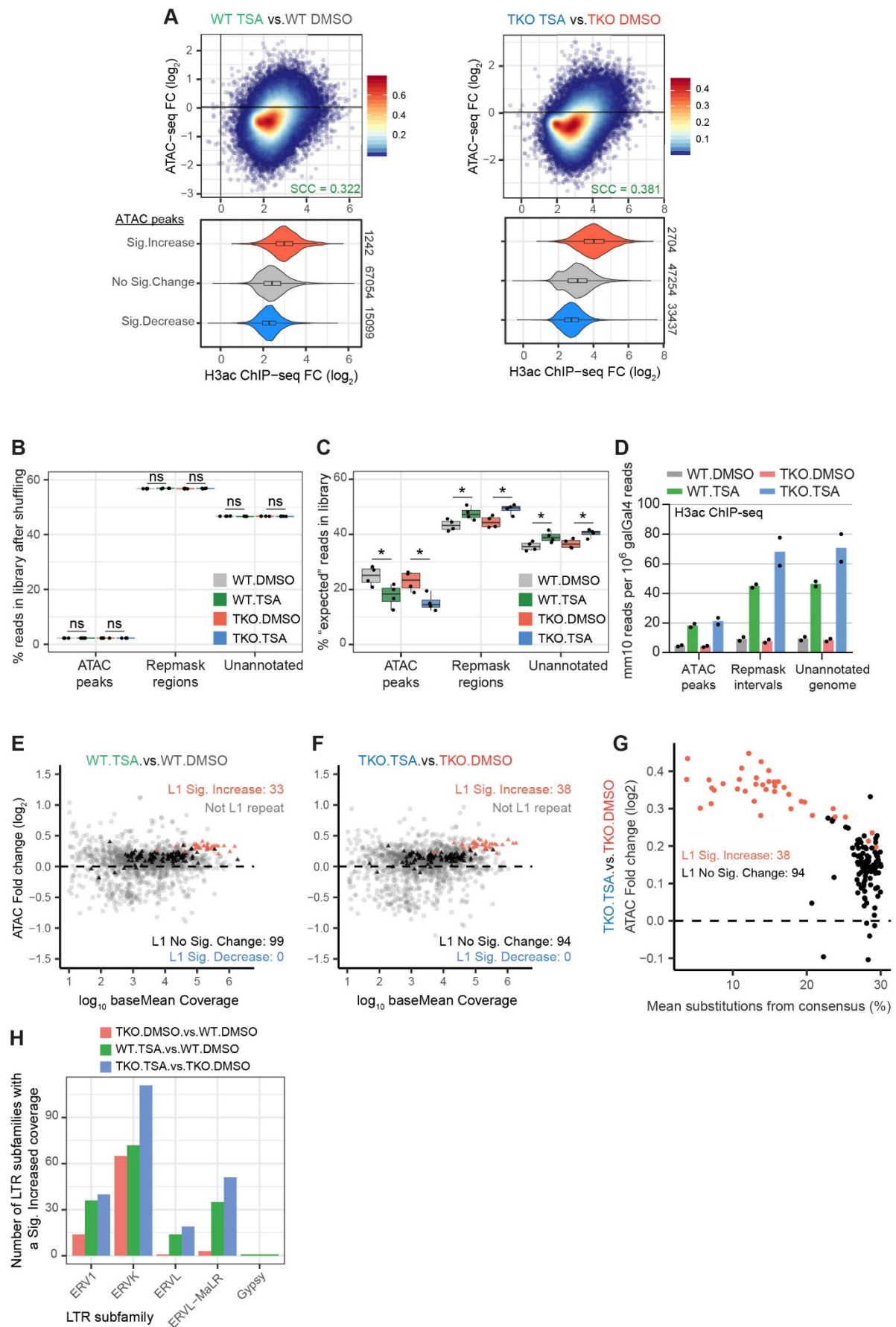


J

Known motif	HOMER database motif entry	Enrichment p-value	% of target sequences	% of background sequences
	NRF1 (NRF)	1e-67	10.42	2.09
	NFY (CCAAT)	1e-47	26.58	13.35
	NRF (NRF)	1e-42	10.94	3.48
	E-box (bHLH)	1e-25	4.58	1.08
	Elk4 (ETS)	1e-24	16.61	8.9

Figure S2. Global comparison of ATAC-seq datasets from different cell lines and conditions Related to Figure 2. (A) Correlation heatmaps comparing the ATAC-seq signal obtained for every sample at all identified THSs. ATAC-seq signal was calculated as the RPKM divided by the RPKM of the input control. Pearson correlation coefficients were calculated for each pair of samples and these values were used for clustering and the colour scale. Correlation heatmaps were generated in R using the `dba.plotHeatmap()` function of the DiffBind package (Stark and Brown, 2011). (B) Scatterplots in the lower triangle compare RPKM values at each THS for each pair of samples (4 replicates per sample) while corresponding Pearson correlation coefficients are shown in the top section. Histograms on the diagonal indicate the distribution of RPKM values. (C) Principle component analysis (PCA) plot generated using regularised log-transformed read counts from all ATAC-seq samples at every THS. The R package DESeq2 (Love et al. 2014) was used for plotting. (D) CpG or (E) GC content in sequences underlying THSs, grouped according to their differential accessibility. (F) Density plots showing the distance of THSs to the nearest transcription start site (TSS). The percentages of THS that are proximal to the TSS (centre of peak within 1500 bp of TSS) are indicated above the plots. (G) Fraction of THSs that overlap different genomic features. (H) THS regions were sorted according to their ATAC-seq signal in wild-type cells and grouped into ten bins of equal size ($n = 8340$). Bar charts indicate the number of THSs called as having significantly increased (red) or decreased (blue) accessibility per bin while pie charts summarise the total number of significantly differential calls. This was repeated using progressively more stringent false discovery rate (FDR) thresholds. (I) Scatterplots comparing the change in ATAC-seq signal with the average CpG methylation levels at 83395 THS intervals. Regions with significantly differential accessibility (fold change > 1.5 and $FDR < 0.01$) are shown in red. The dashed line indicates 60% CpG methylation. (J) The known motifs most enriched in DNMT.TKO specific THSs (target sequences) relative to background sequences. The HOMER motif database entry provides information about the transcription factor's DNA-binding domain (DBD-family in brackets). Target sequences were restricted to those that were methylated ($\geq 60\%$ mean CpG methylation) and had at least a four-fold increase in accessibility compared to WT cells. See also Table S5.

Figure S3

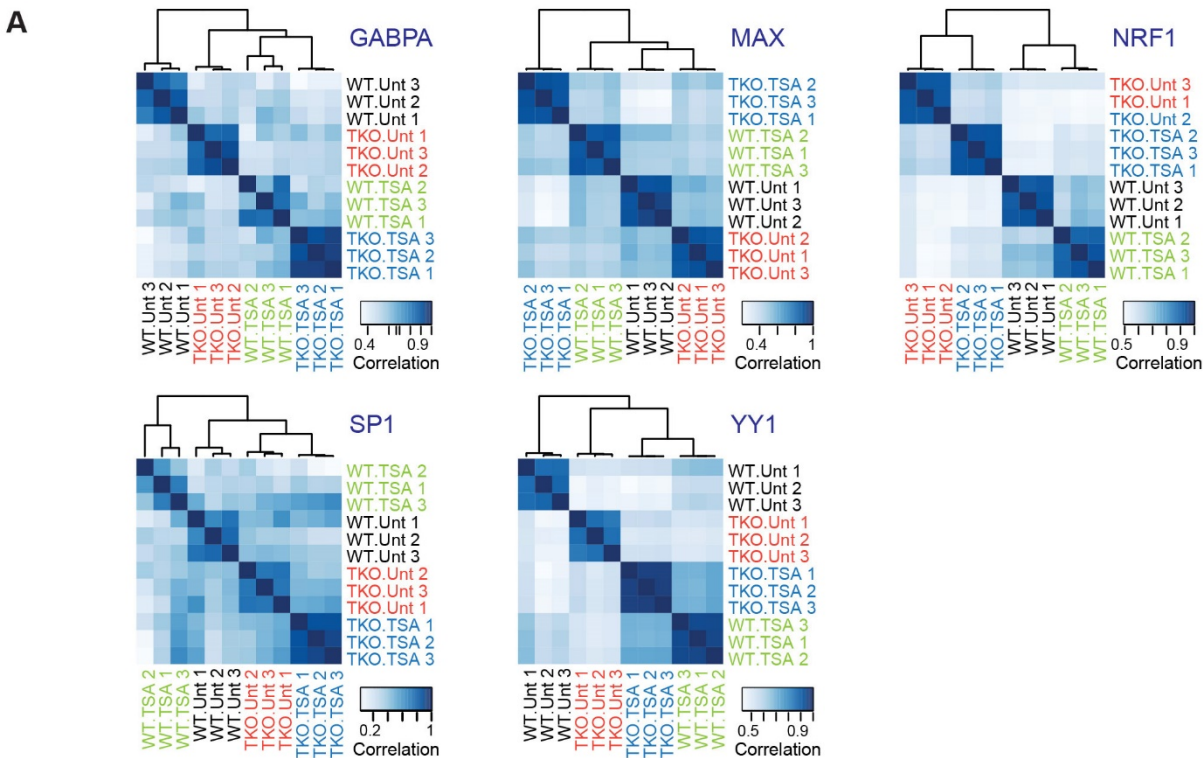


WT TSA vs. WT DMSO	Known motif	HOMER database motif entry	Enrichment p-value	% of target sequences	% of background sequences
		AP-2alpha (AP2)	1e-21	30.65	12.75
		AP-2gamma (AP2)	1e-19	34.79	16.69
		p53 (p53)	1e-14	8.29	1.57
		CTCF (Zf)	1e-08	7.83	2.41
		p53 (p53)	1e-07	2.53	0.27
		BORIS (Zf)	1e-07	8.76	3.28
		Fli1 (ETS)	1e-07	28.34	17.97
		ELF1 (ETS)	1e-07	14.98	7.47
TKO TSA vs. TKO DMSO	Known motif	HOMER database motif entry	Enrichment p-value	% of target sequences	% of background sequences
		AP-2alpha (AP2)	1e-48	35.36	12.84
		AP-2gamma (AP2)	1e-38	38.69	17.06
		p53 (p53)	1e-24	8.65	1.53
		p53 (p53)	1e-15	2.88	0.2
		Jun-AP1 (bZIP)	1e-13	9.56	3.26
		Fra1 (bZIP)	1e-11	16.24	7.91
		TEAD4 (TEA)	1e-11	24.43	14.26
		BATF (bZIP)	1e-10	18.21	9.87

Figure S3. Increased chromatin accessibility at retrotransposons upon HDAC inhibition in mESCs. Related to Figure 2. (A) Top, scatterplots comparing changes in accessibility at every THS ($n = 83395$) to changes in H3ac levels. Colours indicate density of points per graph area. SCC = Spearman Correlation Coefficient. Bottom, distribution of H3ac ChIP-seq fold-change values for THS regions grouped according to their differential accessibility. (B) Boxplots summarising the percentage of reads from each ATAC-seq library that map to intervals split into three mutually exclusive categories, after having randomly shuffled the reads throughout the genome. Two-tailed Student t-tests were used to determine significant differences: ns = non-significant ($p\text{-value} > 0.05$). (C) Boxplots summarising the percentage of reads from each ATAC-seq library that map to intervals split into three mutually exclusive categories, after having randomly shuffled the non-peak reads within the “Repmask” and “Unannotated” regions. Two-tailed Student t-tests were used to determine significant differences: * $p\text{-value} < 0.05$; ** $p\text{-value} < 0.01$; ns = non-significant ($p\text{-value} > 0.05$). (D) 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 mean; points indicate the values for two biological replicates. (E) MA plots comparing for each type of repetitive element, the fold change in ATAC-seq signal across all genomic copies to the baseMean ATAC-seq signal. LINE-1 elements are highlighted. ATAC-seq samples generated from TSA- or DMSO-treated wild-type mESCs were compared. Differential analysis was performed using the R package DESeq2 (Love et al. 2014). Changes were deemed significant

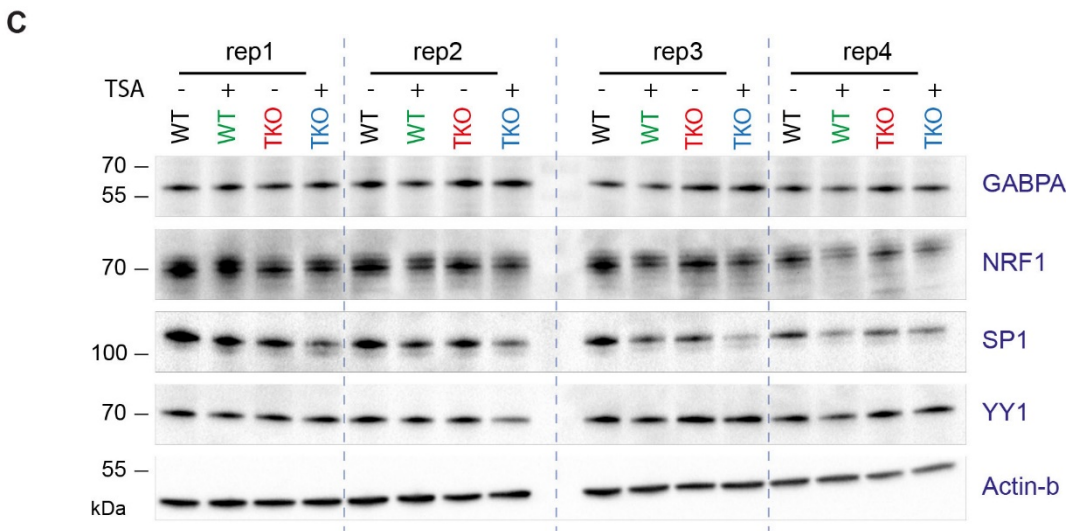
with an FDR < 0.05). See also Table S3. (F) Same as (E), except samples from TSA- and DMSO-treated DNMT.TKO mESCs were compared. (G) Scatterplot comparing for each LINE-1 repeat subtype, the fold change in ATAC-seq signal to the average substitution rate (base mismatches relative to the consensus sequence in parts per 100). See also Table S3. (H) Number of LTR element types, subdivided by family, that show significantly increased ATAC-seq signal for three pairwise comparisons. See also Table S3. (I) The known motifs most enriched in THS regions TSA-treated wild-type cells (top) or TSA-treated DNMT.TKO cells (bottom) (target sequences) relative to background sequences. The HOMER motif database entry provides information about the transcription factor's DNA-binding domain (DBD-family in brackets). Target sequences were restricted to those that had at least a four-fold increase in accessibility compared to WT or untreated cells. See also Table S5.

Figure S4

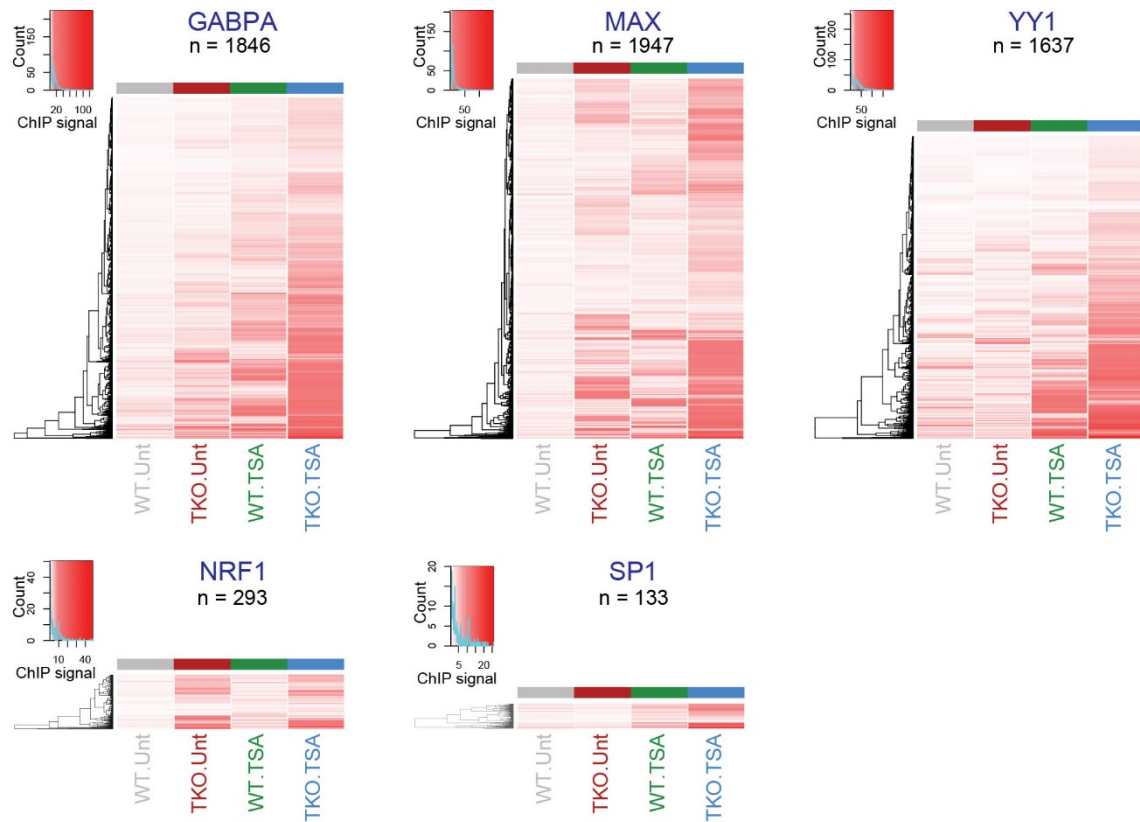


B

Transcription factor	Enrichment p-value	% target sequences	% background sequences	HOMER database entry
GABPA	1e-964	74.68	27.99	GABPA(ETS) Jurkat-GABPa-ChIP-Seq(GSE17954)
MAX	1e-1088	42.31	19.88	Max(bHLH) K562-Max-ChIP-Seq(GSE31477)
NRF1	1e-1723	84.3	13.48	NRF1(NRF) MCF7-NRF1-ChIP-Seq
SP1	1e-469	68.36	40.19	Sp1(Zf) Promoter
YY1	1e-1201	38.45	5.96	YY1(Zf) Promoter



D



E

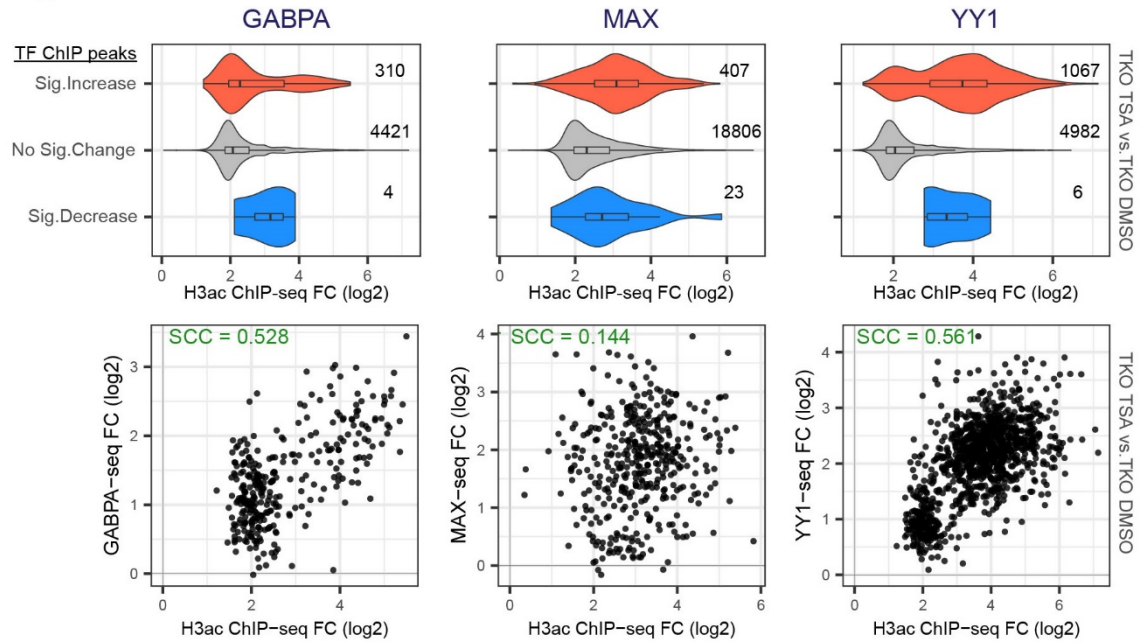


Figure S4: DNA methylation and HDAC activity can modulate transcription factors occupancy. Related to Figure 3. (A) Correlation heatmaps comparing the ChIP-seq signal obtained for every sample at all identified binding sites. ChIP signal was calculated as the RPKM of the TF ChIP-seq divided by the RPKM of the input control. Pearson correlation coefficients were calculated for each pair of samples and these values were used for clustering and the colour scale. Correlation heatmaps were generated in R using the `dba.plotHeatmap()` function of the DiffBind package (Stark and Brown, 2011). (B) Most enriched motifs found within transcription factors ChIP-seq peaks (target sequences) relative to background sequences. For all factors except GABPA, the motif shown is the most enriched compared to all others entries in the HOMER collection. This GABPA motif is ranked 5th most enriched, however the top 10 motifs enriched at GABPA peaks are all highly similar ETS factor motif (See Table S5). (C) Immunoblot analysis of GABPA, NRF1, SP1 and YY1 levels after 36 h TSA treatment in mESCs. Cell lysates were derived from four biological replicate experiments. (C) Heatmaps show the ChIP-seq signal for every peak region at which a significant difference in occupancy was detected between two dataset pairs (TKO.Unt vs. J1.Unt; J1.TSA vs. J1.Unt; TKO.TSA vs. TKO.Unt; or TKO.TSA vs. J1.Unt). ChIP-seq signal was calculated for each sample as the RPKM of the TF ChIP-seq divided by the RPKM of the input control. Values from three biological replicates were averaged for plotting. The heights of the heatmaps are scaled to the number of peak intervals. (E) Top, distribution of H3ac ChIP-seq fold change values for ChIP-seq peaks grouped according to their differential occupancy in TSA- versus DMSO-treated DNMT.TKO cells. Bottom, scatterplots comparing changes in TF occupancy to changes in H3ac levels at ChIP-seq peaks that are gained in TSA- versus DMSO-treated DNMT.TKO cells. SCC = Spearman Correlation Coefficient.

Figure S5

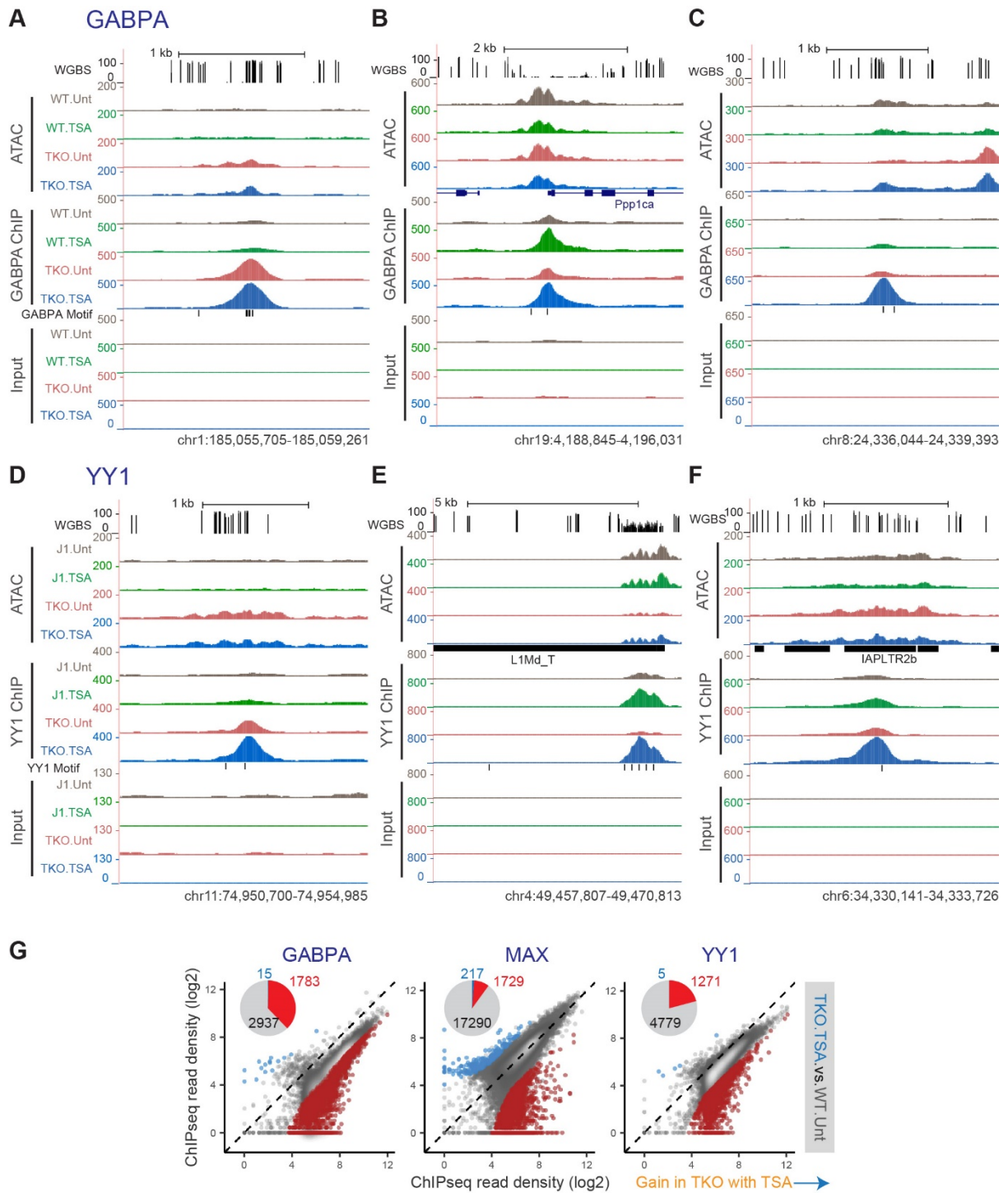


Figure S5. DNA methylation and HDAC activity can modulate transcription factors occupancy. Related to Figure 3. (A) Representative UCSC genome browser snapshot showing CpG methylation levels (Habibi et al., 2013), ATAC-seq, GABPA ChIP-seq and Input ChIP-seq read coverage. The position of genes and that of sequences that match the GABPA motif are shown. ATAC-seq and ChIP-seq coverage graphs were generated after merging alignments from replicate samples. Mm10 coordinates: chr1:185,055,705-185,059,261. (B) Same as (A).

Mm10 coordinates: chr19:4,188,845-4,196,031. (C) Same as (A). Mm10 coordinates: chr8:24,336,044-24,339,393. (D) Representative UCSC genome browser snapshot showing CpG methylation levels (Habibi et al., 2013), ATAC-seq, YY1 ChIP-seq and Input ChIP-seq read coverage. The position of repeat elements and that of sequences that match the YY1 motif are shown. ATAC-seq and ChIP-seq coverage graphs were generated after merging alignments from replicate samples. Mm10 coordinates: chr11:74,950,700-74,954,985. (E) Same as (D). Mm10 coordinates: chr4:49,457,807-49,470,813. (F) Same as (D). Mm10 coordinates: chr6:34,330,141-34,333,726. (G) For every occupancy peak identified for GABPA, MAX or YY1, normalised ChIP-seq signal was plotted for samples generated from TSA-treated DNMT.TKO cells (TKO.TSA) versus untreated wild-type cells (WT.Unt). Regions with significantly differential occupancy (fold change > 4 and adjusted p-value < 10⁻³) are coloured on the scatter plots and their numbers are summarised in the form of pie charts (light blue = significant decrease; red = significant increase).

Figure S6

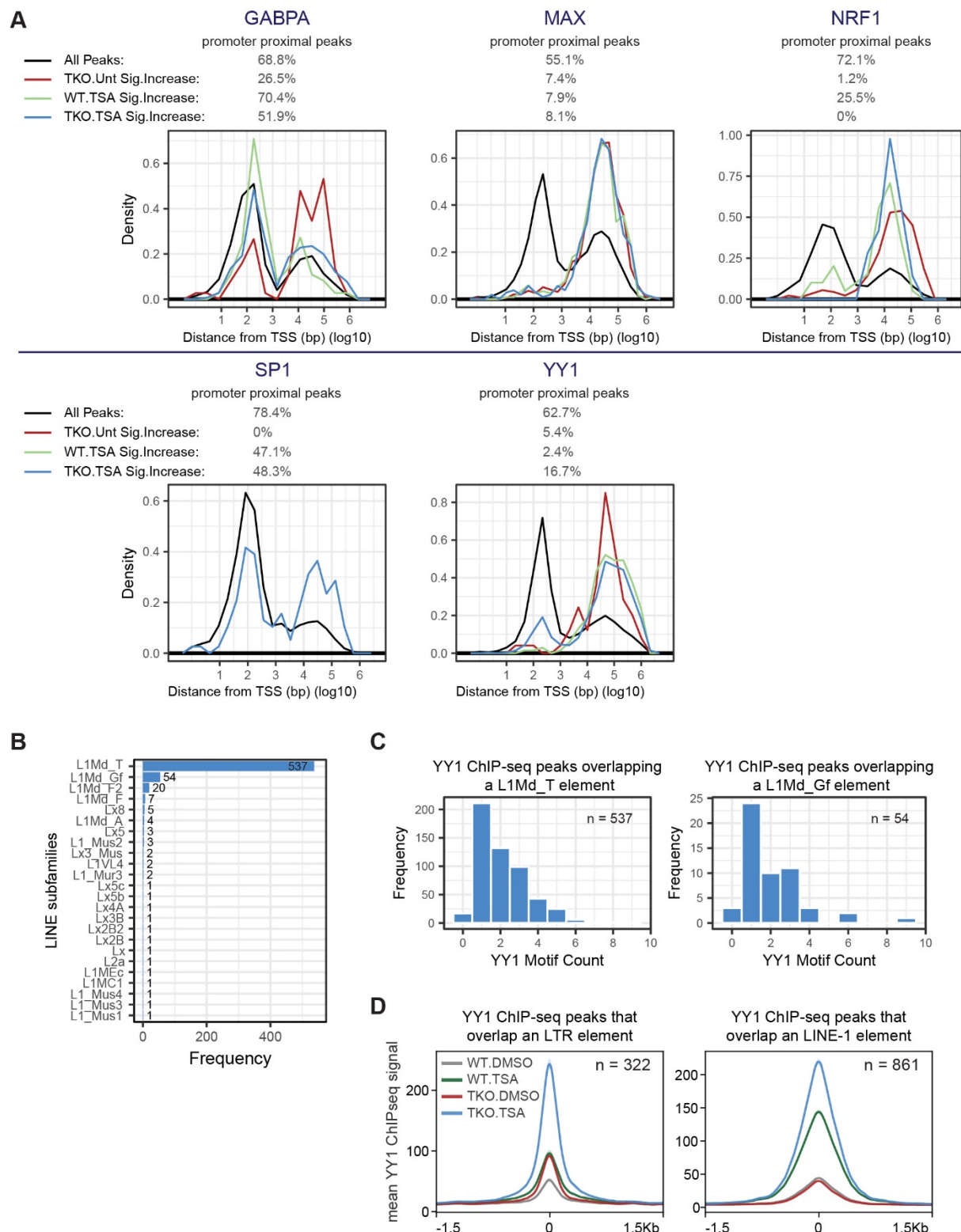


Figure S6. Characteristics of transcription factor binding sites. Related to Figure 4. (A) Density plots showing the distance of ChIP-seq peaks to the nearest transcription start site (TSS). The percentages of peaks that are proximal to the TSS (centre of peak within 1500 bp of TSS) are indicated above the plots. (B) The extent of overlap between TKO.TSA-specific YY1 ChIP-seq

peaks and different subtypes of LINE elements. Only peaks with significantly increased YY1 occupancy in TSA-treated versus untreated DNMT.TKO cells and that overlap a LINE are included (n = 652). (C) Number of TKO.TSA-specific YY1 ChIP-seq peaks that harbour a cognate binding motif. Left: peaks that overlap an L1Md_T element only. Right: peaks that overlap an L1Md_Gf element only. (D) Metaplots showing the average YY1 ChIP-seq signal in 3 kb regions surrounding the centre of ChIP-seq peaks that overlap either an LTR (left) or LINE (right) elements.

Figure S7

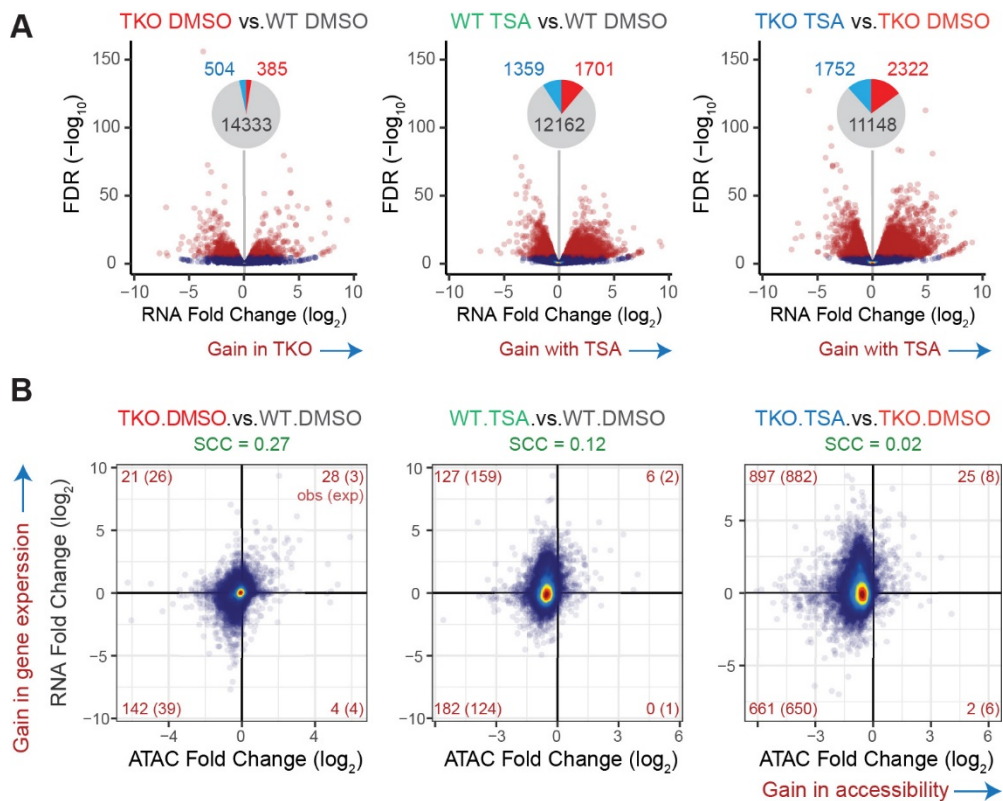


Figure S7. The impact of DNA methylation loss and HDAC inhibition on gene expression. Related to Figure 6. (A) Volcano plots representing the false discovery rate (FDR) and fold change values obtained through pairwise differential analyses of strand-specific RNA-seq read counts at 15222 genes. Genes with significantly differential expression (adjusted p-value < 10⁻⁵) are shown in red on the scatter plots and their numbers are summarised in the form of pie charts (light blue = significant decrease; red = significant increase). (B) Changes in accessibility at promoter-proximal THS regions (centre of peak within 1500 bp of TSS) were compared to changes in expression at their closest gene. The analysis was performed on 13068 THS:gene pairs involving 12356 genes. The numbers of THSs associated with significant changes in both accessibility and gene expression are indicated in red in each quadrant. In brackets are indicated the expected number of sites showing both significant changes in accessibility and gene expression based on the total number of significant differential events. SCC = Spearman Correlation Coefficient.

Figure S8

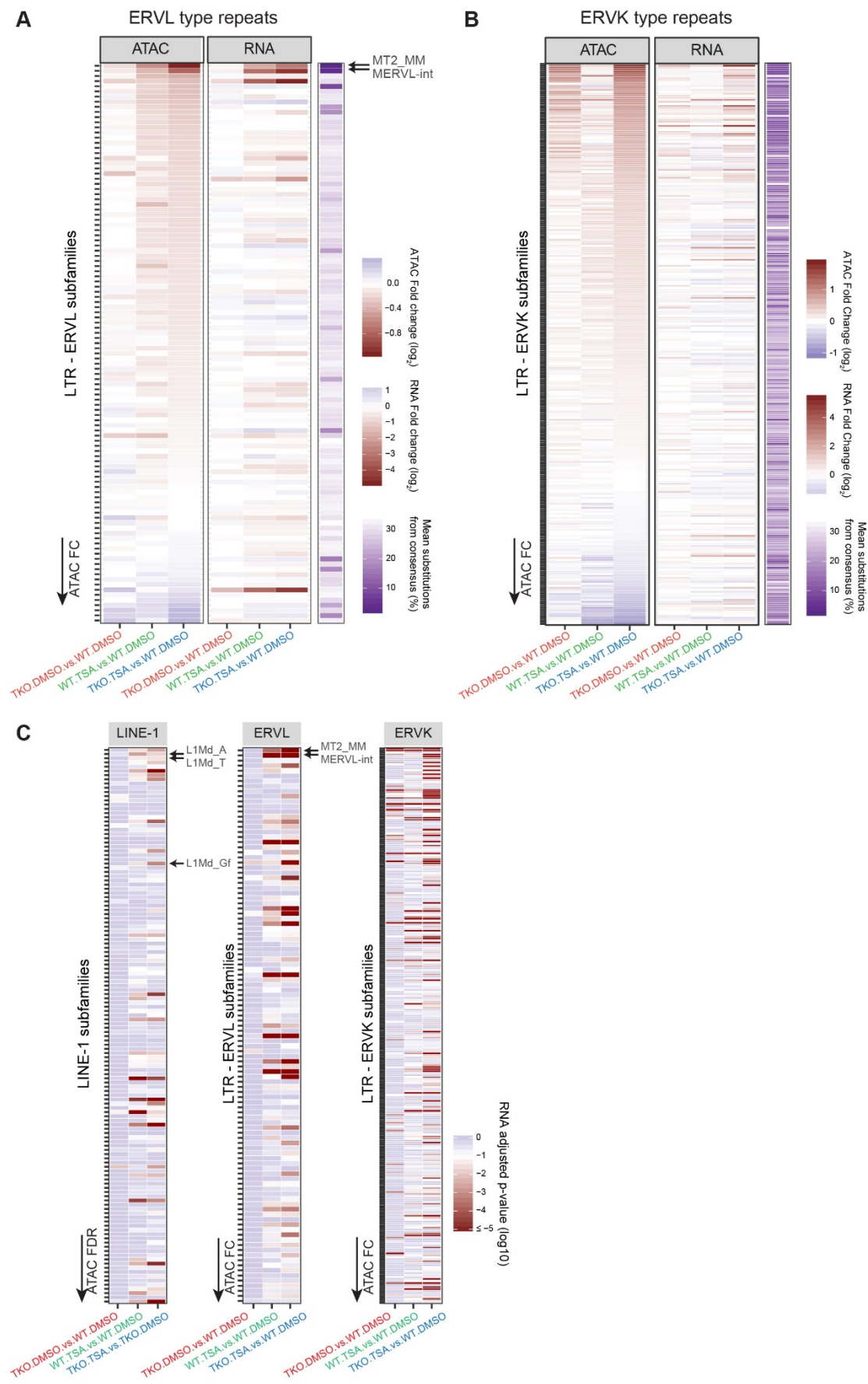


Figure S8. The impact of DNA methylation loss and HDAC inhibition on expression from retrotransposons. Related to Figure 6. (A) For each ERVL subtype (N = 109), we plotted the fold change in ATAC-seq or RNA-seq signal along with scores relating to their sequence conservation (purple). ERVL subtypes were sorted based on the fold change in ATAC-seq coverage when comparing DNMT.TKO-TSA to wild-type cells. See also Table S3. (B) For each ERVK subtype (N = 231), we plotted the fold change in ATAC-seq or RNA-seq signal along with scores relating to their sequence conservation (purple). ERVK subtypes were sorted based on the fold change in ATAC-seq coverage when comparing DNMT.TKO-TSA to wild-type cells. See also Table S3. (C) For each LINE-1 (N = 132), ERVL (N=109) or ERVL subtype (N=231), adjusted p-values for the change in RNA-seq signal were plotted. LINE-1, ERVL and ERVK repeat subtypes were sorted as in the heatmaps shown in figures 6D, S8B and S8A respectively. See also Table S3.

3. Supplementary materials and methods.

Cell cycle analysis

Approximately one million cells were labelled with 10 μ M BrdU for 30 minutes followed by the staining of the incorporated BrdU and total DNA using the APC BrdU Flow Kit (BD Pharmingen). The stained cells were analysed using a Fortessa flow cytometer (APC-BrdU: R670_30, 7AAD: B710_50) and data were evaluated with FlowJo software.

ATAC-seq computational analysis

Whilst we incorporated spike-ins into our ATAC-seq experiments to facilitate calibration, the number of reads mapping to the exogenous galGal4 genome varied widely between the four biological replicates of a same experimental condition (Figure M1 below). Due to this variation and the lack of a consistent trend when comparing samples from different experimental conditions, we omitted the spike-in data from our analyses.

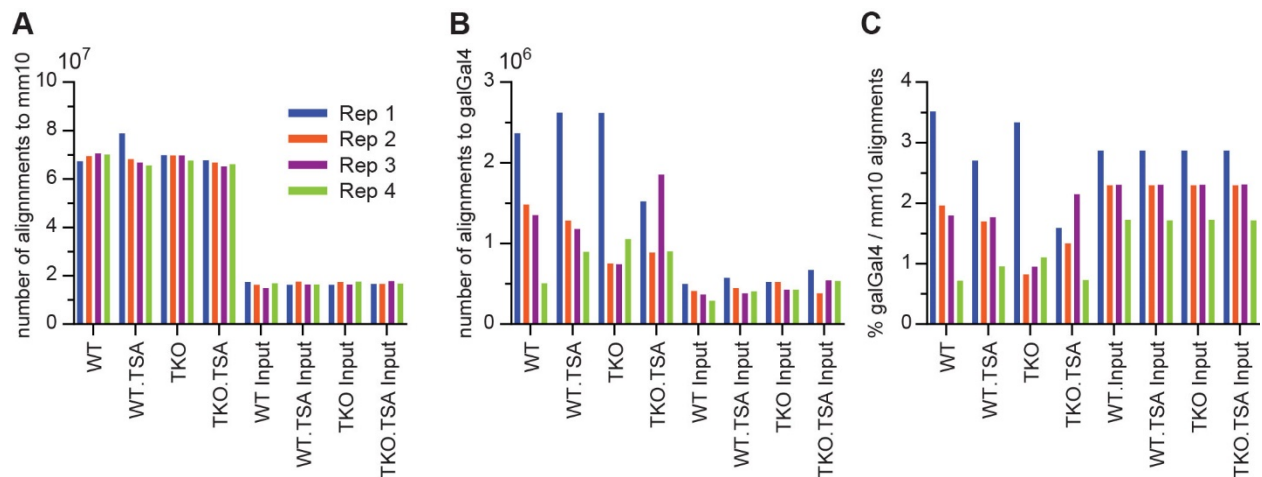


Figure M1. **A.** Number of uniquely mapped reads from each ATAC-seq library that align to the mm10 genome. **B.** Number of uniquely mapped reads from each ATAC-seq library that align to the galGal4 genome. **C.** Number of uniquely mapped reads from each ATAC-seq library that align to the galGal4 genome per 100 reads that align to the mm10 genome.

ATAC-seq 75 bp paired-end reads were trimmed using cutadapt (v1.16) to remove 3' adapter sequences then aligned to a concatenated mm10+galGal4 genome using bowtie2 (v2.2.5.0) with the options *--no-mixed* and *--no-discordant*. Reads that aligned to the exogenous galGal4 genome, reads with multiple alignments and fragments mapping to a custom “blacklist” of artificially high regions of the genome were discarded. Duplicate reads were removed. Alignment files were randomly down-sampled to retain an equivalent number of alignments in all samples. BigWig coverage files were generated using the Deeptools (Ramírez et al. 2016) function ‘bamCoverage’ (v2.4.2) with the option *-bs* set to 2. Enriched regions were identified using the ‘dpeak’ peak-calling algorithm of DANPOS v2.2.2 (Chen et al. 2013) with the options *-m*, *-kd 250*, *-kw 150* and *-p 1e-60*. The R package DiffBind (Stark and Brown, 2011) was used to identify ATAC-seq peak regions (THS) with significant changes in enrichment between experimental conditions. Significant changes were determined using the DBA_DESEQ2 method applying a fold change cut-off of 1.5 and a p-value threshold of 0.01 after multiple-testing correction.

Quantification of ATAC-seq reads within different categories of genomic regions

The genomic space was segmented into three categories of regions: “ATAC-seq peaks” that were called using DANPOS ‘dpeak’ as described above (THS); “Repmask regions” which included all genomic intervals that were annotated as repetitive based on the UCSC RepeatMasker annotation but excluded sequences that overlap with ATAC-seq peaks; and a “Unannotated” category which grouped together all remaining genomic intervals. The bedtools “intersect” command was used to count the number of alignments from each sequencing dataset that overlapped regions within each category. All alignments to the mm10 genome were counted, including those marked as having multiple alignments. Reads that overlapped two regions were counted only once: reads that only partially overlapped with an “ATAC peaks” region were included in this category while reads that overlapped a “Repmask” region and an “Unannotated” region were counted for the “Repmask” region only. We then calculated the percentage of total reads in the dataset that mapped to each category (Figure 2F).

To estimate the percentage of reads mapping outside of “ATAC peak” regions that would align to the “Repmask” and “Unannotated” regions by chance (random or expected distribution), we shuffled the reads with no alignment to ATAC peaks within the combined “Repmask” and “Unannotated” genomic space (using the bedtools “shuffle” command) and repeated the counting procedure (Figure S3C). We calculated “observed/expected” values based on the number of reads mapping to these two categories before and after shuffling (Figure 2G).

Calibrated ChIP-seq computational analysis

ChIP-seq 75-bp paired-end reads were aligned to a concatenated mm10+galGal4 genome using bowtie2 (v2.2.5.0) with the options *--no-mixed* and *--no-discordant*. Reads that had multiple alignments were discarded and duplicate reads were removed. GC composition of reads was similar between samples, conditions and different TF (Figure M2). The number of reads mapping to the mm10 or galGal4 genome were counted and used to calculate the ratio of galGal4 to mm10 reads for both IP and inputs samples. To adjust for the variation in mouse cells used for the four experimental conditions from a single biological replicate, the number of galGal4 reads in the input samples were randomly down-sampled so that the resulting mm10:galGal4 ratio in the four input samples was equal (See Figure S1C and Table S1). Enriched regions were identified using the MACS2 function 'callpeak' (v2.0.10, Zhang et al., 2008) with the genome size set to 1.87×10^9 . Peak calling was performed independently for each biological replicate using all uniquely mapped mm10 reads. Regions identified as enriched in more than one sample from any condition were kept for downstream analyses. Biological replicate samples showed higher sample-to-sample similarity than samples from different experimental conditions (Figure S1F).

To generate bigWig coverage files calibrated to the exogenous spike-in, we first randomly down-sampled BAM files containing the mm10 alignments for a particular sample based on the number of reads mapping to the galGal4 genome for that sample (in a manner which results in an equal mm10:galGal4 reads ratio in all samples). BigWig files were subsequently created using the Deeptools (Ramírez et al. 2016) function 'bamCoverage' (v2.4.2) with the options *-e* and *-bs 2*.

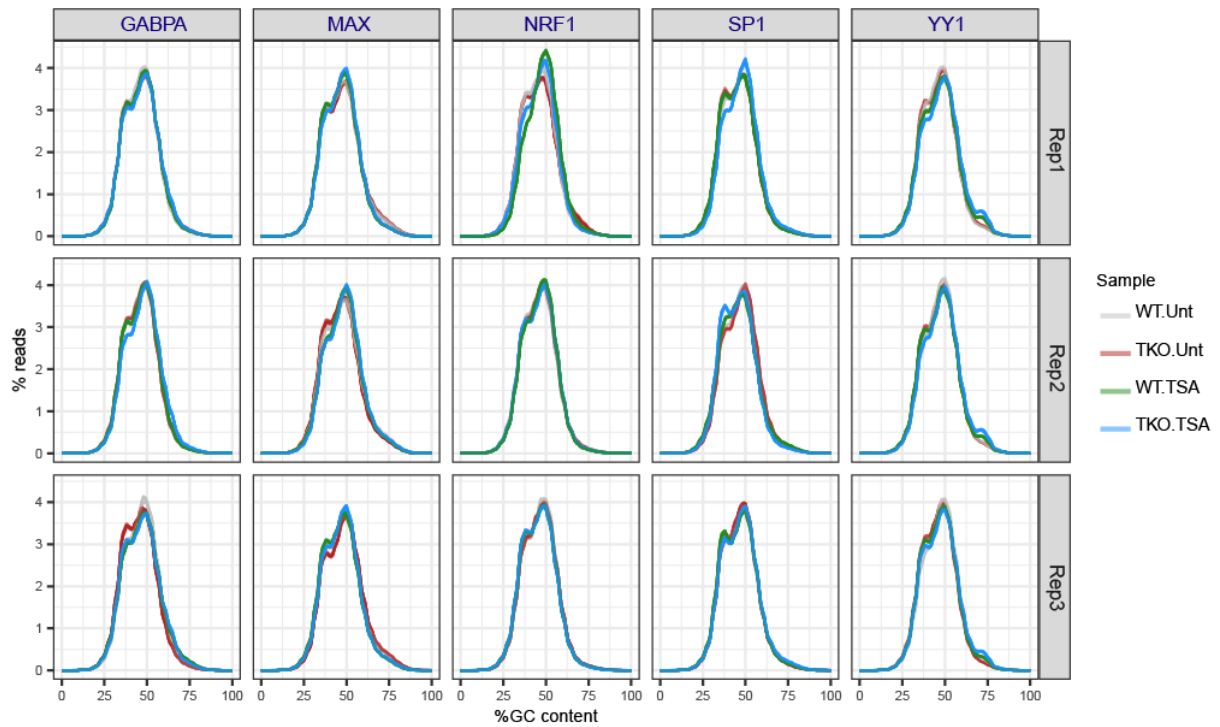


Figure M2. Average GC content for all sequencing reads. *Fastqc* and *multiqc* (Ewels et al. 2016) were used to generate these plots.

ChIP-seq computational analysis

ChIP-seq 75-bp paired-end reads were aligned to the mm10 genome using bowtie2 (v2.2.5.0) with the options *--no-mixed* and *--no-discordant*. Reads that had multiple alignments were discarded, duplicate reads were removed as well as fragments mapping to a custom “blacklist” of artificially high regions of the genome (derived from the ENCODE Consortium Project, 2012). All alignment files from the same transcription factor ChIP were randomly down-sampled to match the number of alignments in the minimal ChIP or input sample. BigWig coverage files were generated using the Deeptools (Ramírez et al. 2016) function ‘bamCoverage’ (v2.4.2) with the options *-e* and *-bs 2*. Enriched regions were identified using the ‘dpeak’ peak-calling algorithm of DANPOS v2.2.2 (Chen et al. 2013) with the options *-m*, *-kd 250*, *-kw 150* and *-p 1e-100*.

When visualising the GABPA ChIP-seq datasets we noticed that sequencing reads were not only enriched in a localised manner, forming the narrow peaks typically seen in transcription factor ChIP-seq data, but also within broader domains whose boundaries coincided closely with the edges of transcribed genes, particularly those with high levels of transcription. For the purposes of this study, we restricted our analysis to localised sites of GABPA ChIP-seq

enrichment. While the 'dpeak' peak-calling algorithm is designed to identify shorter enriched regions rather than larger domains, many of the focal "peaks" called using the parameters noted above were located within the broader GABPA enriched regions. In order to filter these out, we first identified genomic domains >2000bp broadly enriched with GABPA using the DANPOS function 'dregion' with the options *-m 1 -rd 1500 -rw 2000* and *-p 1e-50*. Subsequently, peaks called using dpeak (n = 10022) that overlapped with a called region were discarded unless they intersected a gene promoter (-1kb to 100bp around TSS). 4735 peaks remained, 74.7% contained a GABPA motif, compared to 56.9% before filtering.

The R package DiffBind (Stark and Brown, 2011) was used to identify ChIP-seq peak regions with significant changes in enrichment between experimental conditions. Significant changes were determined using the DBA_DESEQ2 method with the option *bFullLibrarySize=TRUE*, applying a fold change cut-off of 4 and a p-value threshold of 10^{-3} after multiple-testing correction (Bonferroni correction).

RNA-seq computational analysis

RNA-seq 51 bp paired-end reads were aligned to the mouse mm10 genome using TopHat (v2.0.13) (Trapnell et al. 2009) with the *--library-type fr-firststrand*, *--mate-inner-dist* set to 100 and coverage based searching for junctions disabled. FeatureCounts v1.4.5-p1 (Liao et al. 2014) was used to count aligned RNA-seq reads that mapped to the exons of annotated genes. Read counting was strand specific (option *-s* set to 2). Reads with a mapping quality below 1 were excluded. Differential analysis was performed using the R package edgeR (v3.22.3, McCarthy et al., 2012; Robinson et al., 2010) with raw read counts supplied as input. Genes with a low level of expression in all samples were filtered out using a counts per million (cpm) threshold of 1. The generalized linear model (GLM) likelihood test was used to determine differential expression between multiple groups of biological replicate samples. Genes or transcripts were considered as significantly differentially expressed if they had a p-value $< 10^{-5}$ after adjusting for multiple comparisons using the Benjamini and Hochberg procedure. Strand-specific bigWig coverage files were generated using the Deeptools (Ramírez et al. 2016) function 'bamCoverage' (v2.4.2) with the options *--normalizeUsingRPKM*, *--filterRNAstrand* and *-bs 1*.

The enrichment analysis of genes with tissue specificity was performed using the DAVID functional annotation tool (version 6.8, Huang et al., 2009a, 2009b). A functional annotation chart based on the Uniprot UP_TISSUE annotation was reported using default options.

Quantification of reads mapping to repetitive genomic features and differential analysis.

For individual ATAC-seq, ChIP-seq or RNA-seq sample, total read counts for every type of repetitive element were generated using a custom script, as detailed below. All bowtie2 alignments, including those ascribed to multi-mapping reads, were taken into account. The position and annotation of interspersed repeats and low complexity DNA sequences in the mm10 genome was obtained from UCSC (created using RepeatMasker, from the RepBase library of repetitive elements (Jurka et al. 2005)). The custom script makes use of the Samtools “view” command to count reads at genomic intervals, the Bedtools “nuc” command to determine the CpG content of genomic intervals and the ‘st’ tool (<https://github.com/nferraz/st>) to calculate CpG content and nucleotide substitution statistics across all copies of a repeat element type. The R package DESeq2 v1.20.0 (Love et al. 2014) was used to identify repeat element types that show significant differences in coverage between conditions. A p-value threshold of < 0.05 was used to identify significant differences after correction for multiple testing using the Benjamini & Hochberg procedure.

Common computational approaches

Metaplots used to display and summarise ATAC-seq, ChIP-seq or Native H3ac ChIP-seq signals were generated using Deeptools functions (Ramírez et al. 2016). As input, these tools were provided with genome coverage files (bigWigs) created after merging alignments from biological replicate samples.

The gene annotation files for mm10 or galGal4 used in this study were downloaded from UCSC. Genomic regions, for example ChIP-seq or ATAC-seq peaks, were annotated using the HOMER script `annotatePeaks.pl` (Heinz et al., 2010).

The HOMER script `findMotifsGenome.pl` (v4.7, Heinz et al., 2010) was used to define *de novo* motifs that were enriched within particular sets of target genomic intervals. The same script was used to screen for the enrichment of previously known or custom motifs. The maximum log-odds score for a given position weight matrix (PWM) motif within each interval in a .BED file was determined using the HOMER `annotatePeaks.pl` script with the *-mscore* option.

All boxplots were generated using the R package ggplot2. The median is indicated by a bar, the lower and upper hinges correspond to the first and third quartiles while the upper/lower whisker extend from the hinges to the largest value no further than 1.5x the inter-quartile range.

4. Supplementary references.

- Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**: 3047–3048.
- 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.
- Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**: 1–13.
- Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**: 462–467.
- Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.
- McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**: 4288–4297.
- 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.
- Stark R, Brown G. 2011. *DiffBind: differential binding analysis of ChIP-Seq peak data*.
- Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.