Acetylation of H3K115 at the nucleosome dyad is associated with fragile nucleosomes at active regulatory sites.

Dipta Sengupta[1#], Elias T. Friman[1], Robert S. Illingworth[2], Wendy A. Bickmore*[1], Yatendra Kumar*[1],

[1]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK

[2]MRC Centre for Regenerative Medicine, Institute for Regeneration and Repair, University of Edinburgh, Edinburgh EH16 4UU, UK

[#]Current address: MRC Centre for Regenerative Medicine, Institute for Regeneration and Repair, University of Edinburgh, Edinburgh EH16 4UU, UK

*Correspondence to:

Y.K. or W.A.B: MRC Human Genetics Unit, IGC, Crewe Road, Edinburgh EH4 2XU, UK

Tel: +44 131 651 8570

Wendy.Bickmore@ed.ac.uk

Yatendra.Kumar@ed.ac.uk

Keywords: CTCF, embryonic stem cells, nucleosome depleted regions, sub-nucleosomes, transcription start site,

Abstract

Acetylation of lysine residues in the tail domain of histone H3 is well characterized. However, acetyl-lysines in the histone globular domain also hold regulatory potential because of their impact on nucleosome dynamics and stability. In this study, we report the genome-wide distribution of acetylated H3 lysine 115 (H3K115ac), a residue on the lateral surface at the nucleosome dyad, in mouse embryonic stem cells. We find that H3K115ac is associated with highly active promoters, particularly those associated with CpG islands, and with enhancers. During differentiation H3K115ac is dynamic, changing in line with gene activation and chromatin accessibility. Most strikingly, unlike other commonly studied histone acetylation marks, H3K115ac is enriched on “fragile” nucleosomes within the nucleosome depleted regions of active promoters, and enhancers where it coincides with transcription factor binding. Additionally, we detect H3K115ac-marked fragile nucleosomes at sites most strongly occupied by CTCF, within the CTCF footprint and oriented relative to the CTCF motif. This unusual genomic distribution suggests that H3K115ac could have a role in nucleosome destabilization and that it might be a valuable marker for identifying functionally important regulatory elements in mammalian genomes.

Introduction

Post-translational modifications (PTMs) of histone tails play key roles in chromatin structure and gene regulation. Modifications of the N-terminal histone tails have attracted attention due to the unstructured nature of the domain (Ghoneim et al., 2021; Peng et al., 2021) and accessibility to chromatin modifiers and readers. However, more recently it has become apparent that modifications in the globular domains of core histones are also associated with various chromatin functions (Tropberger & Schneider, 2013; Pradeepa et al., 2016; Zorro Shahidian et al., 2021), but they remain relatively understudied.

Modification of residues on the lateral surface of the histone octamer are of particular interest because these can directly affect nucleosome-DNA interactions and nucleosome dynamics (Cosgrove et al., 2004; Lawrence et al., 2016). For example, acetylation of lysine-56 on histone H3 (H3K56ac) is predicted to increase accessibility at both the entry and exit sites of the nucleosome, impacting gene expression, DNA replication and repair (Rajagopalan et al., 2017; Rodriguez et al., 2019). Lysine-64 on H3 is close to the dyad axis of the nucleosome and its acetylation (H3K64ac) decreases nucleosome stability (diCerbo et al., 2014). Lysine-122 of H3 (H3K122) is also located close to the dyad axis and its acetylation impacts nucleosome stability and weakens histone-DNA interactions (Manohar et al., 2009; Rajagopalan et al., 2017) with a direct consequence on transcription from a chromatinised template in vitro (Tropberger et al., 2013). H3K122ac also makes nucleosomes more susceptible to disassembly by chromatin remodellers (Chatterjee et al., 2015). Other acylation events, such as succinylation of H3K122 similarly impact nucleosome stability (Zorro Shahidian et al., 2021).

H3K115 is located in the globular domain of H3 at the nucleosome dyad. Both copies of H3K115 in the histone octamer are juxtaposed to each other on the lateral surface of the nucleosome in close contact with the overlying DNA (Figure 1A). Because of this, acetylation of H3K115 has a high potential to disrupt histone-DNA interactions and therefore to influence nucleosome dynamics, nucleosome breathing and histone release (Zhou et al., 2019). Consistent with this, acetylation of H3K115 (H3K115ac) favours nucleosome disassembly and nucleosome sliding in vitro, by itself or together with H3K122ac (Manohar et al., 2009; Simon et al., 2011).

Promoters have a specific nucleosome organization: two well-positioned nucleosomes (+1 and -1) flank a nucleosome free (NFR) or nucleosome depleted region (NDR) immediately upstream of the transcription start site (TSS). This view has been revised with the detection of micrococcal nuclease- (MNase) and salt-sensitive nucleosome species within the NDRs termed as fragile, unstable, or partially unwrapped nucleosomes in high-throughput sequencing data generated after limited chromatin digestion with MNase (Henikoff et al., 2011; Kent et al.,

2011). Fragile nucleosomes can be isolated by MNase digestion under mild conditions and contain the H3.3 and H2AZ histone variants (Jin and Felsenfeld, 2007). In mouse embryonic stem cells (mESCs), chemical cleavage methods have also identified sub-nucleosomes at TSSs (Ishi et al., 2015; Voong et al., 2016). To the best of our knowledge, fragile nucleosomes have not yet been associated with a specific histone modification in mammalian cells.

Previous research identified H3K64ac and H3K122ac at the TSS of active genes and at some active enhancers in mammalian cells (Tropberger et al., 2013; diCerbo et al., 2014; Pradeepa et al., 2016). In this study we report the distribution of H3K115ac across the mammalian genome in mESCs, and during the differentiation to neural progenitor cells, revealing the focal association of H3K115ac with fragile nucleosomes at active promoters, enhancers, and sites important for genome architecture. Our data suggest that H3K115ac could be an important new tool for the functional annotation and understanding of mammalian genomes.

Results

H3K115ac is associated with highly active promoters and enhancers

To determine where H3K115 acetylated nucleosomes can be found across the mammalian genome we performed native chromatin immunoprecipitation and next generation sequencing (ChIP-seq) on mouse embryonic stem cells (mESCs). We used a commercially available antibody and confirmed its specificity for H3K115ac and not for unmodified H3K115, H3K115R, or H3K122ac (Figure S1A). Analysis of the distribution of H3K115ac across the mouse genome relative to other active and inactive histone modifications showed that H3K115ac correlates with H3K64ac and H3K122ac - H3 lateral surface modifications associated with active promoters and enhancers (Pradeepa et al., 2016), with H3K27ac, and with chromatin accessibility as assayed by ATAC-seq (Figure S1B). However, H3K115ac exhibits a much greater degree of overlap with active regions, particularly promoters, compared to H3K122ac and H3K27ac, as defined by a chromHMM segmentation approach. Approximately one third of H3K115ac peaks are associated with active TSS (Figure 1B).

Levels of H3K115ac at TSS correlate with levels of transcription, as assayed by 4SU-seq (Benabdallah et al., 2019) (Figure 1C), with H3K115ac most enriched on the most active subset of TSSs (Figure 1D). In mESCs there are fewer called peaks for H3K115ac (~39000), compared to other histone H3 lysine acetylation marks (H3K122ac ~80,000; H3K64ac ~65,000; H3K27ac >100,000).

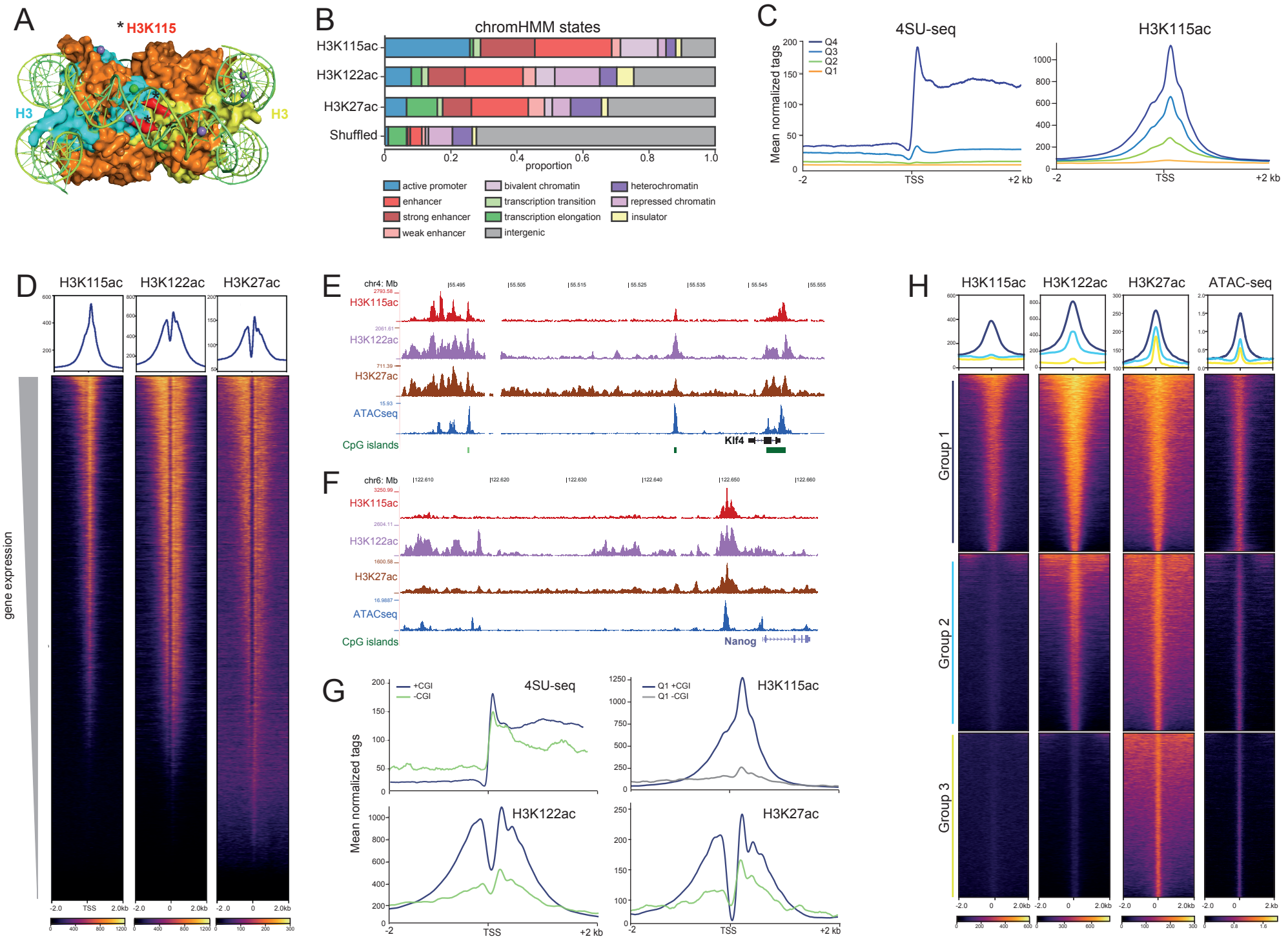

A
* H3K115
H3
H3
B
chromHMM states
H3K115ac
H3K122ac
H3K27ac
Shuffled
0
0.2
0.4
0.6
0.8
1.0
proportion
active promoter
enhancer
strong enhancer
weak enhancer
bivalent chromatin
transcription transition
transcription elongation
intergenic
heterochromatin
repressed chromatin
insulator
C
4SU-seq
H3K115ac
Mean normalized tags
Q4
Q3
Q2
Q1
-2
TSS
+2 kb
D
H3K115ac
H3K122ac
H3K27ac
gene expression
-2.0
TSS
2.0kb
E
chr4: Mb
H3K115ac
H3K122ac
H3K27ac
ATACseq
CpG islands
Klf4
F
chr6: Mb
H3K115ac
H3K122ac
H3K27ac
ATACseq
CpG islands
Nanog
G
4SU-seq
+CGI
-CGI
H3K115ac
Q1 +CGI
Q1 -CGI
H3K122ac
H3K27ac
Mean normalized tags
-2
TSS
+2 kb
H
H3K115ac
H3K122ac
H3K27ac
ATAC-seq
Group 1
Group 2
Group 3
-2.0
0
2.0kb

Figure 1

Figure 1. H3K115ac is associated with active chromatin.

A) Nucleosome structure looking down at the dyad axis (modified from PDB-5X7X, Taguchi et al., 2017). The two histone H3 molecules are shown in cyan and yellow, other histones are in orange and DNA in green. N-terminal histone tails are hidden. Both copies of H3K115 (red and asterisked) are juxtaposed at the dyad axis close to the overlying DNA. B). Proportions of H3K115ac, H3K122ac and H3K27ac peaks that overlap the genomic segments defined by chromHMM (Ernst and Kellis, 2012; Pintacuda et al., 2017). C) Average profiles (mean normalised reads) of, (left) 4SU-seq centred at protein-coding TSS, divided into quartiles (Q1-Q4) of gene expression in mESCs (4SU-seq tags in TSS +500bp); (right) H3K115ac signal around TSS quartiles defined by 4SU. D) Heatmap of ChIP-seq for H3K115ac (this study), H3K122ac and H3K27ac (Pradeepa et al., 2016) in mESCs with respect to TSS of the top 50% of genes by expression and sorted by decreasing gene expression. E) UCSC genome browser screenshot of *Klf4* locus showing ChIP-seq data for H3K115ac, H3K122ac, H3K27ac and ATAC-seq in mESCs. CpG islands (CGI) are indicated. Genome co-ordinates (Mb) are from the mm10 assembly of the mouse genome. F) As for (E) but at the *Nanog* locus. G) Mean normalised reads of, (upper left) 4SU-seq centred at TSS of top 50% genes by expression. Genes are divided into TSS that do (CGI+) or do not (CGI-) overlap with CpG islands (4SU-seq tags in TSS +500bp region). Upper right and lower panels show average profiles of H3K115ac, H2K122ac and H3K27ac read-density at the TSS of the top 25% highly expressed genes (Q4 in panel C), divided into TSS that do (CGI+) or do not (CGI-) overlap with CpG islands. The higher H3K115ac read-density at CGI+ TSS is not due to sample size (Wilcox, p-value < $2.2e^{-16}$, normalized read coverage within a 500bp window centred at TSS). H) Heatmap showing the coverage of H3K115ac, H3K122ac, H3K27ac and ATAC-seq centred on promoter distal accessible peaks (putative enhancers) in mESCs. Data is grouped in enhancers marked by all three H3 acetylation marks (Group 1; H3K27ac+ H3K122ac+ H3K115ac+), just H3K27ac together with H3K122ac (Group 2), or H3K27ac alone (Group 3).

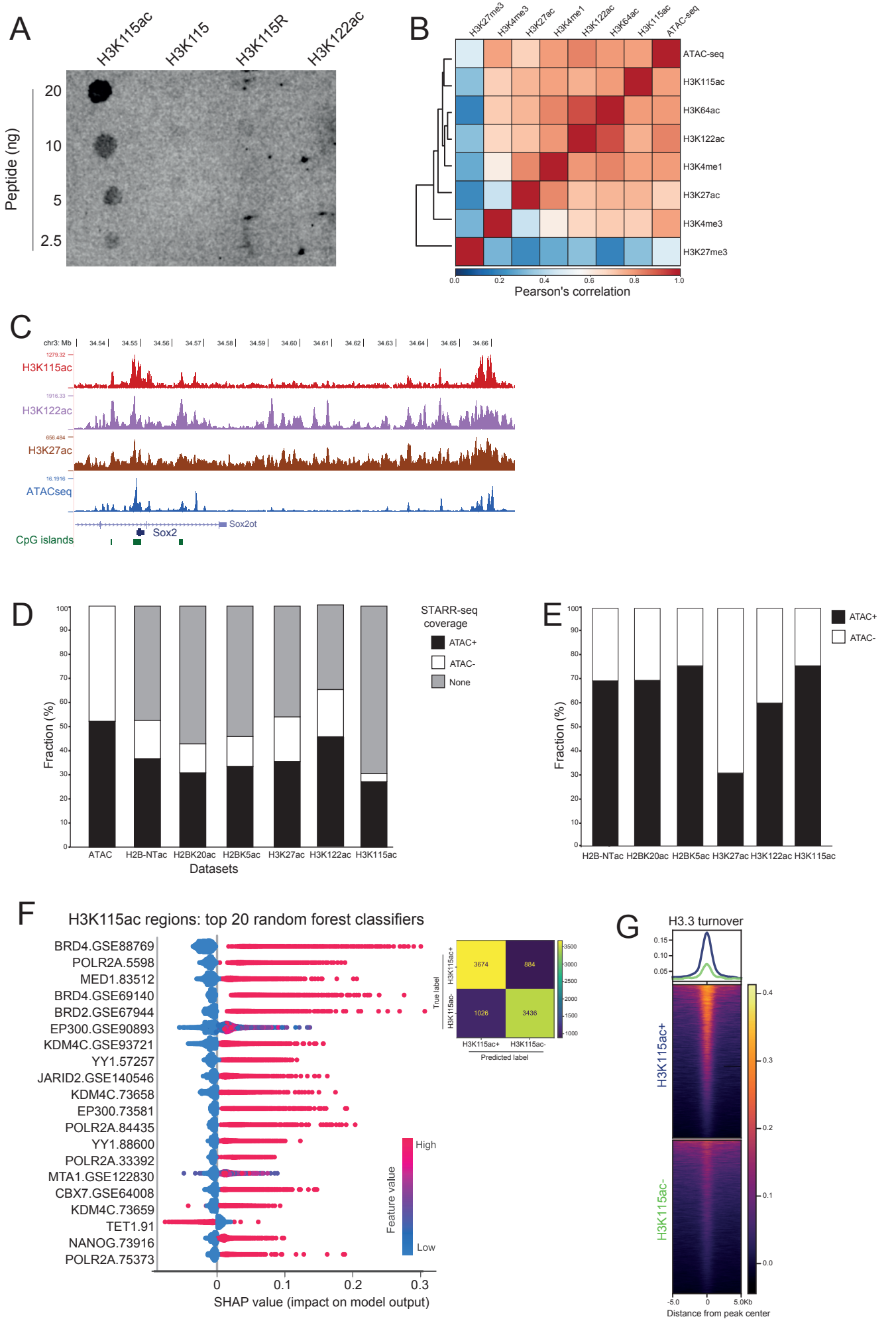

A
H3K115ac
H3K115
H3K115R
H3K122ac
Peptide (ng)
20
10
5
2.5
B
H3K27me3
H3K4me3
H3K27ac
H3K4me1
H3K122ac
H3K64ac
H3K115ac
ATAC-seq
Pearson's correlation
C
chr3: Mb
H3K115ac
H3K122ac
H3K27ac
ATACseq
Sox2ot
Sox2
CpG islands
D
STARR-seq coverage
ATAC+
ATAC-
None
Fraction (%)
ATAC
H2B-NTac
H2BK20ac
H2BK5ac
H3K27ac
H3K122ac
H3K115ac
Datasets
E
ATAC+
ATAC-
Fraction (%)
H2B-NTac
H2BK20ac
H2BK5ac
H3K27ac
H3K122ac
H3K115ac
F
H3K115ac regions: top 20 random forest classifiers
BRD4.GSE88769
POLR2A.5598
MED1.83512
BRD4.GSE69140
BRD2.GSE67944
EP300.GSE90893
KDM4C.GSE93721
YY1.57257
JARID2.GSE140546
KDM4C.73658
EP300.73581
POLR2A.84435
YY1.88600
POLR2A.33392
MTA1.GSE122830
CBX7.GSE64008
KDM4C.73659
TET1.91
NANOG.73916
POLR2A.75373
High
Feature value
Low
SHAP value (impact on model output)
True label
H3K115ac+
H3K115ac-
3674
884
1026
3436
Predicted label
G
H3.3 turnover
H3K115ac+
H3K115ac-
Distance from peak center

Figure S1

Figure S1. ***H3K115ac is associated with active chromatin. Related to Figure 1.***

A) Dot blot for specificity of H3K115ac antibody against H3 peptides as indicated. B) Pearson's correlation matrix with hierarchical clustering in mESCs. Correlation is computed for read counts in 10 Kb windows across the genome for ATAC-seq data, ChIP-seq data for active and repressive histone H3 marks (H3K122ac/H3K27ac/H3K27ac: GSE66023; H3K4me3: GSM1003756; H3K4me1: GSM1003750; H3K27me3: GSM1276707), and for H3K115ac. C) UCSC genome browser screenshot of *Sox2* locus showing ChIP-seq data for H3K115ac, H3K122ac, H3K27ac and ATAC-seq in mESCs. CpG islands (CGI) are indicated. Genome co-ordinates (Mb) are from the mm10 assembly of the mouse genome. D) Overlap of STARR-seq based enhancers (Peng et al., 2020) with different histone aceyl-lysine peaks in mESCs. H2B-NTac refers to regions marked with at least one of the acetyl-lysines in the N-terminal of histone H2B (K5, K11, K12, K16 and K20, Narita et al., 2023). Enhancers are classified as corresponding to accessible (ATAC+) or closed (ATAC-) sites in the mESC genome. Overlap with various histone acetylations is plotted as percentage of total. None; fraction of enhancers with no overlap in STARR-seq. E) Percentage of overlap with ATAC-seq peaks in mESCs for the different histone acetylation peaks as in (D). F) A Random Forest model was trained to predict H3K115ac-positive versus -negative peaks (but marked by H3K27ac and/or H3K122ac) based on ChIP-seq overlap. SHAP values were calculated to determine the impact on the model (top 20 features shown, left). The prediction confusion matrix is shown for the test data (right). G) Heatmap depicting H3.3 turnover index (Deaton et al., 2016; GSM2080325_TI_ES.wig in GEO accession GSE78910) centered at the same peaks as in (F) for H3K115ac-+ve and -ve regions.

Although H3K115ac is enriched, along with H3K64ac and H3K122ac, at the TSS of *Klf4* (Figure 1E) and *Sox2* (Figure S1C), it is notably low at the active *Nanog* promoter (Figure 1F). The *Sox2* and *Klf4* promoters are located within CpG islands (CGIs), but the *Nanog* promoter lacks a CGI (Figure 1F). This prompted us to examine H3K115ac at CGI and non-CGI TSS genome wide. For the most highly expressed genes (Q4 in Figure 1C), transcription levels are slightly lower for nonCGI TSS than for CGI TSS (Figure 1G), and H3K122ac and H3K27ac enrichment somewhat reduced at these non-CGI TSS compared to CGI TSS. In contrast, H3K115ac is very low at non-CGI active TSS (Figure 1G). Overall, ~60% of CGI TSSs in mouse ESCs are marked with H3K115ac compared to just 2% for non-CGI promoters.

The largest proportion of H3K115ac peaks are identified as enhancer elements in overlap analysis with chromHMM segmentation (Figure 1B). We did not find any promoter-distal DNase hypersensitive sites marked by H3K115ac exclusively. H3K115ac peaks overlap with a subset of enhancers that are marked by H3K27ac and H3K122ac and that exhibit the highest chromatin accessibility as assayed by ATAC-seq compared to enhancers marked with H3K27ac and H3K122ac, alone and in combination (Group 1 in Figure 1H). Sequences capable of driving enhancer-like activity in mESCs have been identified using the STARR-seq assay (Peng et al., 2020). Only half of these sequences exhibit chromain accessibility in their native context in the mESC genome, as determined by ATAC-seq (ATAC+, Figure S1D). While peaks of acetyl-lysines generally exhibit significant overlap with accessible enhancers, H3K115ac peaks shows the smallest overlap with ATAC- enhancers compared with H3K27ac and various acetyl-lysines in histone H2B (Narita et al., 2023), even though the proportion of total H3K115ac and H2B acetylated lysines peaks that are ATAC+ do not drastically differ (Figure S1E). Thus, compared to H3K27ac or H2B acetylation, H3K115ac is a better discriminator in identifying whether the endogenous counterparts of STARR-seq active sequences are accessible by ATAC-seq. Association of H3K115ac with active transcriptional regulation is further substantiated by the enrichment of transcription factors, co-activators and RNA polymerase II at H3K115ac regions as revealed by a random forest approach using published ChIP-seq datasets from mESCs (Figure S1F). Furthermore, H3K115ac-marked distal elements display high levels of H3.3 turnover (Deaton et al., 2016) (Figure S1G). Together, these features suggest that H3K115ac may contribute to chromatin accessibility and nucleosome dynamics at regulatory sites.

Changes in H3K115ac during differentiation correlate with transcription and chromatin accessibility.

To ascertain how H3K115ac distribution changes as genes are activated and repressed, we differentiated 46c mESC cells to neural progenitor cells (NPCs) monitoring differentiation by Sox1-GFP fluorescence. ChIP-seq for H3K115ac and ATAC-seq were carried out after 7 days of differentiation and nascent RNA was quantified (4SU-seq) on days 0, 3 and 7 (Benabdallah et al., 2019). Substantial redistribution of H3K115ac is observed during this differentiation, with only 36% of the H3K115ac peaks in mESCs retained in NPCs, and 25682 new peaks gained (Figure 2A).

Promoter regions were classified into those that gained or lost H3K115ac during differentiation if the TSS (-500 to +500 bp) showed at least a 2-fold change in H3K115ac signal compared with day 0. Transcription from TSS that gain H3K115ac is significantly upregulated during differentiation and down-regulated from TSS that lose H3K115ac (Figure 2B). This indicates that the gain of H3K115ac correlates with transcriptional activation during differentiation.

Enhancers make up the majority of the H3K115ac peaks (~60%) compared to promoters (~35%). H3K115ac peaks shared between ESC and NPC are more commonly classified as promoters, whereas dynamic peaks (either ESC-specific or NPC-specific) are more frequently classified as enhancers (Figure 2C). GREAT analysis of the annotation of nearby genes (McLean et al., 2010) show that ESC-specific H3K115ac peaks are enriched in terms associated with embryonic development, stem cell development and maintenance, whereas NPC-specific peaks are enriched in terms associated to neuronal development (Figure S2A). During differentiation, H3K115ac marked dynamics enhancers also show changes in chromatin accessibility, as assayed by ATAC-seq (Figure 2D). Sequentially acting SOX TFs are known to drive gene expression during neuronal differentiation (Bergsland et al., 2011) and indeed NPC-specific H3K115ac peaks are strongly enriched in binding motifs for the SOX family of pioneer transcription factors (Figure S2B). These analyses indicate that H3K115ac is associated with the activation and the accessibility of enhancers bound by pioneer TFs during differentiation.

H3K115ac is associated with sub-nucleosome sized fragments

We noted that, over promoters, H3K115ac localization, appears much more focal and centred over the TSS compared with both H3K27ac and H3K122ac that exhibit a depletion of signal over the TSS (Figure 1D). This suggests that H3K115ac may be associated with MNase-sensitive nucleosomes in the NDR of active promoters. These nucleosomes can be detected

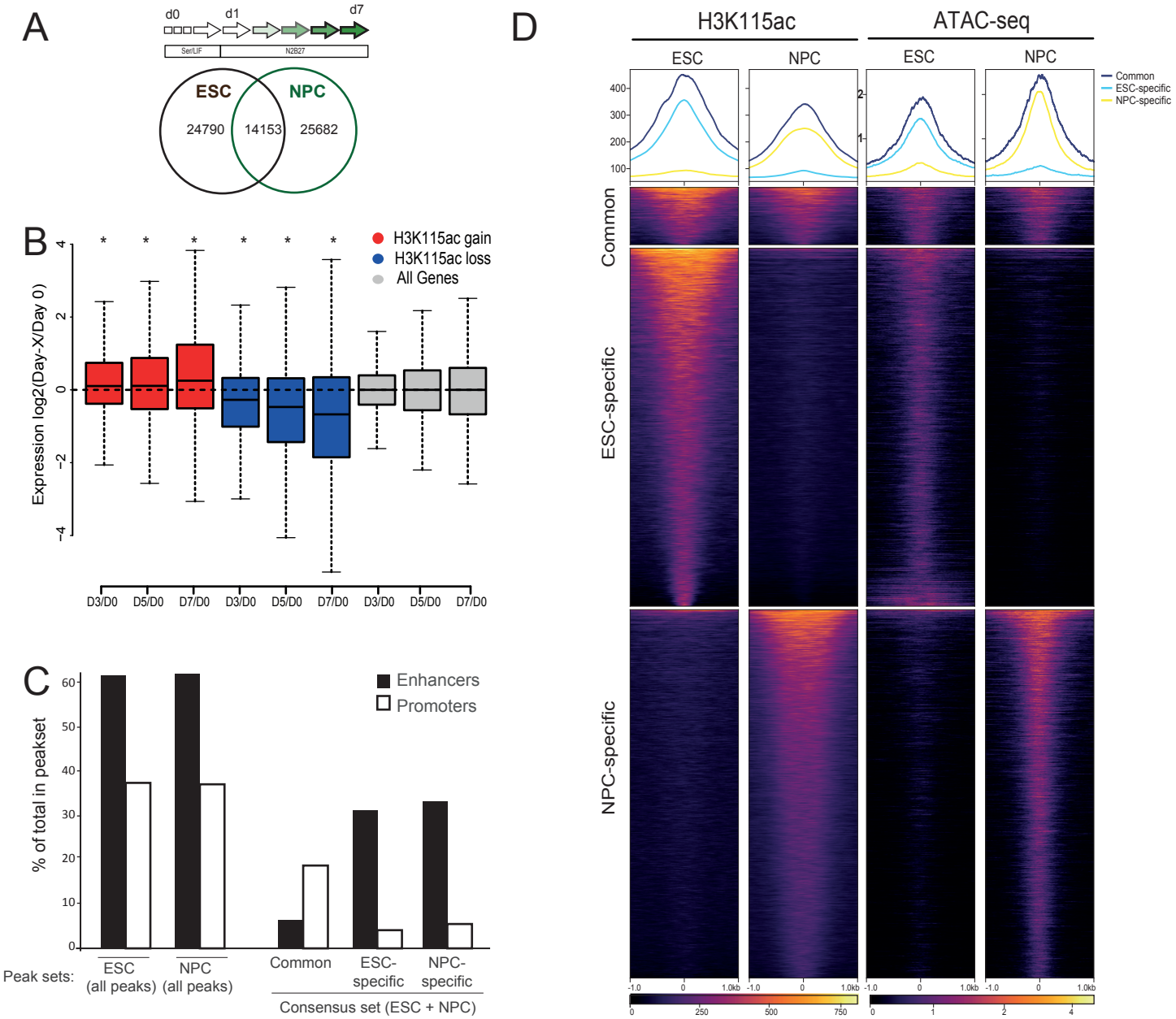
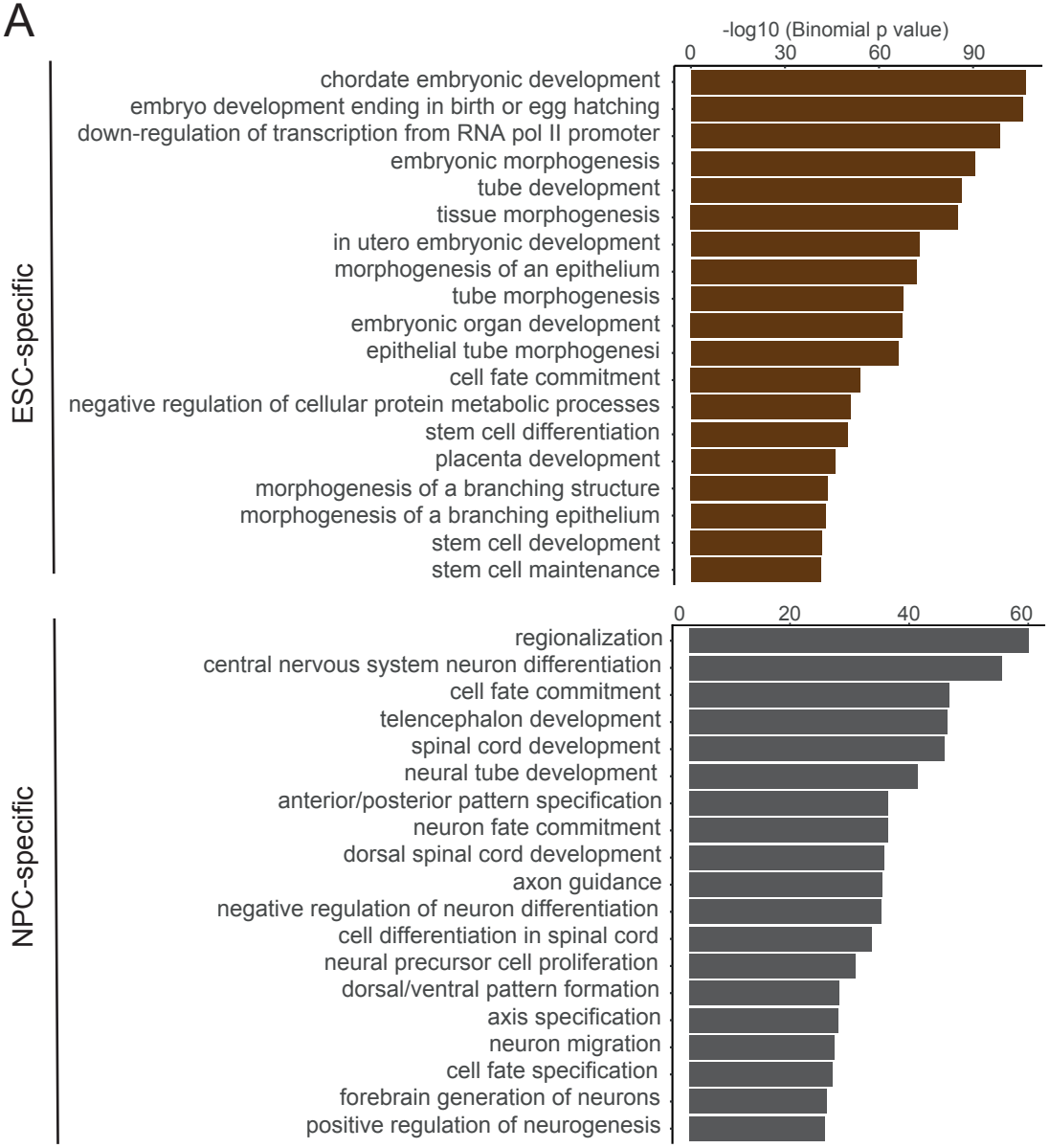


Figure 2. ***H3K115ac is highly dynamic during differentiation.***

A) Venn diagram depicting the number of common and unique peaks of H3K115ac between undifferentiated mESCs and Sox1-eGFP+ve NPCs differentiated for 7 days. B) Boxplots displaying the changes in activity (4SU-seq) of promoters that gain or lose H3K115ac across the 7 days of differentiation. Log2 fold change is shown relative to day 0. *Paired Wilcox, p <0.01. C) Histogram showing % of promoter- or enhancer-associated peaks in mESC and NPCs. Nearly 33% of all H3K115ac peaks are associated with promoters in both ESCs and NPCs with 70% of them in common between ESCs and NPCs. In contrast, the majority of dynamic H3K115ac peaks are associated with enhancers (ESC-specific; n=1881, NPC-specific; n=2811). D) Heatmaps showing H3K115ac and ATAC-seq signal for common and dynamic enhancers between ESC and NPC. Loss/gain of H3K115ac correlates with loss/gain in chromatin accessibility.

A



B

Rank	Motif	TF
1		Sox3
2		Sox10
3		Sox4
4		Sox2
5		Sox15
6		Rfx2
7		RFX (HTH)
8		Sox6
9		X-box (HTH)
10		Sox17

Figure S2. ***NPC-specific H3K115ac peaks correspond to binding motifs for neuronal transcription factors. Related to Figure 2.***

A) GO term (Biological Process) enrichments from peak annotations in GREAT (McLean et al., 2010, http://great.stanford.edu/public/html/) for the H3K115ac peaks assigned to ESC-specific (top) or NPC-specific (bottom) enhancers.

B) Top 10 transcription factor motifs detected in H3K115ac marked NPC-specific enhancer set.

as sub-nucleosomes (footprint < 147bp) obtained from small fragments sequenced from partial MNase digestion (Carone et al., 2014) or mono-nucleosomes (footprint ~ 147 bp) obtained under low salt conditions (Jin & Felsenfeld, 2007). Both species of nucleosome particles are referred to as fragile nucleosomes (Carone et al., 2014; Ishi et al., 2015; Voong et al., 2016).

Our ChIP protocol involves mild MNase digestion of native chromatin without size selection, thus, our ChIP libraries could contain fragments originating from fragile nucleosomes, but single-end sequencing data cannot be used to decipher nucleosome boundaries at base pair resolution. To investigate the distribution of H3K115ac in relation to nucleosome size, we repeated the H3K115ac and H3K27ac ChIP on partially MNase-digested native chromatin from mESCs and sequenced the libraries in paired-end mode. Normalized coverage of end-to-end ChIP fragments showed that, in comparison with the H3K27ac profile which shows enrichment over the +1 and -1 nucleosomes, H3K115ac is highly enriched within the NDR over the TSS of highly active genes (Figure 3A). Comparing the fragment length distributions of the H3K115ac ChIP library, the corresponding input MNase-digested library, and the fragment length distribution of the H3K27ac ChIP library, H3K115ac ChIP preferentially selects nucleosomes from the pool of input nucleosomes that are approximately 15bp shorter than the average and that are shorter than those nucleosomes purified by H3K27ac ChIP (Figure 3B). Around the most active TSSs (Q4), signal from small fragments accumulate inside the TSS NDR (Figure S3A). Signal from fragments > 150bp exhibits the characteristic nucleosome depletion, with the NDR becoming more pronounced with increasing fragment lengths. We used this rationale to split ChIP fragments into sub-nucleosomes (length ≤150bp) and mono-nucleosomes (151-230bp) (Figure S3B). H2K27ac ChIP is depleted in these smaller nucleosome particles compared with the input, whereas H3K115ac is enriched (Figure S3C). Compared with input and H3K27ac ChIP, H3K115ac ChIP fragments are also depleted in dinucleosome-length fragments, consistent with their sensitivity to MNase digestion (Figure S3C).

Nucleosomes associated with A/T-rich sequences are known to be more sensitive to MNase compared to sites of higher G/C content (Chereji et al., 2019). We noticed that a significant proportion of H3K27ac marked nucleosomes in our MNase ChIP-seq data are also sub-nucleosome-sized. To investigate whether H3K115ac and H3K27ac marked nucleosomes might originate from the same population of sequences, we compared the A/T content of sub- and mono-nucleosome-sized fragments carrying these histone modifications. H3K115ac sub-nucleosomes have significantly higher A/T content than H3K27ac marked sub-nucleosomes (Figure 3C, $p < 0.01$, Wilcox Test), showing that they originate from different populations of fragile nucleosomes. Similar observations were made when comparing A/T content between mono-nucleosomes carrying H3K115ac and those marked with H3K27ac (Figure 3C). This bias does not originate from differences in MNase digestion levels, because the matched input

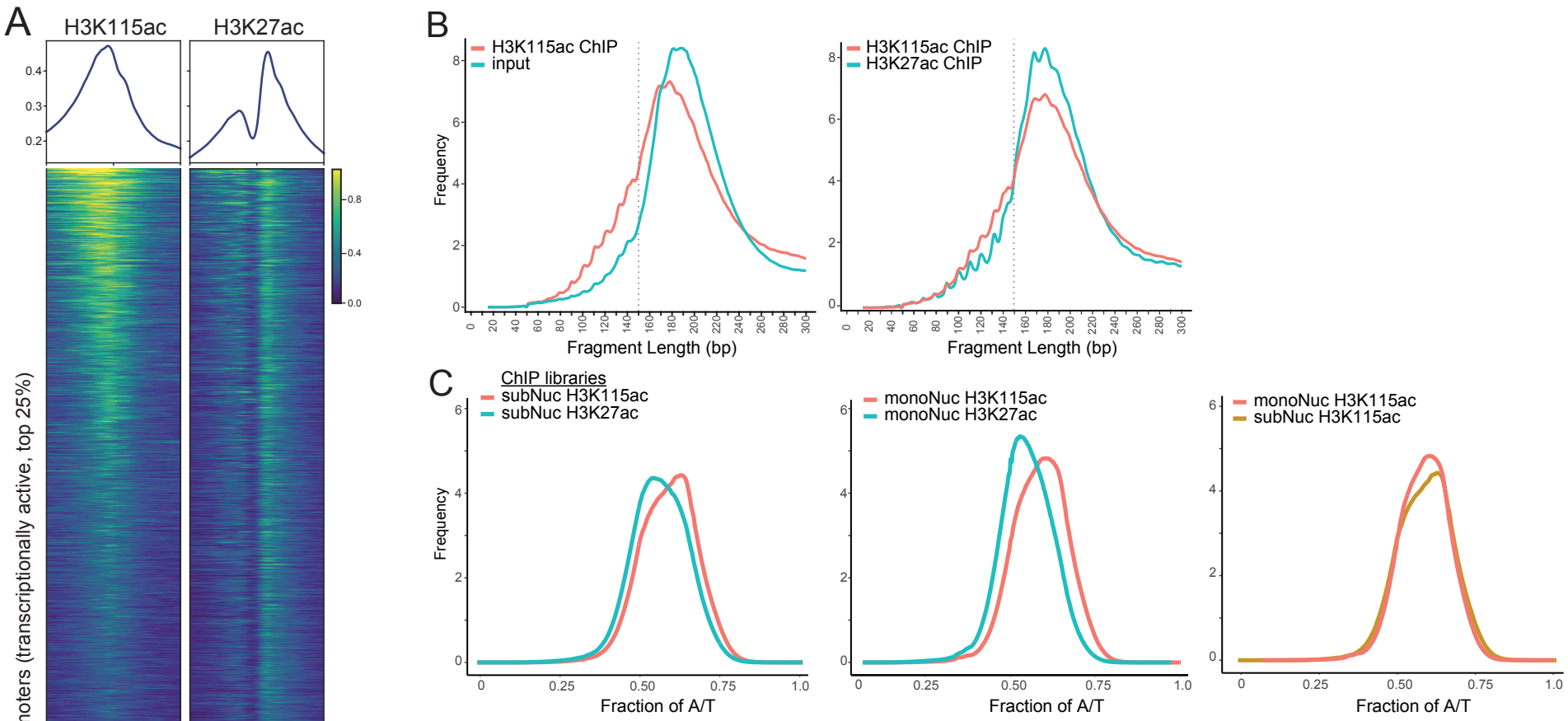


Figure 3 ***H3K115ac is associated with sub-nucleosome sized fragments***

A) Heat map and average profile plots from native ChIPs-seq for H3K115ac and H3K27ac in mESCs using partial MNase digestion and paired-end sequencing. Coverage is generated after piling up the whole span of ChIP fragments, normalized to library size, and plotted upstream and downstream of the TSSs associated with high transcriptional activity (Q4). H3K115ac signal spans the nucleosome depleted region (5' NDR) and TSS. In contrast, H3K27ac is depleted over the 5' NDR. B) Frequency of fragment lengths (bp) from paired-end libraries for; (left) native ChIP of H3K115ac and its input sample, (right) H3K115ac and H3K27ac ChIP libraries. Distribution is scaled to library size to enable comparison. C) Density plots of A/T content from ChIP libraries; (left) H3K115ac marked subnucleosomes have a higher A/T content than subnucleosomes marked with H3K27ac (Wilcox test, $p < 0.01$), (centre) mono-nucleosomes marked with H3K115ac have higher A/T content than those marked with H3K27ac (Wilcox test, $p < 0.01$), (right) the A/T content of sub- and mono-nucleosomal particles marked with H3K115ac are not significantly different (Wilcox test, $p > 0.05$).

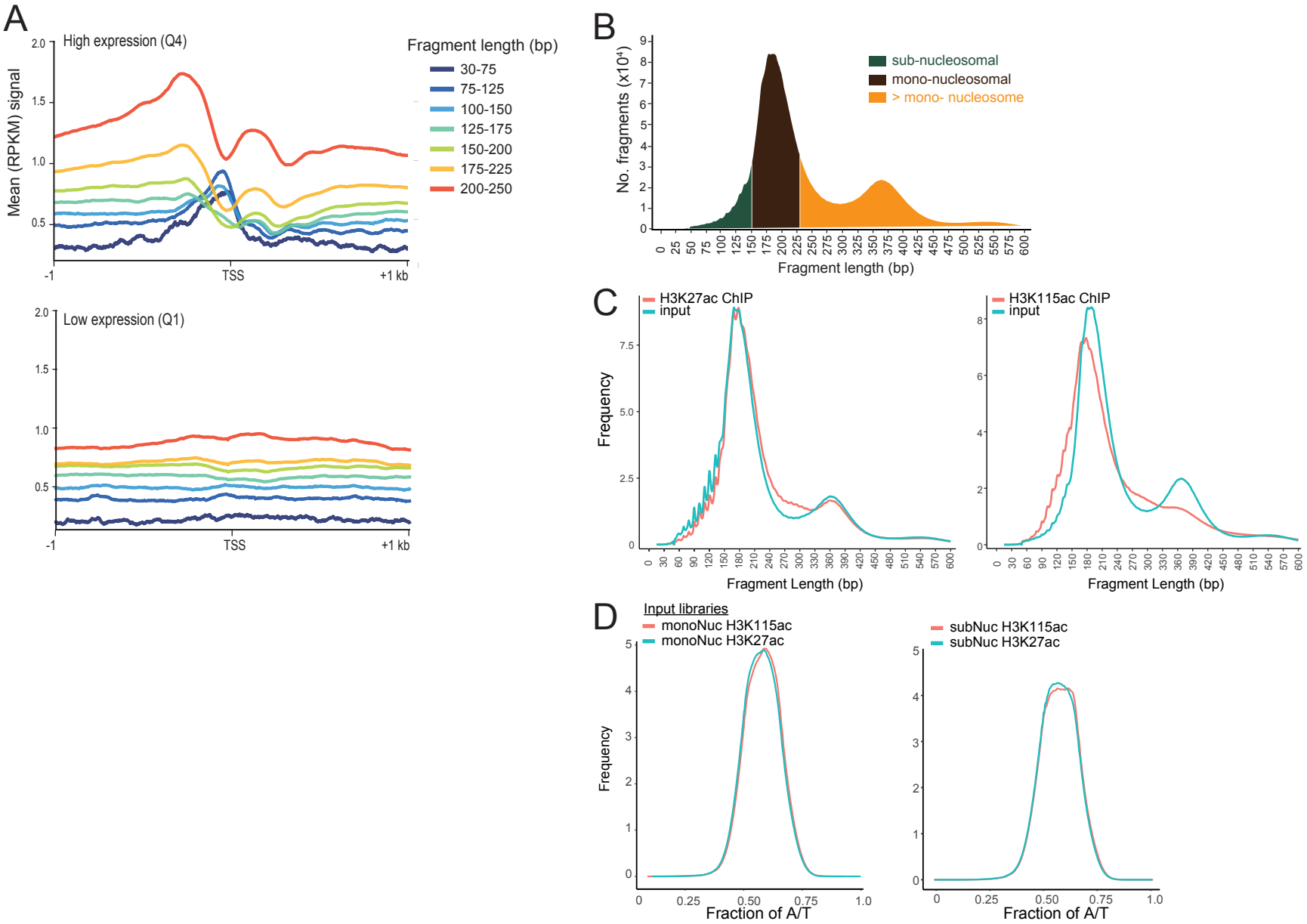


Figure S3. ***H3K115ac is associated with sub-nucleosome sized fragments. Related to Figure 3***

A) Mean coverage of fragments in MNase-digested input libraries at (top) highly active TSSs (Q4) and (bottom) minimally active TSSs (Q1) in mESCs binned into different fragment lengths (bp). B) Schematic of selected fragment lengths to define sub-nucleosomes and mono-nucleosomes from input libraries. C) Fragment length distribution (bp) of; (left) H3K27ac and (right) H3K115ac ChIP fragments plotted against matched input samples. D) Density profiles of A/T content for; (left) mono-nucleosomes and (right) sub-nucleosomes from the matched input libraries used for ChIP-seq of H3K115ac and H3K27ac. A/T content does not differ between different input libraries (Wilcox test, $p > 0.01$).

libraries do not show such a difference (Figure S3D). H3K115ac modified sub-nucleosomes and mono-nucleosomes have the same A/T content, indicating that they likely originate from comparable genomic regions (Figure 3C).

H3K115ac marks fragile-nucleosomes at active TSSs

To explore H3K115ac ChIP fragment size in relation to promoter NDRs and flanking nucleosomes (Figure S4A), we calculated the distance between the centre of fragments from the native ChIP libraries and the nearest TSS. We plotted these distances and fragment lengths as contours of highest density regions for a conservative set (see Methods) of TSSs in mESCs. The resulting 2D contour V-plots (Henikoff et al., 2011) from input libraries show the promoter NDR flanked by positioned nucleosomes with smaller nucleosome particles of varying sizes inside the NDR (Figure 4A). H3K27ac signal is most abundant on the +1 and -1 nucleosomes with little signal detected inside the NDR. In contrast, H3K115ac signal is enriched on the -1 nucleosome and on subnucleosome-sized fragments inside the NDR (Figure 4A, bottom panel). We then defined NDRs by taking nucleosome positions from published datasets (West et al., 2014; Voong et al., 2016), scaled them to the same length, and plotted the coverage of input (MNase-seq), H3K27ac and H3K115ac fragments (Figure 4B). H3K27ac shows a general depletion from the NDR, but both H3K115ac modified mono- and subnucleosome-sized fragments are enriched within the NDR. Notably, within the NDR, subnucleosome-sized H3K115ac MNase fragments are TSS proximal whereas mononucleosome fragments are displaced toward the -1 nucleosome position (Figure 4B). Subnucleosomal H3K115ac remains TSS-proximal even when the NDR is wide (Figure S4B,C). This suggests that H3K115ac deposition is tightly regulated and does not extend into the NDR just because there is more space available there.

We next analysed H3K115ac localisation relative to TSS with respect to the transcriptional activity of the associated gene. As expected from classical MNase-seq experiments, highly active genes exhibit nucleosome depletion at the promoter with strongly phased nucleosomes on both sides (Figure 4C, Q4, top). Sub-nucleosomal, as well as mono-nucleosomal, H3K115ac modified fragments inside the NDR correlate with promoter activity (Figure 4C). In contrast, H3K27ac levels correlate to transcription, but on nucleosomes at the start of the gene body (Figure S4D). Interestingly, compared to mono-nucleosomal H3K27ac, sub-nucleosomal H3K27ac shows stronger phasing in the gene body (Figure S4D), suggesting that sub-nucleosomal sized H3K27ac particles may result from remodeling of the nucleosomes by elongating RNA polymerase II.

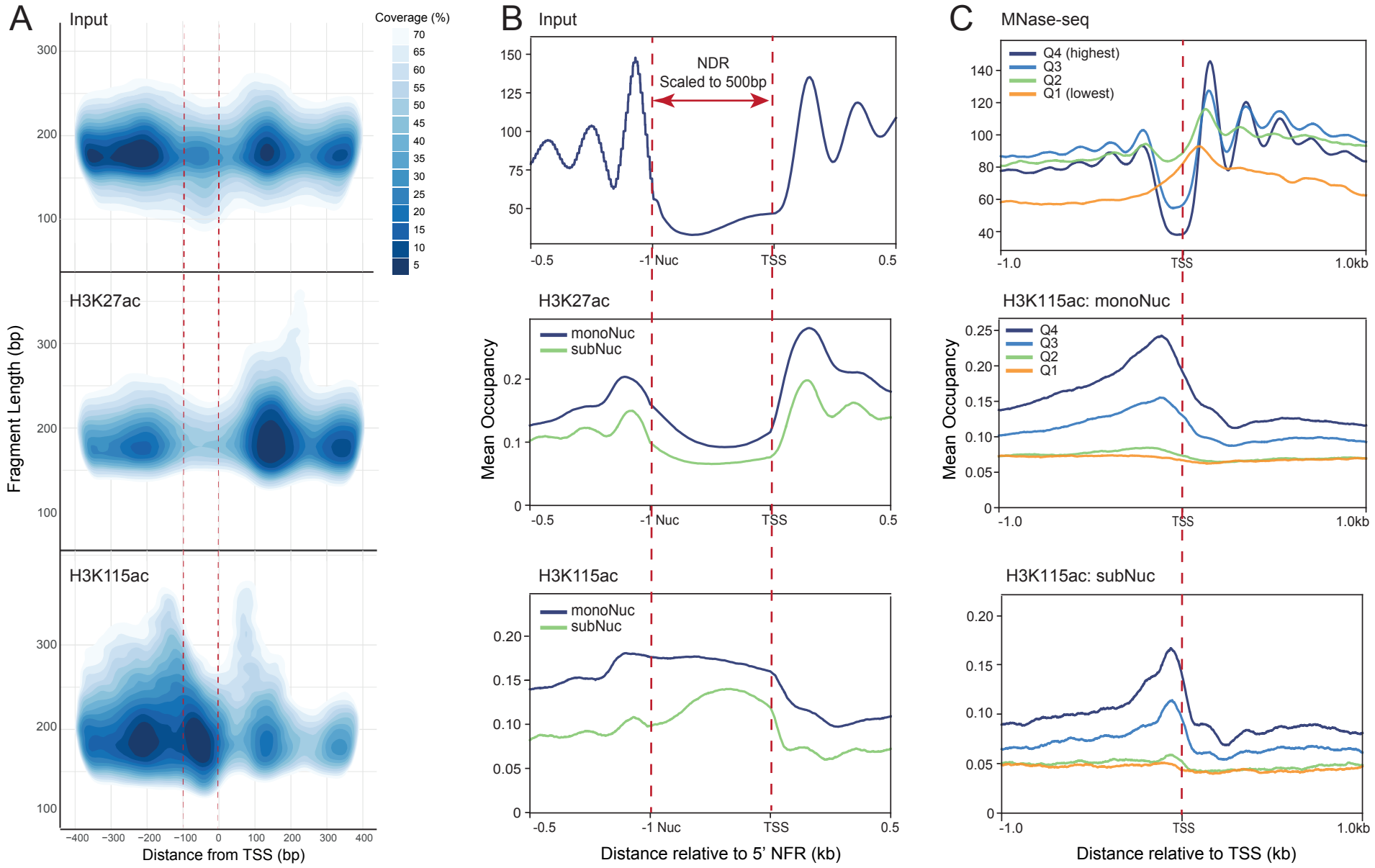


Figure 4 ***H3K115ac is associated with fragile nucleosomes in the NDR of highly transcribed genes***

A) Contour plots depicting the high density regions of chromatin fragments around TSS (+/-400bp) as a function of fragment length (bp) generated from; (top) Input MNase library, (centre) H3K27ac ChIP and (bottom) H3K115ac ChIP. Coverage refers to the proportion of ChIP-fragments. The region from 100bp upstream to the TSS is indicated with red dashed lines. B) Mean nucleosome occupancy around mouse TSSs plotted with respect to the NDRs, scaled to a length of 500bp. MNase-seq data (top, West et al., 2014) were used to define NDR as the region between the 3' boundary of -1 nucleosome and the TSS. The mean occupancy of (centre) H3K27ac or (bottom) H3K115ac ChIP-seq fragments is split into sub- and mono-nucleosomes around the scaled NDRs. C) Mean nucleosome occupancy from MNase-seq data (top) around mouse TSS divided into quartiles of nascent transcription in mESCs from high (Q4) to low (Q1). The mean H3K115ac signal arising from the (centre) mono-nucleosomal or (bottom) sub-nucleosomal immunoprecipitated fragments around the TSSs are shown split into quartiles of nascent transcription.

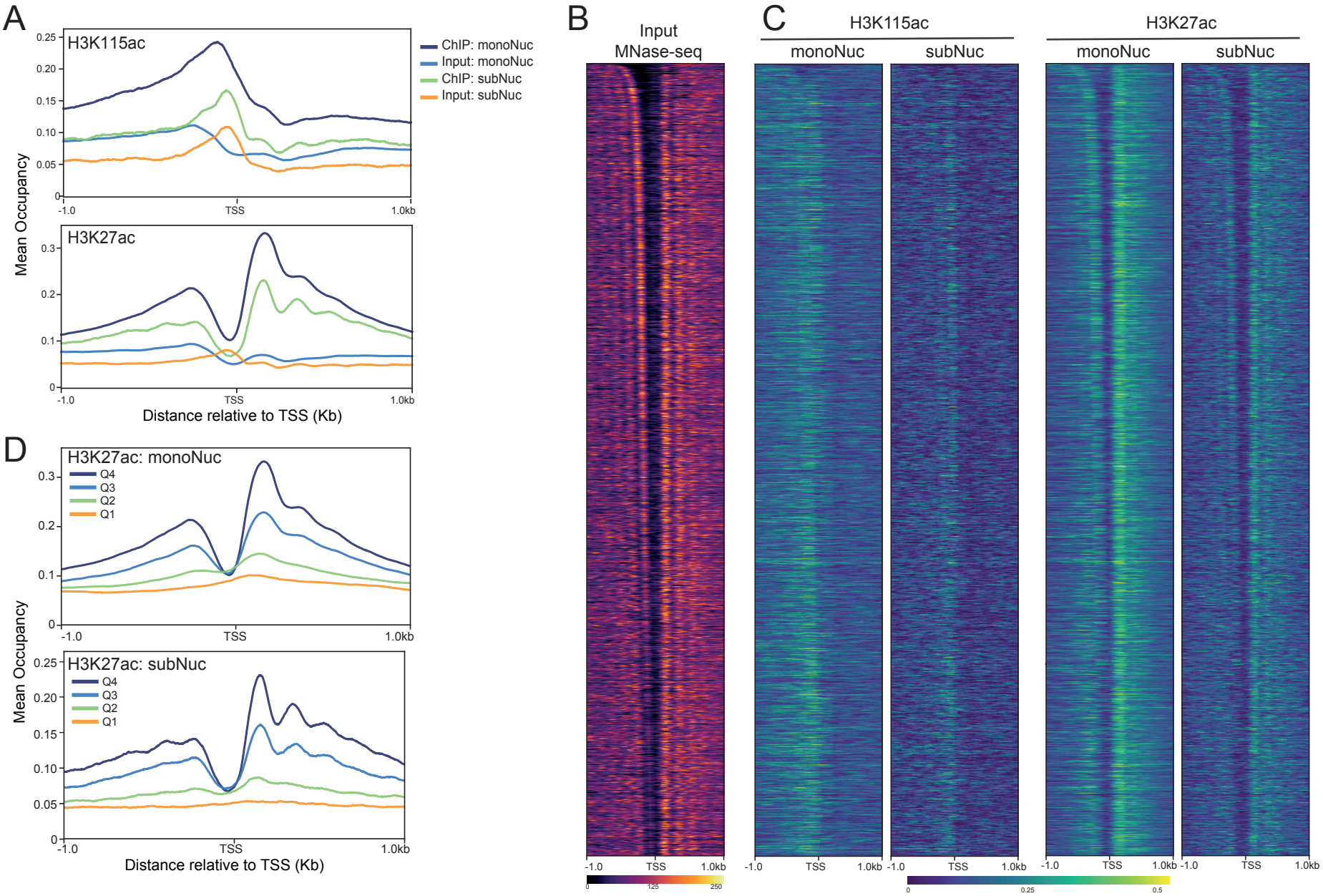


Figure S4. ***H3K115ac is associated with fragile nucleosomes in the NDR of highly transcribed genes. Related to Figure 4.***

A) Mean occupancy of H3K115ac (top) or H3K27ac (below) ChIP-seq from mESCs plotted around (+/-1kb) TSS. Sub-nucleosomal and mono-nucleosomal signals are plotted with together with the profile from their respective input libraries. B) Heatmap displaying occupancy and positioning of MNase-resistant nucleosomes around TSSs sorted by decreasing NDR width in mESCs. Only highly expressed genes (Q4) with no other coding- or noncoding TSS within 500 bp are included. C) Heatmaps of mono- and sub-nucleosomal signal from (left) H3K115ac and (right) H3K27ac ChIP-seq around TSSs in the same order as in (B). D) Mean H3K27ac signal arising from the (top) mono-nucleosomal and (below) sub-nucleosomal chromatin fragments around TSSs split into quartiles of nascent transcription as in Fig. 4C.

H3K115ac correlates with OCT4 and NANOG occupancy at mESC enhancers

H3K115ac marked enhancers exhibit the highest chromatin accessibility compared to enhancers marked by H3K27ac and H3K122ac (Figure 1H). This emphasizes the potential role of H3K115ac, positioned on the nucleosome lateral surface, in facilitating nucleosome remodeling and accessibility at active enhancers. Enhancers are activated by the binding of pioneer TFs, which can recognise and bind their recognition sequences buried in nucleosomes resulting in nucleosome destabilisation and priming the site for the binding of additional TFs (Vernimmen and Bickmore, 2015).

We analysed H3K115ac and H3K27ac paired-end ChIP-seq data at mESCs enhancers occupied by the pioneer pluripotency TF OCT4 (King and Klose, 2017; MacCarthy et al., 2022). H3K27ac occupancy correlates with OCT4 binding in regions immediately flanking the OCT4 binding site but is depleted over the binding site itself (Figure S5A). In contrast, H3K115ac profiles also correlate with OCT4 occupancy levels, but both mononucleosomal and subnucleosomal H3K115ac marked fragments are enriched at the OCT4 motif itself (Figure S5A). Enhancers with different levels of OCT4 binding show the corresponding expected levels of chromatin accessibility, as assayed by ATAC-seq (Figure S5B). A dip in accessibility centered at the OCT4 motif likely indicaties the protection offered by OCT4 binding.

As OCT4 binding alters nucleosome structure (Sinha et al., 2023), facilitating co-operative binding by additional OCT4 molecules, we also analysed enhancers with two OCT4 motifs (n=650) scaling the distance between the two motifs to a window of the same width. Mean occupancy plots centred on this window show that the presence of two OCT4 motifs in close proximity creates sharp protected footprints in ATAC-seq data, with high chromatin accessibility also between the two sites (Figure 5, bottom panel). Absence of stable nucleosomes between the two motifs is evident in MNase-seq, with strongly positioned nucleosomes flanking the ATAC-seq accessible region. H3K27ac is enriched on these flanking nucleosomes but is depleted between the OCT4 motifs (Figure 5, middle panel). In contrast, H3K115ac enrichment spans the entire region between the two motifs on both mono- and sub-nucleosome-sized fragments (Figure 5, top panel). As expected, given the co-operativity between pluripotency TF binding, ATAC-seq (Figure S5C), H3K27ac and H3K115ac (Figure S5D), are all more highly enriched at enhancers occupied at both OCT4 and NANOG, compared to sites occupied by either TF alone. These data show that H3K115ac, and its association with sub-nucleosomes, is linked to binding of pioneer TF complexes at enhancers.

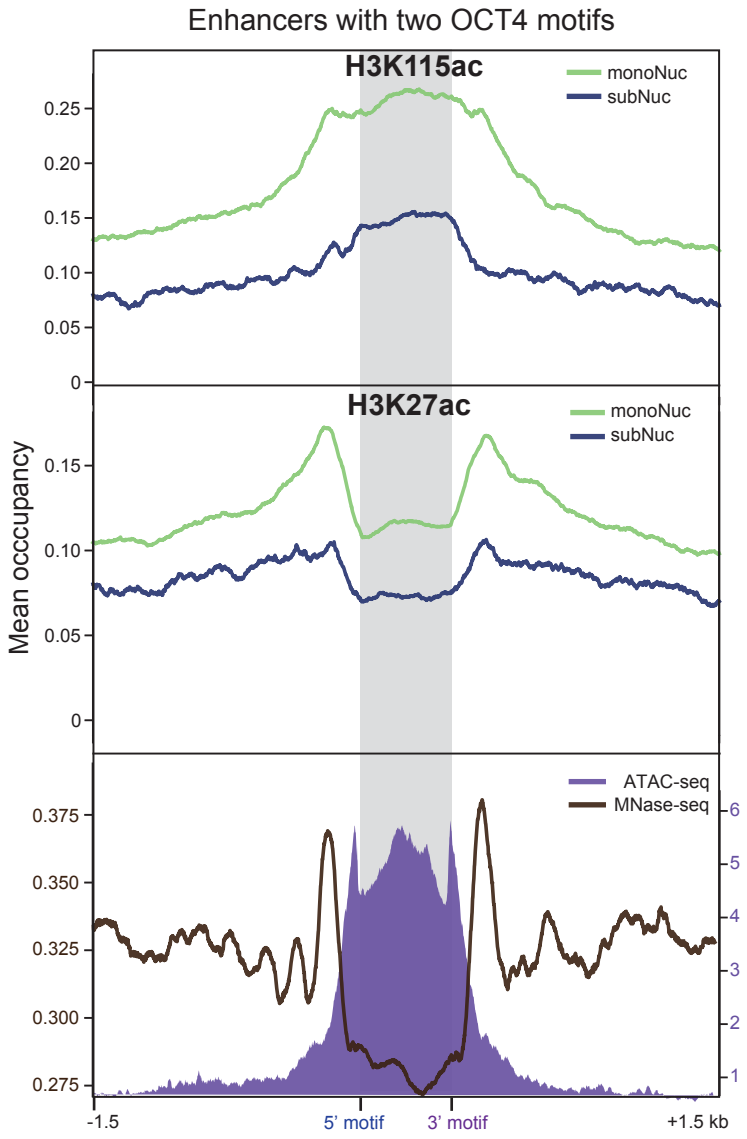
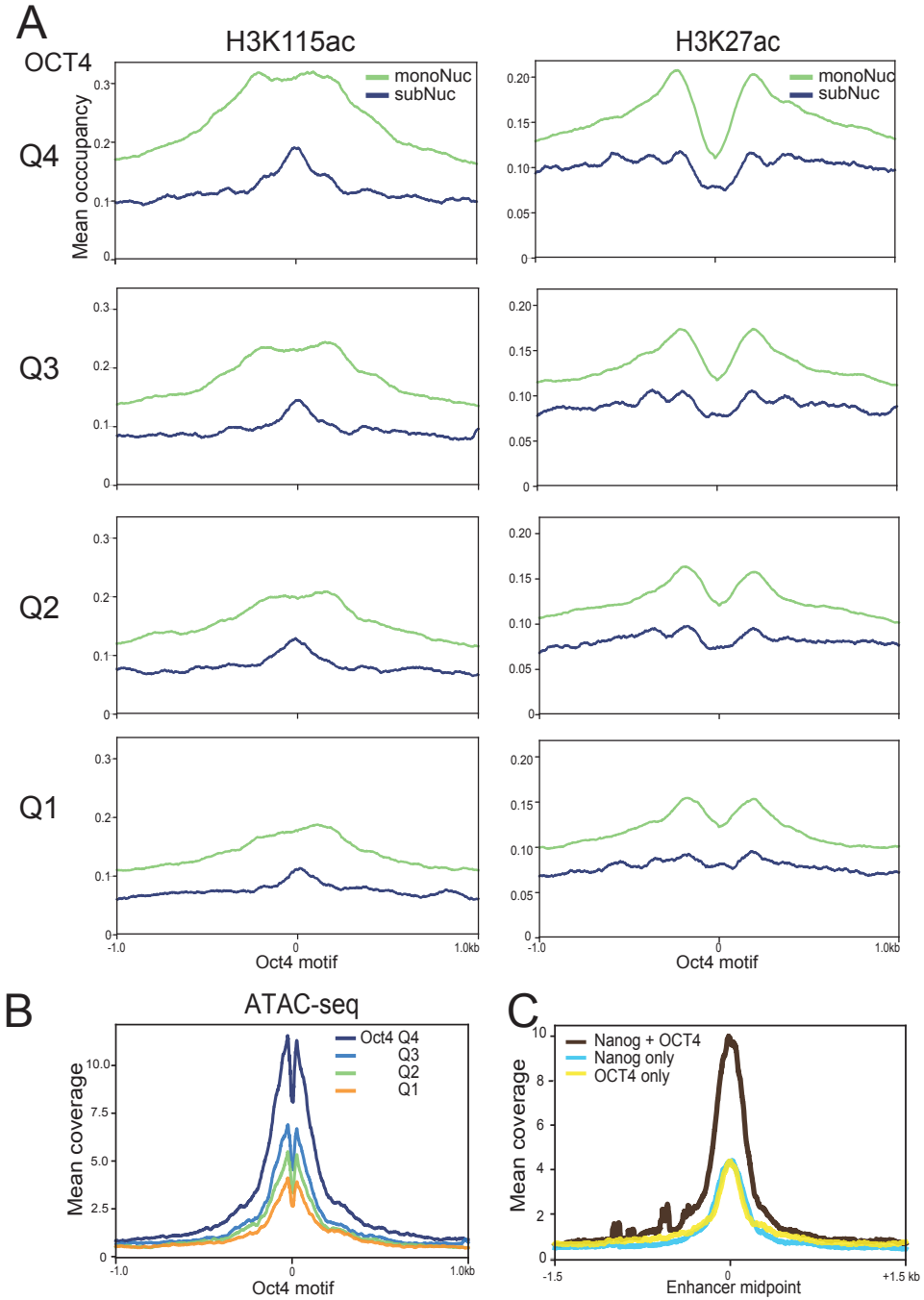


Figure 5. H3K115ac correlates with pioneering activity at mESC enhancers

For mESC enhancers selected based on the presence of two OCT4 motifs within the Tn5 accessible region, the region between the two motifs is scaled to the same length (shaded grey region). Top; H3K115ac ChIP-seq, middle; H3K27ac ChIP-seq on mononucleosomes (green lines) or subnucleosome-sized fragments (blue lines). Bottom panel; ATAC-seq, MNase-seq signal with y-axis values for MNase-seq on the left and for ATAC-seq on the right.



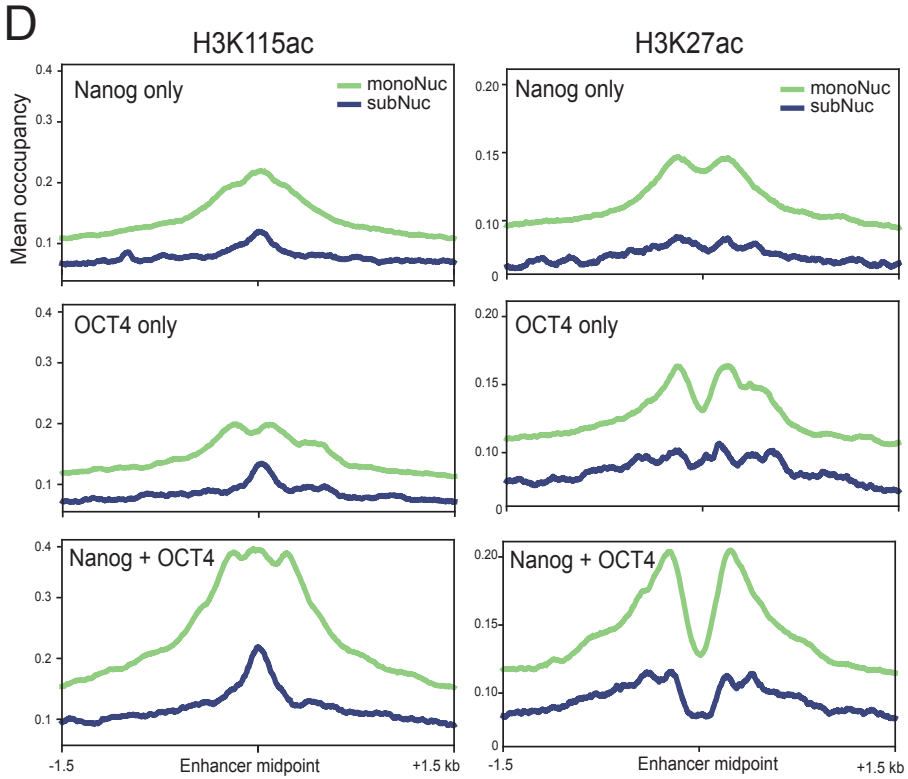


Figure S5 ***H3K115ac and H3K27ac correlates with pioneering activity at mESC enhancers. Related to Figure 5.***

A) H3K115ac (left) and H3K27ac (right) paired-end ChIP signal around OCT4-occupied enhancers in mESCs divided into quartiles (Q1-Q4, low to high) of OCT4 ChIP-seq peak strength. ChIP-seq reads are separated according to fragment length into mononucleosomes (green; monoNuc) and subnucleosomes (blue;subNuc). B) ATAC-seq signal around OCT4-occupied enhancers for the quartiles of OCT4 ChIP-seq peak strength as in (A). C) Mean ATAC-seq signal around mESC enhancers occupied by OCT4 and NANOG alone or in combination. Enhancer midpoint is taken as the centre of the chromatin accessibility peak. D) H3K115ac (left) and H3K27ac (right) paired-end ChIP-seq signal plotted for putative enhancers in mESCs occupied by (top) NANOG alone, (middle) OCT4 alone and (bottom) OCT4+NANOG. Signal is plotted separately for subnucleosomes (blue) and mononucleosomes (green).

H3K115ac marks fragile nucleosomes at and around CTCF binding sites

CTCF binding is known to create arrays of well positioned nucleosomes on both sides of its binding site (Fu et al., 2008; Owens et al., 2019). Though CTCF binding was originally thought to exclude nucleosomes, fragile nucleosomes and very small sub-nucleosomal fragments have been detected at CTCF binding sites (Voong et al., 2016; Klein et al., 2023) (preprint; Nocente et al., 2022). To examine the distribution of H3K115ac and its relationship to sub-nucleosomes at and around CTCF sites, we called CTCF motifs and intersected them with published CTCF ChIP-seq peaks in mESC (Mas et al., 2018). We selected peaks with a single CTCF motif, divided these into quartiles of CTCF ChIP-seq peak scores, and used the coordinates of the oriented CTCF motifs in V-plots and heatmaps. We noticed a wider range of nucleosome sizes in MNase input libraries in Q4 (top 25%) compared to Q1 of CTCF sites (Figure 6A, left panels). In the regions immediately flanking (+100bp and -100bp) the CTCF motif we detect fragments larger than a nucleosome, which might signify MNase protection mediated by interactions between nucleosomes and CTCF or CTCF-interacting proteins. Within the CTCF NDR we detect a range of smaller fragments around the motifs with strongest CTCF binding. Enrichment of H3K115ac and H3K27ac both correlate with the strength of CTCF binding. H3K27ac enrichment is largely limited to mononucleosomes immediately flanking the CTCF site (Figure 6A, right; Figure S6A). Mononucleosomes strongly marked by H3K115ac are also found at these same positions, but in addition H3K115ac is detected on fragments larger than nucleosomes located just inside the canonical flanking nucleosomes (Figure 6A, middle). Moreover, sub-nucleosomal fragments marked by H3K115ac can be detected in the centre of the CTCF site itself (Figure 6A,B).

V-plots reveal a subtle asymmetry in the distribution of sub-nucleosomal H3K115ac fragments within the CTCF site, skewed toward the 3' end of the CTCF motif (Figure 6A, middle). H3K27ac on flanking nucleosomes is also skewed in the same direction (Figure S6A). To determine whether canonical nucleosomes exhibit a similar trend, we plotted nucleosome occupancy around CTCF motifs from a classical MNase-seq experiment (West et al., 2014). While upstream nucleosomes maintain phasing, nucleosomes downstream of the CTCF motif get fuzzier with decreasing CTCF binding score (Q1) and the +1 nucleosome begins to encroach into the CTCF NDR (Figure 6C, Figure S6B). This directionality is lost when CTCF motif orientation is randomized, with enchroachment into the NDR from both sides at CTCF Q1 sites (Figure S6B, C). Directionality bias is not linked to TAD boundary formation. Using proximity to the nearest TAD boundary (Bonev et al.,2017) as a measure of the probability with which a given CTCF site participates in TAD boundary formation, and dividing CTCF motifs into deciles of this distance, for each decile we sorted for descending CTCF binding score and plotted occupancy of canonical nucleosomes (Figure S6B: right). The fuzziness of downstream

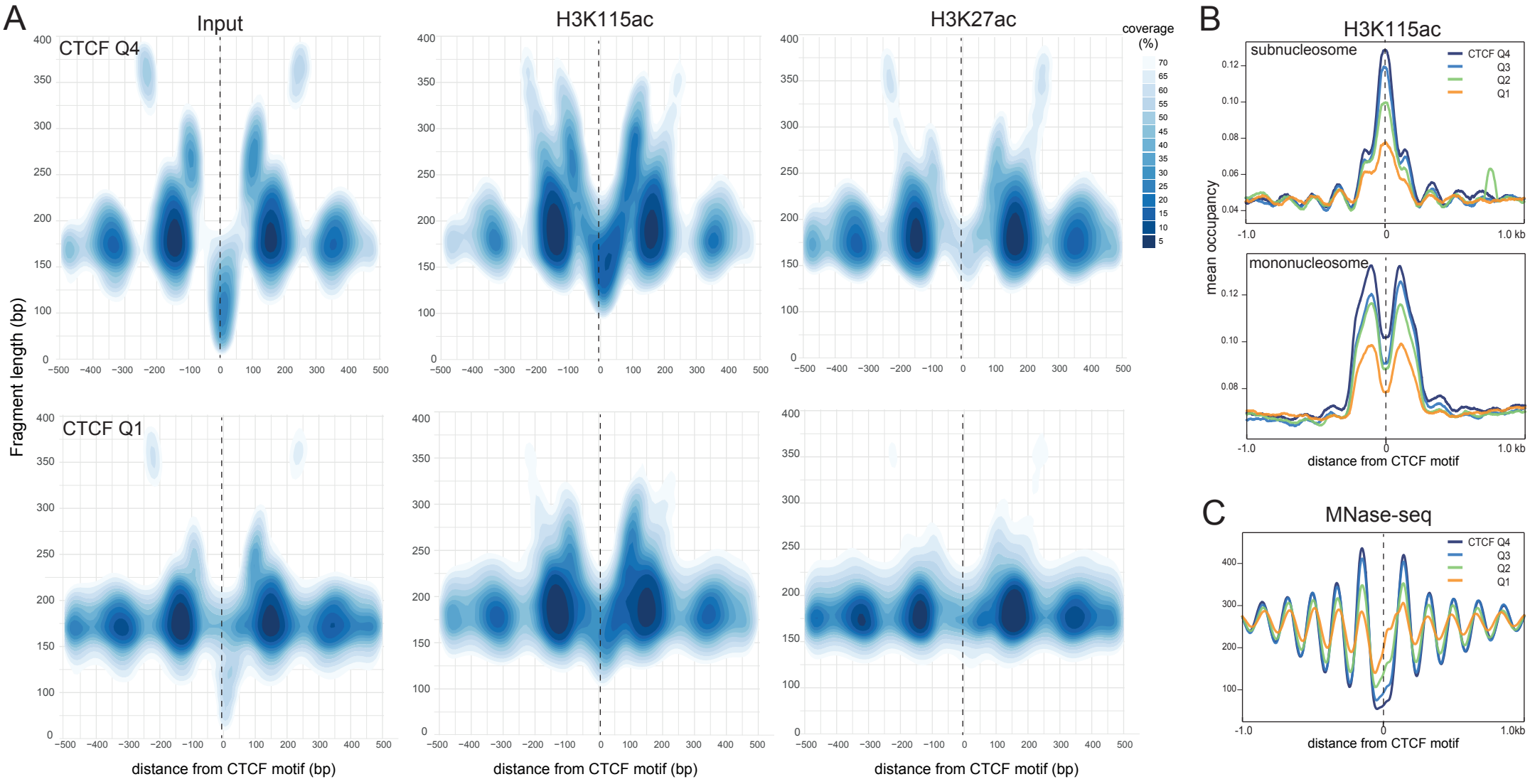


Figure 6 H3K115ac marks fragile nucleosomes at CTCF sites

A) Contour plots depicting high density regions of (left) input, (centre) H3K115ac ChIP, (right) H3K27ac ChIP paired-end sequenced MNase fragments around top (Q4) and bottom (Q1) quartiles of CTCF ChIP-seq peaks in mESCs (data from Mas et al., 2018) as a function of fragment length. All CTCF motifs are oriented in the same direction. B) H3K115ac ChIP-seq signal from subnucleosomal (top) and mononucleosomal (bottom) MNase fragments in mESCs around CTCF motifs across the four quartiles of CTCF ChIP-seq peak strength. All CTCF motifs are oriented in the same direction. C) as in (B) but for MNase resistant, mono-nucleosomal signal from classic MNase-seq (data from West et. al., 2014).

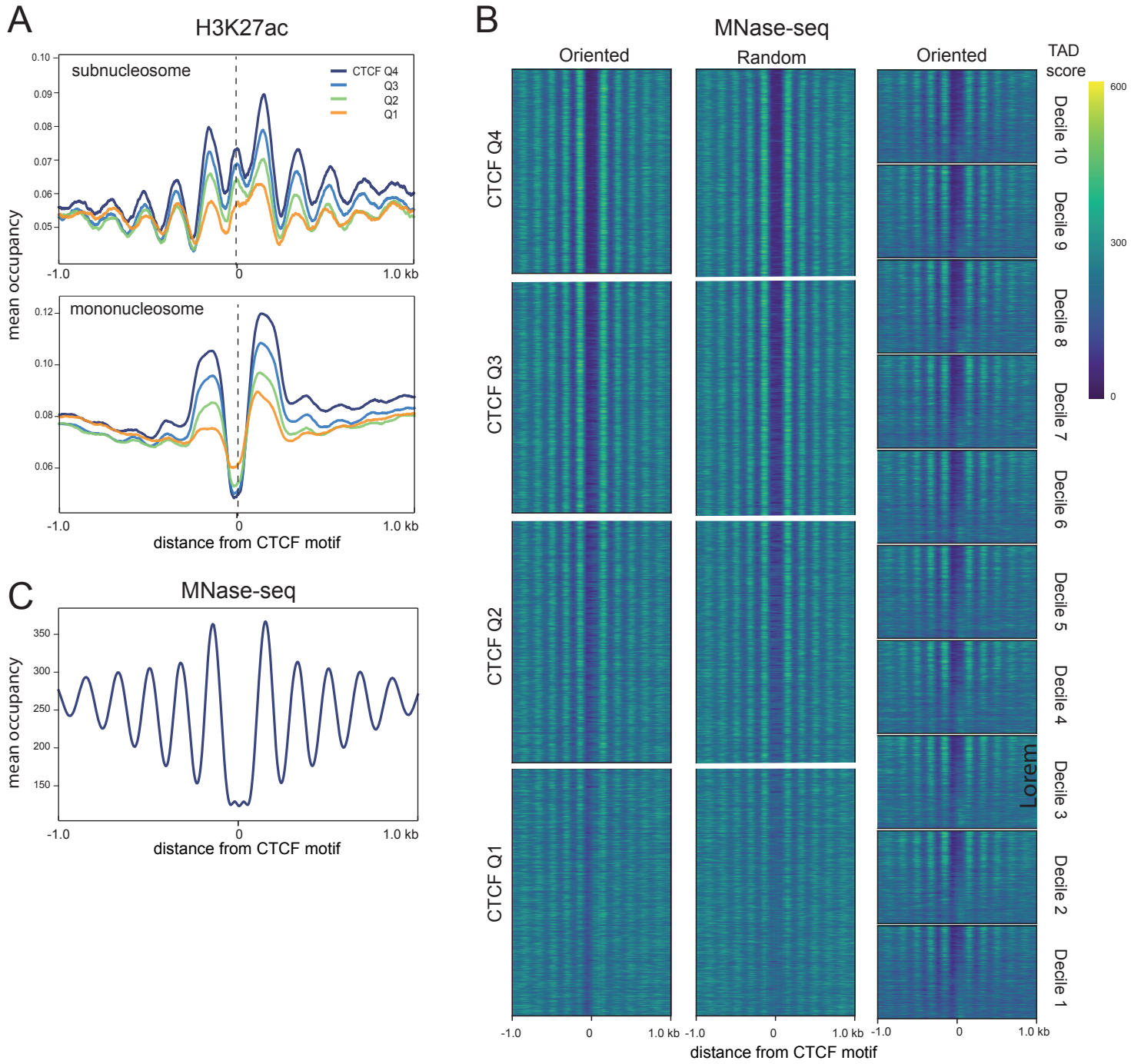


Figure S6 ***H3K27ac and H3K115ac relative to CTCF binding sites and motif orientation. Related to Figure 6.***

A) H3K27ac ChIP-seq signal from subnucleosomal (top) and mononucleosomal (bottom) MNase fragments in mESCs around CTCF motifs across the four quartiles of CTCF ChIP-seq peak strength. All CTCF motifs are oriented in the same direction. B) Heatmap of MNase-seq data ranked by quartiles of CTCF occupancy with CTCF motifs oriented in the same direction (left) or with randomised orientation (middle). The heatmap on the extreme right displays MNase-seq signal over CTCF sites divided into deciles of distance to the closest TAD boundary in mESCs (CTCF sites in decile1 are closest, TAD boundaries from Bonev et al., 2017). Within each decile, sites are sorted by CTCF occupancy in descending order. C) Mean occupancy of nucleosomes derived from classical MNase-seq around all CTCF sites centred at the motif.

nucleosomes correlates with lower CTCF binding scores at all deciles of distance to TAD boundary. These data suggest an H3K115ac modified fragile nucleosomes are associated with the mode of CTCF binding.

Discussion

Histone modifications are commonly used to identify putative active regulatory elements in the mammalian genome. Acetylation of the histone H3 N-terminal tail at lysine 27 (H3K27ac) is the most widely used acetyl-lysine histone modification used for this purpose (Moore et al., 2020; Stunnenberg et al., 2020). However, H3K27ac is dispensible both for gene activation in mESCs during the exit from pluripotency (Sankar et al., 2022), and for the accessibility of enhancers (Zhang et al., 2020), calling into question the causal role of H3K27ac in enhancer function.

Beyond the histone H3 tail, lysine acetylation in the globular domain of histones, and particularly at the lateral surface of the nucleosome where histones are in close contact with overlying DNA, are important for nucleosome dynamics, nucleosome stability (diCerbo et al., 2014; Chatterjee et al., 2015; Lawrence et al., 2016; Rajagopalan et al., 2017) and transcription (Tropberger et al., 2013). We have previously shown that acetylation of H3K64 and H3K122 marks a subset of active enhancers in mESCs (Pradeepa et al., 2016). Whilst in *Drososphila melanogaster* H3K122R mutation – rendering the residue unacetylatable - is viable, H3K115R and H3K115Q mutations both show 100% embryonic lethality in the fruit fly (Graves et al., 2016). In *Saccharomyces cerevisiae* H3K115A and H2K115Q abrogate transcriptional silencing and result in a hypersensitivity to DNA damage (Hyland et al., 2005). In this study we explore the acetylation of H3 at lysine 115 (H3K115ac) in mammalian cells, showing that it also marks a specific subset of active promoters and putative enhancers in mESCs and during differentiation. Compared with other histone modifications, H3K115ac better discriminates whether sequences showing regulatory activity in mESCs by STARR-seq (Peng et al., 2020) will also show evidence of chromatin accessibility by ATAC-seq in their endogenous context.

It is now recognised that, rather than being nucleosome free, active promoters are characterised by MNase sensitive nucleosomes that are termed fragile nucleosomes (Jin & Felsenfeld 2007; Carone et al., 2014; Ishi et al., 2015; Voong et al., 2016). Some fragile nucleosomes exist as subnucleosomes missing certain histones. Though fragile nucleosomes have been associated with H3.3 and H2AZ histone variants (Jin et al., 2013), to date it has not been clear if nucleosome fragility is also linked to histone modification, perhaps influencing the unwrapping or breating of DNA from the core particle, or the release of histone subunits (Zhou et al., 2019). Here we have shown that H3K115ac specifically marks the fragile nucleosomes

at the TSS of the most active genes. The DNA fragments protected at these H3K115ac marked TSS nucleosomes are approximately 15bp shorter than the flanking mononucleosomes, suggesting that H3K115 acetylated TSS nucleosomes may be partially unwrapped. These fragile nucleosomes are devoid of H3K27ac but are flanked by H3K27ac nucleosomes at the +1 and -1 positions.

For the most highly expressed genes we noted that the enrichment of H3K115ac is restricted to those TSS associated with CpG islands (CGI), whereas H3K27ac is still enriched, albeit at lower levels, on the nucleosomes flanking non-CGI TSS. This is consistent with the known correlation between CGI and nucleosome depletion (Tazi and Bird, 1990; Fenouil et al., 2012). CGI promoters have other distinguishing features, including the presence of distinct binding proteins and transcription regulation mechanisms (Hughes et al., 2023). Even though CGIs are generally GC-rich, at the very local level the DNA sequences associated with the H3K115ac nucleosomes have a higher A/T content than those associated with H3K27ac, consistent with the enhanced MNase sensitivity of nucleosomes associated with more A/T-rich sequences (Chereji et al., 2019).

We also detected H3K115ac-marked fragile nucleosomes at sites co-occupied by pluripotency factors at enhancers in mESCs. In budding yeast, fragile nucleosomes are generated by the action of chromatin remodellers of the SWI/SNF family (Brahma and Henikoff, 2019). In mESCs, BRG1 has recently been suggested to act on fragile nucleosomes at enhancers to generate MNase-sensitive sub-nucleosomal hemisome particles that facilitate OCT4 binding (preprint; Nocente et al., 2022). Given the impact of H3K115 acetylation on nucleosome stability in vitro (Manohar et al., 2009; Simon et al., 2011), it is plausible that H3K115 acetylation facilitates the action of BRG1 in further destabilisation of fragile nucleosomes at enhancers.

Chemical cleavage methods have detected sub-nucleosomal particles at CTCF binding sites (Voong et al., 2016). In this study, we detect H3K115ac modified fragile nucleosomes at sites strongly occupied by CTCF, with a directional bias toward the downstream (3') end of the CTCF motif. We also detect asymmetry in the nucleosome phasing surrounding CTCF motifs, with strong phasing being preferentially lost on the 3' side of CTCF motifs as CTCF occupancy decreases. The zinc fingers of CTCF bind their recognition site such that it is the N-terminus of CTCF that is oriented toward the 3' end of the motif (Yang et al., 2023). It is the N-terminal domain of CTCF that is also important for the interaction with cohesin, for the impeding or reversal of cohesin-mediated loop extrusion and for the insulation of genome architecture at TAD boundaries (Li et al., 2020; Nora et al., 2020; Davidson et al., 2023; Zhang et al., 2023). However, we did not detect a correlation between the strength of nucleosome phasing and TAD boundaries (insulation score) independent of CTCF occupancy. Together these data suggest

that remodelling of nucleosomes and nucleosome positioning are linked to the orientation and strength with which CTCF binds its motif.

To the best of our knowledge, H3K115ac is the first histone modification enriched on a fragile nucleosome at TSS, at TF binding sites of active enhancers, and at CTCF sites. In the future, it will be interesting to establish the functional significance of H3K115ac in modulating nucleosome structure, genome regulation and genome organisation and to ascertain the value of H3K115ac profiling for the functional annotation of genomes.

Materials and Methods

Cell culture and neural differentiation

Cell lines were validated and tested for mycoplasma at IGC, University of Edinburgh. Undifferentiated 46C *Sox1-GFP* mESCs (Ying et al., 2003) were cultured in GMEM-BHK21 medium (Gibco, #21710-025) supplemented with Leukaemia Inhibitory factor (10^3 units/ml. in-house), 10% foetal calf serum (Hyclone, #SH3007003), non-essential amino acids (Sigma, #M7145), sodium pyruvate (Sigma, #S8636), β-Mercapto-ethanol (GIBCO, #31350010). 46C mESCs were differentiated to neural progenitor cells (NPCs) as described previously (Ying et al., 2003; Benabdallah et al., 2016). Cells were seeded at $3X10^6$ 24 hours (h) prior to differentiation on 0.1% gelatin-coated T75 Corning flasks. Differentiation media ((1:1DMEM/F12: Neuro basal medium (GIBCO, #31330-032 and 21103-049 respectively) supplemented with 0.5x B27 (Invitrogen, #17504044), 0.5x N2 (Invitrogen, #17502048), L-glutamine and 50mM 2-β-mercaptoethanol (GIBCO, #31350010)) was prepared fresh on the day of differentiation. On D0, ESCs were harvested, washed twice with PBS, twice with differentiation media, then seeded at a density of $1x10^6$ ESCs on 0.1% gelatin-coated T75 Corning flasks. Cells were cultured for 7 days with daily media changes after day 2. NPC differentiation was monitored visually through expression of the Sox1-GFP reporter.

Peptide Dot Blots

Custom peptides (ThermoFisher Scientific) with the sequence N-CAIHA$\mathbf{K}_{115}$RVTIMPK-C were synthesized as unmodified (H3K115), acetylated K115 (H3K115ac) and lysine to arginine mutation (H3K115R). For H3K122ac the peptide sequence was N-CGGVTIMP$\mathbf{K}_{122}$DIQLA-C. 2.5, 5,10 and 20 ng of peptide were spotted on a nitrocellulose membrane and allowed to dry at room temperature (r.t.). The blot was incubated in blocking buffer (5% non-fat dry milk in TBST: 20mM Tris, 150mM NaCl, + 0.15% Tween-20) for 30 minutes (mins), followed by primary antibody (1:1500 dilution of anti- H3K115ac, catalogue # PTM170, PTM Bio) in blocking buffer at r.t. before washing 3 x with TBST for 5 min each. The blot was then incubated with a 1:10000 dilution of secondary antibody (HRP-linked anti-Rabbit IgG; Cell Signalling Technology, catalogue # 7074S) in blocking buffer for 45 mins at r.t. followed by washing as in previous steps. The blot was developed with HRP substrate (# 34096, ThermoFisher) and imaged on an ImageQuant LAS-4000 Imager.

Native ChIP-seq with partial MNase digestion

mESCs and NPCs were harvested with Accutase (Lifetech) and washed twice with cold PBS. 5 million cells were permeabilized by resuspending in 100μL cold NBA buffer (5.5% Sucrose, 85 mM NaCl, 10 mM TrisHCl pH 7.5, 0.2 mM EDTA, 0.2 mM PMSF, 1 mM DTT, 1X Protease

Inhibitors) and 100μL of NBB buffer (NBA buffer + 0.1% NP40) and incubated for 10 mins on ice. Nuclei were then collected by centrifugation at 1000g for 5 min at 4°C followed by washing with 200 μl NBR buffer (5.5% Sucrose, 85 mM NaCl, 10mM TrisHCl pH7.5, 3 mM $MgCl_2$, 1.5 mM $CaCl_2$, 0.2 mM PMSF, 1 mM DTT). Nuclei were resuspended in 100μl NBR buffer and incubated with 1μl of RNase-A (10mg/ml) for 5 min at r.t. followed by treatment with 4U MNase (Sigma) for 10 min at 20°C while mixing at 500 rpm. MNase was stopped by addition of an equal volume of ice cold 2x STOP buffer (215 mM NaCl, 5.5 % sucrose, 10 mM TrisHCl pH 8, 20 mM EDTA, 2 % Triton X100, 0.2 mM PMSF, 1 mM DTT, 2X Protease Inhibitors) and placing on ice. The reaction was diluted to 500μl total volume by adding NBR:STOP buffer and incubated on ice for 14 h to release digested fragments. Samples were centrifuged at 10000g for 10 mins at 4°C and solubilized chromatin was recovered as the supernatant. At this point, 10% of the chromatin was saved as Input. Dynabeads-ProteinA (Invitrogen) were blocked and pH-calibrated as per the manufacturer's protocol and 5 μl bed volume of beads were loaded with 2 μg of ChIP antibody (H3K27ac; Abcam-AB4729, H3K115ac; PTM BIO-PTM-170). Beads were added to the solubilized chromatin, supplemented with BSA to 0.1mg/ml and rotated for 4 h at 4°C. Beads were collected by a pulse spin and concentrated on a magnetic rack, washed 3x for 5 min each at r.t with 500 μL ChIP WASH buffer (10mM Tris pH 8, 2mM EDTA, 150mM NaCl, 1% NP-40, 1 % Na-Deoxycholate), followed by a single wash with TE buffer. Bead bound chromatin was eluted with 100μl elution Buffer (0.1 M $NaHCO_3$, 1 % SDS) at 37°C while mixing at 1000 rpm for 15 mins. Input chromatin was diluted to 100 μl with elution buffer and the pH of Input and IP samples was adjusted by adding 7 μl of 2M TrisCl pH 6.8 and both were treated with 20 μg Proteinase-K for 2 h at 55°C while mixing at 1000 rpm. DNA was purified with Qiagen MinElute columns and sequencing libraries constructed using the NEBNext® Ultra™ II DNA library kit per the manufacturer's instructions. Samples were sequenced for 50 cycles in paired-end mode on an Illumina HiSeq-2000.

ATAC-seq sample preperation

ATAC-seq libraries were prepared from $5x10^4$ cells as previously described (Buenrostro et al., 2013) with some modifications (Wong et al 2023). We used 15 mins incubation on ice in the nuclei preparation step and the Tn5 reaction was performed in 50 μl of custom transposition buffer (10 mM Tris pH 8, 5 mM $MgCl_2$ and 10% dimethylformamide) with 2.5 μl Tn5 transposase (Illumina, 20034197) at 37°C while mixing at 1000 rpm for 30 mins. The reaction was stopped and purified with a MinElute PCR Purification Kit (Qiagen, 28004). To minimize PCR duplication, we used 20% of tagmented DNA for a test qPCR reaction to determine the minimum number of PCR cycles required to amplify the DNA without entering the plateau phase. The rest of the tagmented DNA was then amplified in identical conditions of dilution and

PCR cycles. PCR reactions were purified and unused adaptors removed with an equal volume of AMPure XP beads (Beckman, A63880) following the manufacturer's protocol and eluted in 20 µl Tris pH 7.8. Libraries were quantified using a Qubit assay (Invitrogen, Q32851) and were selected for the presence of a clear nucleosomal ladder pattern on a Bioanalyzer High Sensitivity assay (Agilent). Samples were sequenced for 50 cycles in paird-end mode on an Illumina HiSeq-2000.

4SU sequencing (4SU-seq)

4SU-seq was performed as described previously (Boyle et al., 2020). 4-thiouridine (4SU; Sigma T4509) was added to cell cultures to a final concentrarion of 500 µM and incubated for 20 mins at 37°C. Cells were collected by trypsinization and total RNA purified from $5x10^6$ cells with Trizol (Invitrogen, 15596026) and DNase-treated with a Turbo DNA-free kit (Invitrogen, AM1907M). Thirty micrograms of total RNA were biotinylated using 60 µg of Biotin-HPDP (Pierce, 21341, 1mg/ml stock in dimethylformamide) in a 300 µl reaction in biotinylation buffer (10 mM TrisHCl pH 7.4, 1 mM EDTA) for 90 mins at r.t. Uncoupled biotin was removed by chloroform extraction (1:1 volume) and labelled RNA precipitated using isopropanol. RNA was resuspended in 100 µl RNAse-free water and incubated with an equal volume of µMacs streptavidin beads (Miltenyi 130-074-101) for 15 mins at r.t. Beads were captured on a pre-calibrated µMacs column, followed by 3 washes with 900 µl of Wash Buffer pre-heated to 65°C followed by 3 washes at r.t. RNA was eluted with 100 mM DTT and purified using an RNeasy MinElute cleanup kit (Qiagen 74204) as per the the manufacturer's instructions. Ribosomal RNA was removed using a RiboMinus eukaryote system v2 kit (Ambion, A15027) and stranded libraries constructed using the NEBNext Ultra II directional RNA library preparation kit (NEB, E7760) following the protocol for ribosome-depleted RNA and with an 11-min RNA fragmentation step according to the manufacturer's instructions. Libraries were amplified, purified and sequenced in paired-end mode as described for ATAC-seq.

Data Analysis

Quality assessment and pre-processing of sequencing data:

Publicly available datasets were downloaded from the NCBI SRA repository with SRA toolkit (fastq-dump). We used FASTQC v0.11.7 for quality control metrics. Filtering of poor-quality reads and trimming of adaptor sequences was performed using *fastp* (Chen et. al., 2018) using default settings.

ChIP-seq analysis

Single-end ChIP libraries: Filtered and trimmed reads were aligned to the mm10 assembly of mouse genome using *bowtie* (Langmead and Salzberg, 2012) and uniquely mapped reads were kept. Alignments were converted to BAM format and sorted using *SAMtools* (Li et al., 2009) and PCR duplicates removed using *Picard* (https://broadinstitute.github.io/picard/). *Bedtools* (Quinlan and Hall, 2010) was used to create BED files followed by conversion into HOMER (Heinz et al., 2010) tag directory format (*makeTagDirectory*) using default parameters. We called peaks using HOMERs *findPeaks* command with options *-F 2 -localSize 50000 -size 150 -minDist 300 -fdr 0.01 –region.* UCSC compatible bigwig files normalized to sequencing depth were generated using *makeUCSCfile* with parameters: *-bigWig -style chipseq –fragLength 150*.

Paired-end ChIP libraries: Filtered and trimmed reads were aligned to the mm10 genome using *bowtie2* with options *(--no-discordant --no-mixed --no-unal -X 2000*) followed by conversion to BAM format and removal of PCR duplicates as above. Read pairs were converted into ChIP fragments using *bedtools bamtobed –bedpe* and fractionated into sub-nucleosomes (<150 bp) and mono-nucleosomes (150-225 bp). Fragment coverage was computed at base-pair resolution with *bedtools genomeCoverageBed* and coverage normalized to sequencing depth and scaled to 10^7. Coverage data was converted to bigwig format using *wigTobigWig* (Kent utilities, UCSC). We computed Pearsons correlation between two replicates in 10Kb windows across the genome using the normalized coverage data and replicates were merged based on high correlation (Pearsons coeff. > 0.9).

ATAC-seq analysis

Trimming, filtering, and alignment steps are as described for paired-end ChIP-seq. The 5′ ends of the reads were offset by +4 bases for reads on the Watson strand, and by −5 bases for the reads on the Crick strand, to reflect the exact location of Tn5 insertion. Peaks were called using *macs2 callpeak–nomodel–extsize 150–shift -75 -g 'mm' -p 0.01.* Irreproducible discovery rate (IDR, ≤0.05, Li et al., 2011) was used to select reproducible peaks from two biological replicates. For visualization, a 20 bp fragment was centered on each Tn5 insertion site and its coverage was calculated at single-base resolution in the whole genome. Scaling and normalization steps were as for paired-end ChIP-seq above.

4SU-seq analysis

Trimming, filtering and alignment steps are as described for paired-end ChIP-seq. Alignments were converted into HOMER tag directory format with options suitable for stranded libraries *-format sam -flip –sspe*. Strand-specific genome wide coverage (bigwig format) was computed

with parameters: *makeUCSCfile –bigwig -fsize 1e20 -strand + (or −) -norm 1e8.* Read coverage for promoters was computed using HOMER's *annotatePeaks.pl* with parameters: *-size "given" -noann -nogene -len 0 -strand both -norm 0.* Read counts were then converted into reads per kilobase per million mapped reads (RPKM).

MNases-seq data processing and nucleosome calls

Classical MNase-seq (fragments within mononucleosomal range) data for mESCs was downloaded from NCBI SRA (GSE59062, West et. al., 2014). Trimming, filtering, and alignment steps are as described for paired-end ChIP-seq. High quality alignments (MAPQ >=30) from two biological replicates were merged using *SAMtools*. Nucleosome peaks were called using the DANPOS3 algorithm (Chen et. al., 2015) with default parameters. Single base resolution wig files from DANPOS3 were converted to bigwig files for visualization in the UCSC genome browser and plotting coverage using *deeptools* (Ramírez et. al., 2016).

Promoter and TSS definitions

Gene annoatations were downloaded from the UCSC table browser for genome build: mm10, track: NCBI RefSeq, table: UCSC RefSeq (refGene). For a given TSS (coding and non-coding), 1000 bp upstream and downstream were added and used as a promoter region to filter out promoter proximal peaks in downstream analyses. To create a reference set of unique protein-coding TSSs without duplicated entries, we collapsed the transcripts with identical TSS location (mouse TSS set_v1, supplementary data). For use in V-plots, a more conservative set of TSSs was defined such that: i) for genes with multiple TSSs, a single TSS with highest 4SU-seq count in mESCs was retained, ii) genes with single TSSs were included, iii) TSSs were filtered out if there was another coding/non-coding TSS within 1 Kbp. This ensured that the local and directional observations in our analyses are not affected by transcriptional activity in the vicinity.

Random forest model

ATAC-seq accessible non-promoters were overlapped with H3K115ac, H3K122ac, and H3K27ac peaks and classified as H3K115ac-positive or negative. Peaks from the cistromeDB (Mei et al., 2017) and ReMap2022 (Hammal et al., 2022) were filtered as described previously (Friman et al., 2023). The PeakPredict package (https://github.com/efriman/PeakPredict) command *overlap_peaks* was run with settings *'--predict_column H3K115ac --model RandomForestClassifier --balance --shap'* , where balancing downsamples the groups to have the same size prior to splitting into test and training sets. The PeakPredict package implements scikit-learn (Pedregosa et al., 2011) and SHAP values (Lundberg and Lee, 2017).

CTCF ChIP-seq peaks and motif calls

Replicate merged CTCF ChIP-seq peaks for mESCs were downloaded from Remap2022 (Mas et al.,2018; Hammal et. al., 2022; GSE99530). ChIP-seq peak scores were used to split data into quartiles in R. The mouse genome was scanned for concensus CTCF motifs (JASPER: MA0139.1) using *matchPWM* in the Biostrings package in Bioconductor. We removed any peaks that overlapped with > one CTCF motif and CTCF binding scores divided into quartiles.

Contour plots

For each reference (TSS, enhancer, CTCF motif etc) we plotted signed distance (x-axis) and fragment length (y-axis). We defined equal number of high density regions as contours using the package ggdensity in ggplot2 (Otto and Kahle, 2023). Contours refer to different levels of fragment density as shown by color density. Each contour is associated with a probability (%) with which it is bound to data.

Datasets

The following publically accessible datasets were used in this study; NCBI GEO GSE66023 (H3K122ac, H3K27ac), GSM1003756 (H3K4me3), GSM1003750 (H3K4me1), GSM1276707 (H3K27me3), GSE59062 (MNase), and GSE115774 (4SU-seq). Sequencing data generated in this study are submitted to NCBI GEO with the Accession number GSExxx.

Author Contributions

DS, YK: Designed and performed the experiments; YK, EF, RI: Performed data analyses; DS, WB, YK: conceived the project; YK, WB: wrote the manuscript; all authors edited the manuscript.

Acknowledgements.

We thank Nick Gilbert for discussions. This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/) for sequencing data analysis and the Wellcome Trust Clinical Research Facility, Edinburgh (https://clinical-research-facility.ed.ac.uk/) for sequencing.

Funding statement.

DS was supported by a Newton Fellowship from the Royal Society. Work in the WAB lab is funded by MRC University Unit grants MC_UU_00007/2 and and MC_UU_00035/7. E.T.F. was supported by the Swiss National Science Foundation (P500PB_206805).

References

Benabdallah NS, et al. (2016) SBE6: a novel long-range enhancer involved in driving sonic hedgehog expression in neural progenitor cells. *Open Biol*. 6: 160197.

Benabdallah NS, Williamson I, Illingworth RS, Kane L, Boyle S, Sengupta D, Grimes GR, Therizols P, Bickmore WA. (2019) Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Mol. Cell* 76:473-484.e7.

Bergsland M, Ramskold D, Zaouter C, Klum S, Sandberg R, Muhr J. (2011) Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev.* 25:2453–2464.

Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A, Cavalli G. (2017) Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*. 171:557-572.e24.

Boyle S, Flyamer IM, Williamson I, Sengupta D, Bickmore WA, Illingworth RS. 2020. A central role for canonical PRC1 in shaping the 3D nuclear landscape. *Genes Dev*. 34:931-949.

Brahma S, Henikoff S. 2019. RSC-Associated Subnucleosomes Define MNase-Sensitive Promoters in Yeast. *Mol. Cell* 73:238-249.e3.

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. Methods*. 10:1213-1218.

Carone BR, Hung JH, Hainer SJ, Chou MT, Carone DM, Weng Z, Fazzio TG, Rando OJ. (2014) High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev. Cell* 30:11-22.

Chatterjee N, North JA, Dechassa ML, Manohar M, Prasad R, Luger K, Ottesen JJ, Poirier MG, Bartholomew B. (2015) Histone Acetylation near the Nucleosome Dyad Axis Enhances Nucleosome Disassembly by RSC and SWI/SNF. *Mol. Cell. Biol*. 35:4083-4092.

Chen K, Chen Z, Wu D, Zhang L, Lin X, Su J, Rodriguez B, Xi Y, Xia Z, Chen X, et al. (2015) Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat. Genet*. 47:1149-1157.

Chen S, Zhou Y, Chen Y, Gu J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 34:i884-i890.

Chereji RV, Bryson TD, Henikoff S. (2019) Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biol*. 20:198.

Cosgrove MS, Boeke JD, Wolberger C. (2004) Regulated nucleosome mobility and the histone code. *Nat. Struct. Mol. Biol*. 11:1037-1043.

Davidson IF, Barth R, Zaczek M, van der Torre J, Tang W, Nagasaka K, Janissen R, Kerssemakers J, Wutz G, Dekker C, Peters JM. (2023) CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* 616:822-827.

Deaton AM, Gómez-Rodríguez M, Mieczkowski J, Tolstorukov MY, Kundu S, Sadreyev RI, Jansen LE, Kingston RE. (2016) Enhancer regions show high histone H3.3 turnover that changes during differentiation. *Elife* 5:e15316.

Di Cerbo V. Mohn F, Ryan DP, Montellier E, Kacem S, Tropberger P, Kallis E, Holzner M, Hoerner L, Feldmann A, *et al.* (2014) Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription. *Elife* 3:e01632.

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

Fenouil R, Cauchy P, Koch F, Descostes N, Cabeza JZ, Innocenti C, Ferrier P, Spicuglia S, Gut M, Gut I, Andrau JC. (2012) CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res*. 22:2399-408.

Friman ET, Flyamer IM, Marenduzzo D, Boyle S, Bickmore WA. (2023) Ultra-long-range interactions between active regulatory elements. *Genome Res*. 33:1269-1283.

Fu Y, Sinha M, Peterson CL, Weng Z. (2008) The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet*. 4:e1000138.

Ghoneim M, Fuchs HA, Musselman CA. (2021) Histone Tail Conformations: A Fuzzy Affair with DNA. *Trends Biochem. Sci*. 46:564-578.

Graves HK, Wang P, Lagarde M, Chen Z, Tyler JK. (2016) Mutations that prevent or mimic persistent post-translational modifications of the histone H3 globular domain cause lethality and growth defects in Drosophila. *Epigenetics Chromatin* 9:9.

Hammal F, de Langen P, Bergon A, Lopez F, Ballester B. (2022) ReMap 2022: a database of Human, Mouse, Drosophila and Arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucl. Acids Res*. 50:D316-D325.

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.

Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. (2011) Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci U S A*. 108:18318-18323.

Hughes AL, Szczurek AT, Kelley JR, Lastuvkova A, Turberfield AH, Dimitrova E, Blackledge NP, Klose RJ. (2023) A CpG island-encoded mechanism protects genes from premature transcription termination. *Nat. Commun*. 14:726.

Hyland EM, Cosgrove MS, Molina H, Wang D, Pandey A, Cottee RJ, Boeke JD. (2005) Insights into the role of histone H3 and histone H4 core modifiable residues in Saccharomyces cerevisiae. *Mol. Cell. Biol*. 25:10060-10070.

Ishii H, Kadonaga JT, Ren B. (2015) MPE-seq, a new method for the genome-wide analysis of chromatin structure. *Proc Natl Acad Sci U S A.* 112:E3457-65.

Jin C, Felsenfeld G. (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev*. 21:1519-1529.

Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G. (2009) H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat. Genet*. 41:941-945.

Kent NA, Adams S, Moorhouse A, Paszkiewicz K. (2011) Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing. *Nucl. Acids Res*. 39:e26.

King HW, Klose RJ. (2017) The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *Elife*. 13;6:e22631.

Klein DC, Troy K, Tripplehorn SA, Hainer SJ. (2023) The esBAF and ISWI nucleosome remodeling complexes influence occupancy of overlapping dinucleosomes and fragile nucleosomes in murine embryonic stem cells. *BMC Genomics*. 24:201.

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

Lawrence M, Daujat S, Schneider R. (2016) Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet*. 32:42-56.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078-2079.

Li Q, Brown JB, Huang H, Bickel PJ. (2011) Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* 5:1752–1779.

Li Y, Haarhuis JHI, Sedeño Cacciatore Á, Oldenkamp R, van Ruiten MS, Willems L, Teunissen H, Muir KW, de Wit E, Rowland BD, Panne D. (2020). The structural basis for cohesin-CTCF-anchored loops. *Nature* 578:472-476.

Lundberg S, Lee S-I. (2017) A Unified Approach to Interpreting Model Predictions *arXiv*:1705.07874

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol*. 28:495-501.

MacCarthy CM, Huertas J, Ortmeier C, Vom Bruch H, Tan DS, Reinke D, Sander A, Bergbrede T, Jauch R, Schöler HR, Cojocaru V (2022) OCT4 interprets and enhances nucleosome flexibility. *Nucl. Acids Res*. 50:10311-10327.

Manohar M, Mooney AM, North JA, Nakkula RJ, Picking JW, Edon A, Fishel R, Poirier MG, Ottesen JJ. (2009) Acetylation of histone H3 at the nucleosome dyad alters DNA-histone binding. *J. Biol. Chem.* 284:23312-23321.

Mas G, Blanco E, Ballaré C, Sansó M, Spill YG, Hu D, Aoi Y, Le Dily F, Shilatifard A, Marti-Renom MA, Di Croce L. (2018) Promoter bivalency favors an open chromatin architecture in embryonic stem cells. *Nat. Genet*. 50:1452-1462.

Mei S, Qin Q, Wu Q, Sun H, Zheng R, Zang C, Zhu M, Wu J, Shi X, Taing L, et al. (2017) Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucl. Acids Res*. 45(D1):D658-D662.

Moore JE, ENCODE Project Consortium; Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, Kawli T, Davis CA, Dobin A, Kaul R, et al., (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583:699-710.

Narita T, Higashijima Y, Kilic S, Liebner T, Walter J, Choudhary C. (2023) Acetylation of histone H2B marks active enhancers and predicts CBP/p300 target genes. *Nat. Genet*. 55:679-692.

Nocente MC, Karamitsos AM, Drouineau E, Albawardi W, Dulary C, Ribierre F, Picaud H, Alibert O, Acker J, Aude J-C, et al. (2022) BRG1 generates subnucleosomes that expand OCT4 binding and function beyond DNA motifs at enhancers. *bioRxiv* 2022.09.15.507958.

Nora EP, Caccianini L, Fudenberg G, So K, Kameswaran V, Nagle A, Uebersohn A, Hajj B, Saux AL, Coulon A, et al., (2020) Molecular basis of CTCF binding polarity in genome folding. *Nat. Commun*. 11:5612.

Otto J, Kahle D (2023) ggdensity: Interpretable Bivariate Density Visualization with 'ggplot2'. https://jamesotto852.github.io/ggdensity/, https://github.com/jamesotto852/ggdensity/.

Owens N, Papadopoulou T, Festuccia N, Tachtsidi A, Gonzalez I, Dubois A, Vandormael-Pournin S, Nora EP, Bruneau BG, Cohen-Tannoudji M, Navarro P. 2019. CTCF confers local nucleosome resiliency after DNA replication and during mitosis. *Elife* 8:e47898.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, 2011. Scikit-learn: Machine learning in Python. *J. Machine Learning Res. 12*: 2825-2830.

Peng T, Zhai Y, Atlasi Y, Ter Huurne M, Marks H, Stunnenberg HG, Megchelenbrink W. (2020) STARR-seq identifies active, chromatin-masked, and dormant enhancers in pluripotent mouse embryonic stem cells. *Genome Biol*. 21:243.

Peng Y, Li S, Landsman D, Panchenko AR. (2021) Histone tails as signaling antennas of chromatin. *Curr. Opin. Struct. Biol*. 67:153-160.

Pintacuda G, Wei G, Roustan C, Kirmizitas BA, Solcan N, Cerase A, Castello A, Mohammed S, Moindrot B, Nesterova TB, Brockdorff N. (2017) hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish Polycomb-Mediated Chromosomal Silencing. *Mol. Cell* 68:955-969.e10.

Pradeepa MM, Grimes GR, Kumar Y, Olley G, Taylor GCA, Schneider R, Bickmore WA. (2016) Histone H3 globular domain acetylation identifies novel enhancers. *Nat. Genet.* 48:681-686.

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

Rajagopalan M, Balasubramanian S, Ioshikhes I, Ramaswamy A. (2017) Structural dynamics of nucleosome mediated by acetylations at H3K56 and H3K115, 122. *Eur. Biophys J*. 46:471-484.

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. *Nucl. Acids Res*. 44:W160-165.

Rodriguez Y, Horton JK, Wilson SH. (2019) Histone H3 Lysine 56 Acetylation Enhances AP Endonuclease 1-Mediated Repair of AP Sites in Nucleosome Core Particles. *Biochemistry* 58:3646-3655.

Sankar A, Mohammad F, Sundaramurthy AK, Wang H, Lerdrup M, Tatar T, Helin K. (2022) Histone editing elucidates the functional roles of H3K27 methylation and acetylation in mammals. *Nat. Genet*. 54:754-760.

Simon M, North JA, Shimko JC, Forties RA, Ferdinand MB, Manohar M, Zhang M, Fishel R, Ottesen JJ, Poirier MG. (2011) Histone fold modifications control nucleosome unwrapping and disassembly. *Proc. Natl. Acad. Sci U S A.* 108:12711-12716.

Sinha KK, Bilokapic S, Du Y, Malik D, Halic M. (2023) Histone modifications regulate pioneer transcription factor cooperativity. *Nature* 619:378-384.

Stunnenberg HG; International Human Epigenome Consortium; Hirst M. (2016) The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* 167:1145-1149.

Taguchi H, Xie Y, Horikoshi N, Maehara K, Harada A, Nogami J, Sato K, Arimura Y, Osakabe A, Kujirai T, et al., H. (2017) Crystal Structure and Characterization of Novel Human Histone H3 Variants, H3.6, H3.7, and H3.8. *Biochemistry* 56:2184-2196.

Tazi J, Bird A (1990) Alternative chromatin structure at CpG islands. *Cell* 60:909-920.

Tropberger P. Pott S, Keller C, Kamieniarz-Gdula K, Caron M, Richter F, Li G, Mittler G, Liu ET, Bühler M, et al., (2013) Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. *Cell* 152:859–872.

Tropberger P, Schneider R (2013) Scratching the (lateral) surface of chromatin regulation by histone modifications. *Nat. Struct. Mol. Biol.* 20:657–661.

Vernimmen D, Bickmore WA. (2015) The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. *Trends Genet*. 31:696-708.

Voong LN, Xi L, Sebeson AC, Xiong B, Wang JP, Wang X. (2016) Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell* 167:1555-1570.e15.

West JA, Cook A, Alver BH, Stadtfeld M, Deaton AM, Hochedlinger K, Park PJ, Tolstorukov MY, Kingston RE. (2014) Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nat. Commun*. 5:4719.

Wong YY, Harbison JE, Hope CM, Gundsambuu B, Brown KA, Wong SW, Brown CY, Couper JJ, Breen J, Liu N, et al. (2023). Parallel recovery of chromatin accessibility and gene expression dynamics from frozen human regulatory T cells. *Sci Rep*. 13:5506.

Yang J, Horton JR, Liu B, Corces VG, Blumenthal RM, Zhang X, Cheng X. (2023). Structures of CTCF-DNA complexes including all 11 zinc fingers. *Nucl. Acids Res*. 13:gkad594.

Ying QL, Stavridis M, Griffiths D, Li M, Smith A. (2003) Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol*. 21:183-186.

Zhang, T., Zhang, Z., Dong, Q. Xiong J, Zhu B. (2020) Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biol*. 21:45.

Zhang H, Shi Z, Banigan EJ, Kim Y, Yu H, Bai XC, Finkelstein IJ. (2023) CTCF and R-loops are boundaries of cohesin-mediated DNA looping. *Mol. Cell* 28:S1097-2765.

Zhou K, Gaullier G, Luger K. (2019) Nucleosome structure and dynamics are coming of age. *Nat. Struct. Mol. Biol*. 26:3-13.

Zorro Shahidian L, Haas M, Le Gras S, Nitsch S, Mourão A, Geerlof A, Margueron R, Michaelis J, Daujat S, Schneider R. (2021) Succinylation of H3K122 destabilizes nucleosomes and enhances transcription. *EMBO Rep*. 22:e51009.