Published in final edited form as:
Nat Genet. 2016 June ; 48(6): 681–686. doi:10.1038/ng.3550.

Histone H3 globular domain acetylation identifies a new class of enhancers

Madapura M Pradeepa[1,2,*], **Graeme R Grimes**[1], **Yatendra Kumar**[1], **Gabrielle Olley**[1], **Gillian C A Taylor**[1], **Robert Schneider**[3,4], and **Wendy A Bickmore**[1,*]

[1]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine at University of Edinburgh, Crewe Road, Edinburgh, UK

[2]School of biological sciences, University of Essex, Colchester, UK

[3]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104/Inserm U964/Université de Strasbourg, 67400 Illkirch, France

[4]Helmholtz Zentrum Munich, Institut of Functional Epigenetics, 8574 Neuherberg, Germany

Abstract

Histone acetylation is generally associated with active chromatin, but most studies have focused on the acetylation of histone tails. Various histone H3 and H4 tail acetylations mark the promoters of active genes1. This includes acetylation of H3 on lysine 27 (H3K27ac), which blocks the deposition of polycomb mediated H3K27me32. H3K27ac is also widely used to identify active enhancers3,4, and the assumption has been that profiling of H3K27ac is a comprehensive way of cataloguing the set of active enhancers in mammalian cell types. Here we show that acetylation of lysine residues in the globular domain of H3 (H3K64ac and H3K122ac) marks active gene promoters and also a subset of active enhancers. Moreover, we find a novel class of active functional enhancers that are marked by H3K122ac but lack H3K27ac. This work suggests that, to identify enhancers, a more comprehensive analysis of histone acetylation is required than was previously considered.

Keywords

Histone acetylation; H3K122ac H3K64ac; Enhancers; Polycomb

*Correspondence to: M.M.P (pmadap@essex.ac.uk) or W.A.B. (Wendy.Bickmore@igmm.ed.ac.uk).
MRC Human Genetics Unit, IGMM, Crewe Road, Edinburgh EH4 2XU, UK, Tel: +44 131 332 2471, Fax: +44 131 467 8456

Accession codes
ChIP sequencing data generated in this study have been submitted to the NCBI Gene expression Omnibus (GEO) repository under accession number GSE66023. Other datasets used in this study and their accession numbers are given in Supplementary Table 6.

Competing Interests
The authors declare no competing interests.

Author Contributions
M.M.P., Y.K., G.O. and G.C.A.T. performed the experiments. M.M.P and G.R.G analysed data. R.S provided valuable reagents and discussion. M.M.P and W.A.B conceived the project, designed experiments and wrote the manuscript. All authors contributed to writing, read the paper and provided feedback.

Europe PMC Funders Author Manuscripts

Covalent modifications at the globular domains of the core histones have been implicated in a variety of chromatin functions5. Post-translational modifications (PTMs) located on the lateral (outer) surface of the histone octamer can alter contacts between the histones and the nucleosomal DNA and directly affect chromatin structure5. The acetylation of H3K56 (H3K56ac) is associated with DNA unwrapping from the nucleosome and has been implicated in chromatin assembly and genome stability6. Acetylation of H3 at K64 (H3K64ac), located at the start of the first alpha helix in the histone fold domain (HFD), destabilizes nucleosomes and facilitates nucleosome dynamics *in vitro*7. Methylation of the H3K64 is implicated in heterochromatin establishment8. Histone – DNA interactions reach their maximum strength in the nucleosome dyad and, unlike acetylation on histone tails, H3K122ac is sufficient to stimulate transcription *in vitro* from chromatinized templates9 and promote nucleosome disassembly10.

Metagene analysis of H3K122ac and H3K64ac chromatin immunoprecipitation (ChIP) sequencing reads from mouse embryonic stem cells (mESCs) shows these marks correlate with the magnitude of gene expression (Fig. 1a). Surprisingly, given the link between histone acetylation and active chromatin, we find H3K122ac over a subset of inactive or poised genes that are repressed by polycomb complexes in mESCs (Fig. 1b and c). Sequential ChIP-qPCR confirmed the presence of H3K122ac on bivalently (H3K27me3/H3K4me3) marked nucleosomes (Fig. 1d).

Pearson correlation analysis across multiple histone modifications in mESCs indicates that H3K64ac and H3K122ac cluster with each other and with H3K4me1 (Fig. 1e) – a marker for enhancers11. H3K122ac and H3K64ac reads were also enriched at active promoters and strong enhancers across hidden Markov model based chromatin states (ChromHmm)12,13 (Supplementary Figure 1). Given this, we aligned H3K64ac, H3K122ac and H3K27ac ChIP-seq data with the mid-point of enhancers in mESCs, as defined by the H3K4me1 peaks ± 2 kb away from RefSeq TSSs11 (Fig. 2a). The data clustered into three groups based on the overlap of H3K4me1 peaks with those of H3K27ac and H3K122ac. Group 1, (n = 23,153) are H3K27ac+ and are, for the most part, also marked by significantly high levels of H3K122ac and H3K64ac (Wilcoxon sum rank test, Supplementary Table 1). This group of enhancers would be classified as active based upon their H3K27ac status3,4. At the other extreme, group 3 (n = 5,265) are negative for all three acetylation marks, and would be classified as inactive enhancers. Group 2 enhancers (n = 9,340) are negative for H3K27ac, but are marked by significantly high levels of H3K122ac and, a subset by H3K64ac. Using current methods, these would be classified as inactive enhancers. H3K122ac (which co-occupies promoters with H2A.Zac9 and can induce transcription14) and H2A.Zac are comparably enriched in group 2 as in group 1 enhancers (Fig. 2b). Group 2 enhancers also have high levels of the EP300, which acetylates H3 at K64, K122 and K277,9,15.

We found that group 1 enhancers have high levels of H3K122ac and H3K64ac (Fig. 2b). A subset of the clustered enhancers associated with highly expressed genes in ESCs, which have been termed ‘super-enhancers’ (SEs)16, were also heavily enriched with H3K64ac (Fig. 2a and d and Supplementary Figure 2). Our data suggest that there is a class of putative regulatory elements (Group 2, in Fig. 2) in mESCs that are marked by H3K122ac and/or H3K64ac but that lack H3K27ac that is usually used as a predictor of active enhancers.

Gene-Ontology (GO) analysis of subclasses indicates that both the H3K27ac+ and the H3K27ac–/H3K122ac+ group of enhancers are associated with terms such as ‘stem-cell maintenance’. But the H3K27ac+ enhancers were also significantly enriched with terms associated with cell adhesion, which were lacking in the H3K122ac+/H3K27ac– set. Instead hindbrain morphogenesis, placental development and germ layer formation terms were prominent (Supplementary Figure 3a). A sub-class of group 2 enhancers, which are H3K27me3+, are enriched with terms associated with negative regulation of transcription, differentiation and development (Supplementary Figure 3b).

Transcription factor (TF) motif enrichment analysis indicated SP1, SP2, SP4, KLF5, EGR1, TFAP2a, TFAP2b and TFAP2c binding sites, which we note generally have a high GC content, are enriched in group 2 enhancers (Supplementary Figure 4a). Compared to group 1 and group 3, group 2 enhancers also have higher levels of H3K27me3 and H2A.Z (Fig. 2b) – both markers for poised promoters and enhancers17,18. A subset of group 2 enhancers with H3K27me3 peaks are enriched for un-methylated CpG islands (CGIs) (Supplementary Figure 4a), which are located at promoters and enhancers19. Bidirectional transcription of enhancers correlates with enhancer activity20, however these transcripts are degraded by the exosome complex making them difficult to detect. Analysis of Exosome sensitive RNAs (RNA seq reads from Exosome component 3 (*Exosc3*) knockout ESCs vs wild type)21 shows that group 2 enhancers transcribe high levels of Exosome sensitive eRNAs (Fig. 2c).

We tested the enhancer activity of these elements using luciferase reporter assays in mESCs, a well-characterized *Nanog* enhancer22 (Fig. 2d) served as a positive control. Group 2 genomic regions (H3K27ac–) with enrichment for H3K122ac (Fig. 3a and b) exhibited 4 – 120 fold higher activity compared to negative controls and were equally, or more, active than the *Nanog* enhancer. Similarly, enhancer assays performed in a human breast adenocarcinoma cell line (MCF7) cells showed that H3K27ac–/H3K122ac+ enhancers9 display higher reporter activity than H3K27ac+ enhancers (Fig. 3c and d).

To demonstrate the *in vivo* functional importance of group 2 enhancers, we used CRISPR/Cas923 to delete them from the ESC genome (Fig. 4a and b). As positive controls we also deleted one allele of the SE located near *Nanog* and *Klf4* (Fig. 2c). This led to a significant reduction in *Nanog* and *Klf4* expression, respectively (Fig. 4c), but not of *Dppa3* – located 80kb upstream of *Nanog*. Expression of *Rad23b* – 180 kb downstream of *Klf4* is somewhat affected by the intervening enhancer deletion. Homozygous deletion of the putative group 2 enhancer 42kb downstream of *Lif* (Lif 42k en$^{-/-}$) led to reduced expression of *Lif,* but not of the flanking gene *Hormad2* (Fig. 4c). Similarly, deletion of one allele of the putative enhancer 30kb upstream from *Tbx3* (Tbx3 -30k en) led to down regulation of *Tbx3*.

To examine whether histone acetylation is important for the function of these new regulatory elements we used dCas9 to recruit the Sid4x repressor complex24 to them (Fig. 4d). As positive controls, recruitment of dCas9-Sid4x to the *Nanog* enhancer, and to the SE of *Nanog Klf4,* and *Sox2,* led to significant reduction in expression of the respective target genes but not other nearby genes (Fig. 4e). For the group 2 enhancers analysed, ChIP-qPCR showed that Sid4x recruitment effectively reduced the levels of H3K122ac at the target Tbx3 -30k en, with no effect at the off-target control (Sox2 SE) (Fig. 4d). RT-qPCR analysis

showed reduced expression of putative target genes upon Sid4x recruitment to Foxd3 -57k en, Tbx3 -30k en, Sox2 40k en and Sox2 60k en, but not of the control genes (Fig. 4e). Sox2 40k en also displayed higher activity in reporter assays (Fig. 3b).

In order to investigate H3K122ac as an enhancer mark in more detail, we performed ChIP-seq for H3K122ac, H3K27ac and H3K4me1 in a human erythroleukemic (K562) cell line. As in ESCs, H3K122ac is enriched at active promoter, strong enhancer and poised promoter states (ChromHmm)12 in K562 cells (Fig. 5a and b). H3K122ac is also enriched at SEs, and H3K27ac+ enhancers (Fig. 5c, d & e). Similar to ESCs, a subset of H3K27ac– enhancers is marked with H3K122ac (Fig. 5c, d & e), is DHSs and bound by TFs (Fig. 5e, Supplementary dataset 2). TFs enrichment analysis of ENCODE ChIP-seq shows group 2 enhancers are enriched for CTCF, ZNF143, SMC3, RAD21, EZH2 and USF1 over group 1 (Supplementary Figure 4c).

Rather than a simple definition of active enhancers as being marked by H3K4me1/H3K27ac, a more complex picture of different histone acetylation marks at enhancers is emerging25. Our data suggests that using H3K27ac alone gives an incomplete catalogue of the active enhancer repertoire, and that acetylation of H3 at the lateral surface of the histone octamer can be used to identify a novel class of active enhancers that have no significant H3K27ac enrichment.

Lysine acetyl transferases (KATs) generally have relaxed substrate specificity, with the exception of KAT8, which acetylates H4K1625,26 and is critical for the maintenance of ESC pluripotency and differentiation27,28. H4K16ac marks active enhancers in ESCs, including some that lack H3K27ac25. Like the globular domain acetylations of H3, H4K16ac directly affects chromatin structure by perturbing inter-nucleosomal interactions *in vitro*29, but not higher-order chromatin structure25. The role of most histone acetylation marks at enhancers is unknown, but acetylation in the histone tails can recruit reader proteins such as BRD4 that are thought to be important for enhancer function30. This is unlikely to be the case for H3K64 and H3K122 acetylation due to their location at the lateral surface of the histone octamer. Rather, acetylation of these residues is believed to function by directly altering nucleosomal stability and mobility, and by facilitating the binding of activators5. The finding of H4K16ac and H3 globular domain acetylations at enhancers suggests that opening of local chromatin structure might be an important facet of enhancer function and may stimulate the identification of yet more regulatory histone PTMs that directly affect the physical properties of the nucleosome.

URLs

GREAT: Genomic Regions Enrichment of Annotations Tool, http://bejerano.stanford.edu/great/public/html ; R Project for Statistical Computing, https://www.r-project.org ; BEDTools suite, http://bedtools.readthedocs.org/en/latest. Super-enhancer database http://www.bio-bigdata.com/SEA; Gene expression omnibus (GEO) server, www.ncbi.nlm.nih.gov/geo; ENCODE/Broad –K562 ChromHmm, http://genome.ucsc.edu/ENCODE/downloads.html ; mESC ChromHmm https://github.com/gireeshkbogu/

chromatin_states_chromHMM_mm9. http://rsat.sb-roscoff.fr/; NGS plots, https://github.com/shenlab-sinai/ngsplot ;

Online Methods

Cell culture

46C, Sox1-GFP mouse embryonic stem cells (mESC)31 were cultured as described previously25. Human erythro-myeloblastoid leukemia cells (K562) were cultured in RPMI 1640 with L-Glutamine media containing 10% fetal bovine serum (FBS), L-glutamine, penicillin and streptomycin. Cell lines were validated and Mycoplasma tested at IGMM, University of Edinburgh.

Sequential Chromatin Immunoprecipitation (ChIP)—Antibodies recognising H3K122ac and H3K64ac were previously described7,9. mESCs were cross-linked in 1% formaldehyde for 10 min and then quenched by the addition of glycine to a final concentration of 0.125 M. Chromatin was sheared using a biorupter (Diagenode) to an average fragment length of ~100 – 200bp. Sequential ChIP was performed as described previously32. Briefly, 5 μg antibodies against H3K4me3 (07-473, Millipore) and H3K27me3 (07-449, Millipore) were covalently coupled to Dynabeads using Invitrogen antibody coupling kit (Cat. 14311D) according to the manufacturer's instructions. The first ChIP was performed using either H3K4me3 or H4K27me3 antibodies, and the immunoprecipitated chromatin was then eluted with 10 mM DTT, diluted 30 times with RIPA buffer (1X PBS, 1% NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, *Roche Protease Inhibitor Cocktail) before performing the second ChIP with anti-H3K122ac9. Purified chromatin was quantified by qPCR using the standard curve method and expressed as % of input bound. Primer details are given in Supplementary Table 2.

Native Chromatin Immunoprecipitation

10×10^6 mESCs and K562 cells were centrifuged at 500 *g* for 3 min, washed twice in PBS and then resuspended in 200 μl of NBA buffer (85 mM NaCl, 5.5 % Sucrose, 10 mM TrisHCl pH 7.5, 0.2 mM EDTA, 0.2 mM PMSF, 1 mM DTT, 1x Protease inhibitors (Calbiochem, 539134-1SET). Cells were lysed by the addition of an equal volume of NBA + 0.1 % NP40 and incubated on ice for 3 min. Nuclei were pelleted at 2,000 *g* for 3 min at 4 ºC, then washed with NBR buffer (85 mM NaCl, 5.5 % Sucrose, 10 mM TrisHCl pH 7.5, 3 mM $MgCl_2$, 1.5 mM $CaCl_2$, 0.2 mM PMSF and 1 mM DTT) and pelleted at 2,000 *g* for 3 min at 4 ºC. Nuclei were resuspended (10×10^6 nuclei/ml) in NBR supplemented with RNaseA (20 μ g/ml) and incubated at 20 ºC for 5 min. Chromatin was fragmented for 30 min at 20 ºC using 0.133 U/μl micrococcal nuclease (MNase - Boehringer units; SigmaAldrich - N3755-500UN; titrated to give predominantly mono-nucleosomes). Digestion was stopped with the addition of an equal volume of STOP bufffer (215 mM NaCl, 10 mM TrisHCl pH 8, 20 mM EDTA, 5.5 %, Sucrose, 2 % TritonX 100, 0.2 mM PMSF, 1 mM DTT, 2X Protease Inhibitors) and digested nuclei left on ice overnight to release soluble, fragmented chromatin. Chromatin was pre-cleared by centrifugation at 12,000 *g* for 10 min at 4 ºC and the soluble chromatin (supernatant) transferred to a fresh tube. 5 % of the released chromatin was retained as input and the remainder incubated for 4

h at 4 ºC on a rotating wheel with ~5 μg of antibodies (H3K122ac9; H3K64ac7; H3K4me1 - Abcam ab8895, lot:GR251663-1; H3K27ac - Abcam ab4729, lot:GR254707-1) pre-coupled to protein A dynabeads (Life Technologies; 10002D) in PBS containing 5 mg/ml BSA and 0.1 mM PMSF. Immune complexes bound to beads were washed 5x with wash buffer 1 (150 mM NaCl, 10 mM TrisHCl pH 8, 2 mM EDTA, 1 % NP40 and 1 % sodium deoxycholate) on a rotating wheel for 5 min each and once in room temperature TE buffer for 1 min. Chromatin was released from the beads by incubation with 0.1 M $NaHCO_3$ / 1 % SDS for 30 min at 37 ºC followed by the addition of proteinase K (100 ug/ml) and Tris pH 6.8 (100 mM) and incubation at 55 ºC overnight. For both native and cross-linked ChIP, Dynabeads were removed using a magnetic rack and the chromatin purified using Qiaquick PCR Purification columns (Qiagen) according to the manufacturer's instructions.

ChIP-seq library preparation and Deep Sequencing

Libraries were prepared as previously described33 with the following modifications: No purification was performed between the A-tailing and ligation reactions. After A-tailing reaction, enzymes were inactivated by incubation at 75 ºC for 20 min. and the ligation reaction was supplemented with ligation reagents (400 U of T4 DNA ligase (NEB), 1x buffer 2 (NEB), 7.5 % PEG-6,000, 1 mM ATP and 13.3 nM of annealed Illumina adaptors (AU)) and incubated at 16 °C overnight. Size selection following the ligation and PCR steps was performed with 1x and 0.8x reaction volumes of Agencourt AMPure XP beads respectively (Beckman Coulter - A63880).

Replicate 1 of the H3K122ac and H3K64ac ChIPs was sequenced at The Danish National High-Throughput DNA sequencing Center (Copenhagen; 42 base single end reads). Replicate 2 of the H3K122ac and H3K64ac ChIPs, 2 replicates of H3K27ac ChIPs and all ChIP and input samples prepared from K562 cells were sequenced at Edinburgh Genomics (The University of Edinburgh, 50 base single end reads).

Read mapping

FASTQ files were aligned using Bowtie34 (version 0.12.8) with parameters set to retain uniquely mapped reads with a maximum of two mismatches (bowtie options: -e 40 -m 1 -v 2). For mapping, mm9 and hg18 bowtie indexes were used for mouse (mESC) and human (K562 and MCF7) datasets respectively. Mapped reads from two biological replicates of H3K27ac, H3K122ac and H3K64ac were merged for further analysis.

Peak calling

Peaks were called using SICER35. For mESC, MNase-digested ChIP input DNA (GSM1156619) was used as a background control for H3K27ac, H3K64ac and H3K122ac. For H3K4me1 in ESCs (E14TG2a; GSM1003750), Input (GSM1003746) was used as a background control. mESC biological replicates were merged using SAMtools (v0.1.19) prior to peak calling with SICER (v1.1). SICER parameters: window size – 200 bp; fragment size – 150 bp; false discovery rate – 0.01; gap size – 600 bp for H3K122ac, H3K64ac, and H3K4me1 and a 200bp window size for H3K27ac.

Generation of Bedgraphs for visualisation on UCSC genome browser

Bedgraphs for each histone mark were generated from the aligned read files using the HOMER software suite (v4.7)36, at a resolution of 10 bp and with a normalised tag count of 10 million. Mapped reads from two biological replicates for H3K122ac, H3K64ac and H3K27ac ChIPs in mESCs were combined for the generation of Bedgraphs for Figure 1 to 4. UCSC tracks for individual replicates covering representative loci are shown in Supplementary Figure 2. Similarly, data from single experiments for MCF7 and K562 ChIPseq reads were processed to generate Bedgraphs for visualization in UCSC genome browser.

Heatmaps and average profiles

Heatmaps and average profile for Refseq gene transcription start sites (TSS; ± 2kb), Refseq gene transcription end sites (TES; ± 2kb from), enhancer midpoints (± 2 kb from) and for entire length of super-enhancers (all scaled to an equivalent length ± 2 kb), were generated using ngsplot v2.6137.

For Figure 1a, gene expression quartiles from high (Q4) to low (Q1) were obtained from our previous study25 and used to generate average profile plots for H3K122ac and H3K64ac across TSS and TES as detailed above.

The Heatmap for Figure 1b was generated for TSSs (± 2kb) of genes which have been shown to be repressed by polycomb complexes38.

The average profile plots (Figure 2a) for enrichment of strand specific RNA-seq reads in *Exosc*$^{-/-}$/WT (SRP042355)21 for the 3 enhancer groups were generated using ngsplot37 (v2.61).

Genome-wide correlation analysis of histone marks

Pearson's correlation coefficients were calculated between datasets using bamCorrelate tool39 (version 1.5.9, removing duplicate reads and a using a resolution of 10 kb). The correlation matrix was hierarchically clustered and visualized using the Bioconductor package pheatmap.

ChromHMM analysis

To calculate the distribution of histone marks against different chromatin states the bamCorrelate tool was used to count reads within chromHMM segments for K56212 and mESCs13. Datasets were normalized to read per million (RPM).

Enhancer analysis

Enhancers were defined as H3K4me1 peaks, with gene TSSs (RefSeq TSS ± 2kb) and genome blacklist40 regions removed. Active enhancer regions (group 1) were defined as genomic intervals overlapping both H3K4me1 and H3K27ac peaks. Inactive enhancers, defined as peaks of H3K4me1 with no associated H3K27ac peak, were stratified into group 2 and 3 representing those with and without an associated H3K122ac peak, respectively (Supplementary datasets 1 and 2). Peak intersections were performed using the BEDtools41

(v2.23.0) intersect function. Super-enhancer co-ordinates for K562 cells and mESC were obtained from super-enhancer archive. H3K27me3 peak regions were called using MACS242 (v2.1.0, broadpeak with no input control).

TF motif enrichment analysis

TF motif enrichment analysis was performed using the Regulatory Sequence Analysis Tools (RSAT) server. Nucleotide sequences from group 2 enhancers (H3K122ac+ in ESCs) were used as inputs for TF motif enrichment analysis with group 1 enhancer co-ordinates as the background.

Gene ontology (GO) enrichment analysis

Gene Ontology (Biological Process) enrichment analysis was performed using the Genomic Regions Enrichment of Annotations Tool (GREAT)43. Bed files from group1, group2 enhancers intersecting with H3K27me3 peaks (H3K122ac+/H3K27me3+) and group2 enhancers lacking H3K27me3 (H3K122ac+/H3K27me3–) were used as input and whole genome as background to select significantly enriched GO terms for nearby genes.

Enrichment analysis of DNaseI hypersensitivity sites (DHS) and un-methylated CpG islands (CGIs)19

To determine the enrichment of DHS (GSM1014154) and CGIs at subgroups of enhancers; a Fisher's exact test was performed using BEDtools fisher (default options) 41. Un-methylated CGI for mESCs were obtained from GSE43512.

Dual luciferase enhancer assays

Putative enhancer regions were PCR amplified from mouse (E14TG2a ESC) or human (HepG2) genomic DNA, cloned into pGL4.26 plasmid and sequence verified. Details of enhancers and PCR products used in this assay are given in the Supplementary Table 3. Putative enhancers from mESCs were assayed in E14TG2a mESCs and putative MCF7 cell enhancers from were assayed in MCF7 cells. Forty-eight hours post-transfection, a luciferase assay was performed using the Dual-luciferase Reporter assay (Promega) as per the manufacturer's instructions. Firefly luciferase activity was normalized to transfection efficiency with Renilla luciferase activity using pRL-TK. All values are shown as log2 ratios of enhancer activity vs. empty vector.

Enhancer deletions

Pairs of gRNAs (Supplementary Table 4) designed to direct Cas9 to regions flanking putative enhancers, were cloned into SpCas9-2A-GFP (PX458, Addgene number 48138) and transfected using lipofectamine 2000 (Invitrogen) into 46C ESCs44. 24 hours after transfection, transfected cells were FACS sorted for GFP + and are seeded at the 5,000 cells/ 100mm dish. Surviving colonies were isolated and screened for deletion by PCR and homozygous clones were verified by Sanger sequencing. RT-qPCR was performed as described previously32, altered gene expression upon deletion of enhancer elements was measured v/s wild type control.

Sid4x recruitment to enhancers

The repressive mSin3 Interaction Domain (Sid4x) was cloned C-terminal of dCas9 (pAC-Sid4x) by replacing VP160 from dCas9VP160-2A-puro (pAC94)45. 2-3 guides per enhancer, or 5 – 7 guides per super-enhancer (Supplementary Table 5), were designed and oligos were synthesized from Sigma or IDT and cloned into pSLQ sgRNA expression plasmid as described46. All clones were verified by Sanger sequencing. Equal ratios of guideRNA pools and dCas9-Sid4x plasmids were co-transfected into mESCs using Lipofectamine 2000. 24 hours after transfection puromycin (2μg/ml) was added to the media. Surviving transfected cells were harvested 48 hrs post transfection and RT-qPCR was performed as described32 and native ChIP was performed for H3K122ac and H3K27ac. ChIP enrichment was calculated as the percentage input bound by the standard curve method. As a control pAC-Sid4x was transfected along with non-targeting pSLQ sgRNA plasmid.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Robert Illingworth and Robert Young (MRC HGU, Edinburgh) and Philipp Tropberger for discussions, Uttiya Basu (Albert Einstein College of Medicine) for sharing mapped RNAseq data from Exosome knockout ESCs and Stanley Qi (Stanford University) for sharing CRISPR guideRNA plasmid backbone. This work was supported by the Medical Research Council UK and by a European Research Council advanced grant 249956 (WAB). Work in the RS laboratory is supported by the Fondation pour la Recherche Médicale (FRM), by the Agence Nationale de Recherche (ANR, CoreAc), La Ligue National Contre La Cancer (Equipe Labellise) and INSERM Plan Cancer (épigénetique et cancer).

References

1. Wang Z, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. 2008; 40:897–903. [PubMed: 18552846]
2. Kim TW, et al. Ctbp2 Modulates NuRD-Mediated Deacetylation of H3K27 and Facilitates PRC2-Mediated H3K27me3 in Active ESC Genes During Exit From Pluripotency. Stem Cells. 2015; 33:2442–55. [PubMed: 25944056]
3. Creyghton MP, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010; 107:21931–6. [PubMed: 21106759]
4. Rada-Iglesias A, et al. A unique chromatin signature uncovers early developmental enhancers in humans. Nature. 2011; 470:279–83. [PubMed: 21160473]
5. Tropberger P, Schneider R. Scratching the (lateral) surface of chromatin regulation by histone modifications. Nat Struct Mol Biol. 2013; 20:657–61. [PubMed: 23739170]
6. Neumann H, et al. A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. Mol Cell. 2009; 36:153–63. [PubMed: 19818718]
7. Di Cerbo V, et al. Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription. Elife. 2014; :1–23. DOI: 10.7554/eLife.01632
8. Daujat S, et al. H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming. Nat Struct Mol Biol. 2009; 16:777–81. [PubMed: 19561610]
9. Tropberger P, et al. Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. Cell. 2013; 152:859–72. [PubMed: 23415232]

10. Simon M, et al. Histone fold modifications control nucleosome unwrapping and disassembly. Proc Natl Acad Sci U S A. 2011; 108:12711–12716. [PubMed: 21768347]
11. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007; 39:311–8. [PubMed: 17277777]
12. Ernst J, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–9. [PubMed: 21441907]
13. Bogu GK, et al. Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. Mol Cell Biol. 2015; MCB.00955–15. doi: 10.1128/MCB.00955-15
14. Halley JE, Kaplan T, Wang AY, Kobor MS, Rine J. Roles for H2A.Z and its acetylation in GAL1 transcription and gene induction, but not GAL1-transcriptional memory. PLoS Biol. 2010; 8:e1000401. [PubMed: 20582323]
15. Tie F, et al. Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. Development. 2014; 141:1129–39. [PubMed: 24550119]
16. Whyte, Wa, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–19. [PubMed: 23582322]
17. Zentner GE, Tesar PJ, Scacheri PC. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. 2011; 21:1273–83. [PubMed: 21632746]
18. Voigt P, Tee WW, Reinberg D. A double take on bivalent promoters. Genes Dev. 2013; 27:1318–1338. [PubMed: 23788621]
19. Long HK, et al. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. Elife. 2013; 2:1–19.
20. Andersson R, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014; 507:455–61. [PubMed: 24670763]
21. Pefanis E, et al. RNA Exosome-Regulated Long Non-Coding RNA Transcription Controls Super-Enhancer Activity. Cell. 2015; 161:774–789. [PubMed: 25957685]
22. Jiang J, et al. A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol. 2008; 10:353–60. [PubMed: 18264089]
23. Mali P, et al. RNA-Guided Human Genome Engineering via Cas9. Science. 2013; 339:823–826. [PubMed: 23287722]
24. Konermann S, et al. Optical control of mammalian endogenous transcription and epigenetic states. Nature. 2013; 500:472–6. [PubMed: 23877069]
25. Taylor G, Eskeland R, Hekimoglu-Balkan B, Pradeepa M, Bickmore WA. H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. Genome Res. 2013; 23:2053–2065. [PubMed: 23990607]
26. Smith ER, et al. A Human Protein Complex Homologous to the Drosophila MSL Complex Is Responsible for the Majority of Histone H4 Acetylation at Lysine 16. Mol Cell Biol. 2005; 25:9175–9188. [PubMed: 16227571]
27. Li X, et al. The Histone Acetyltransferase MOF Is a Key Regulator of the Embryonic Stem Cell Core Transcriptional Network. Cell Stem Cell. 2012; 11:163–178. [PubMed: 22862943]
28. Ravens S, et al. Mof-associated complexes have overlapping and unique roles in regulating pluripotency in embryonic stem cells and during differentiation. Elife. 2014; 2014:1–23.
29. Shogren-Knaak M, et al. Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science. 2006; 311:844–7. [PubMed: 16469925]
30. Lovén J, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell. 2013; 153:320–34. [PubMed: 23582323]
31. Ying QL, Stavridis M, Griffiths D, Li M, Smith A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol. 2003; 21:183–186. [PubMed: 12524553]
32. Pradeepa MM, Grimes GR, Taylor GCa, Sutherland HG, Bickmore WA. Psip1/Ledgf p75 restrains Hox gene expression by recruiting both trithorax and polycomb group proteins. Nucleic Acids Res. 2014; 42:9021–32. [PubMed: 25056311]

33. Bowman SK, et al. Multiplexed Illumina sequencing libraries from picogram quantities of DNA. BMC Genomics. 2013; 14
34. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]
35. Zang C, et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics. 2009; 25:1952–8. [PubMed: 19505939]
36. Heinz S, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell. 2010; 38:576–589. [PubMed: 20513432]
37. Shen L, Shao N, Liu X, Nestler E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics. 2014; 15:284. [PubMed: 24735413]
38. Ku M, et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. 2008; 4:e1000242. [PubMed: 18974828]
39. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014; 42:W187–91. [PubMed: 24799436]
40. Dunham I, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
41. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
42. Zhang Y, et al. Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 2008; 9:R137. [PubMed: 18798982]
43. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010; 28:495–501. [PubMed: 20436461]
44. Ran F, Hsu P, Wright J, Agarwala V. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013; 8:2281–308. [PubMed: 24157548]
45. Cheng AW, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. 2013; 23:1163–1171. [PubMed: 23979020]
46. Chen B, et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell. 2013; 155:1479–91. [PubMed: 24360272]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

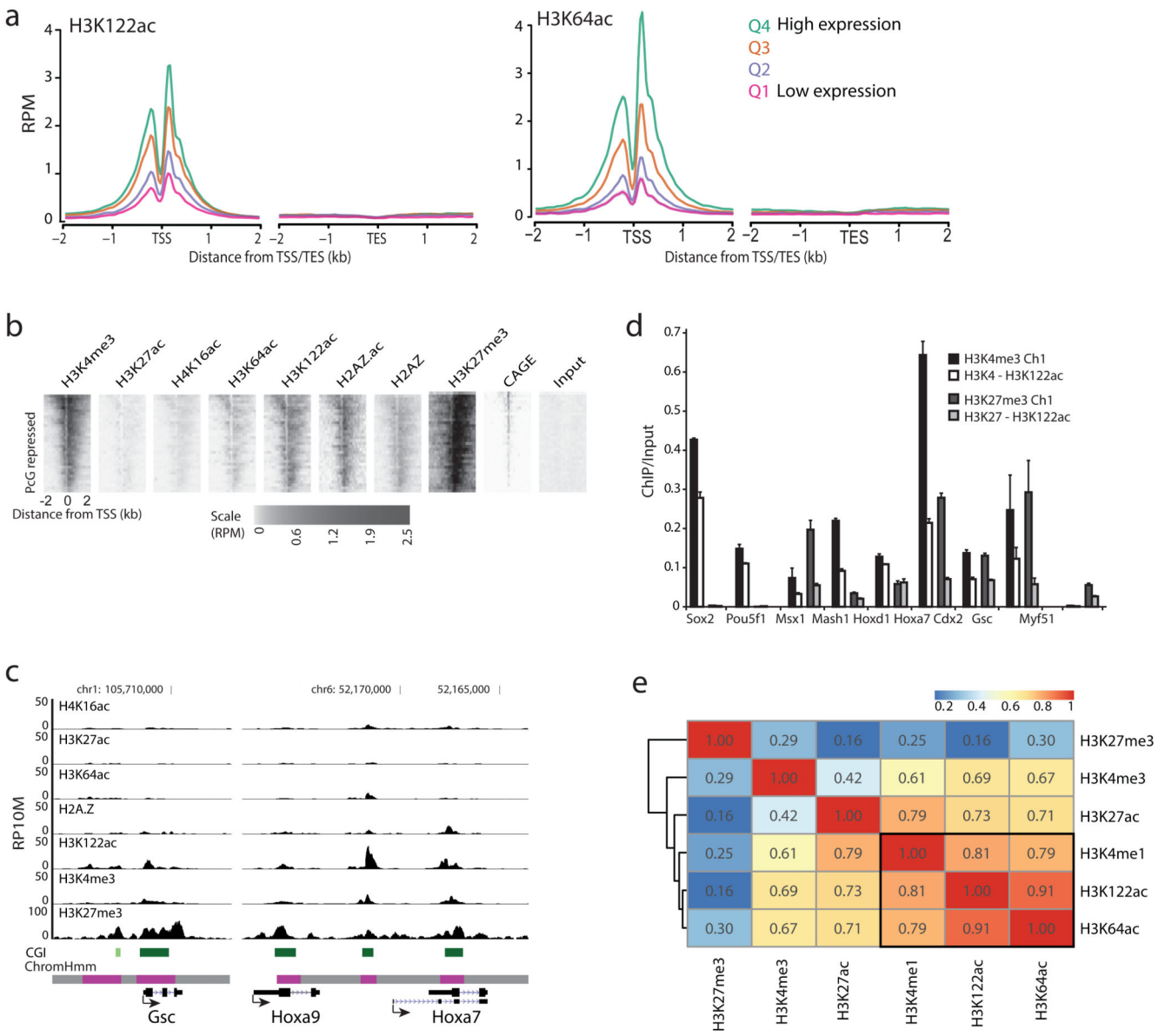

Figure 1. Genomic distribution of H3K122ac and H3K64ac

a) H3K122ac and H3K64ac native ChIP-seq reads per million (RPM) around (± 2kb) the transcription start (TSS) and end (TES) sites of genes, separated into quartiles according to gene expression in ESCs from low to high (Q1 – Q4) (n = 2 biological replicates).

b) Heatmaps of H3K4me3, H3K27ac, H4K16ac, H3K64ac, H3K122ac, H2A.Zac, H2A.Z, H3K27me3 ChIP-seq, CAGE tags and input chromatin RPM around (± 2kb) TSS of polycomb repressed genes in ESCs.

c) Reads per 10 million (RP10M) for ChIP-seq of H4K16ac, H3K27ac, H3K64ac, H2A.Z, H3K122ac, H3K4me3 and H3K27me3 across selected polycomb target genes *Gsc*, *Hoxa9* and *Hoxa7* in ESCs. Genome co-ordinates are from the NCBI37/mm9 assembly of the mouse genome, CpG islands (CGI) and ChromHmm segmentation of these coordinates are shown below (purple=poised promoters; grey=heterochromatin)13.

d) Sequential ChIP-qPCR over promoters of active genes – *Sox2*, *Pou5f1*, polycomb target genes – *Msx1*, *Mash1*, *Hoxd1*, *Hoxa7*, *Cdx2*, *Gsc*, and non-expressed gene *Myf51*. First ChIPs were performed with covalently coupled H3K4me3 (black bars) and H3K27me3 (dark grey bars) antibodies, followed by a second ChIP with H3K122ac antibodies (white and light grey bars for H3K4me3 and H3K27me3 first ChIP, respectively). Primer details given in Supplementary Table 2. Data is a representative of one of two experiments and error bars shows standard error of mean (s.e.m) from 3 technical replicates.

e) Heatmaps showing the hierarchical clustering of ChIP-seq data for H3K27me3, H3K4me3, H3K27ac, H3K4me1, H3K122ac and H3K64ac. Genome-wide Pearson's correlation coefficient was calculated by dividing the genome into 10kb windows, correlation values among histone modifications are shown.

Europe PMC Funders Author Manuscripts

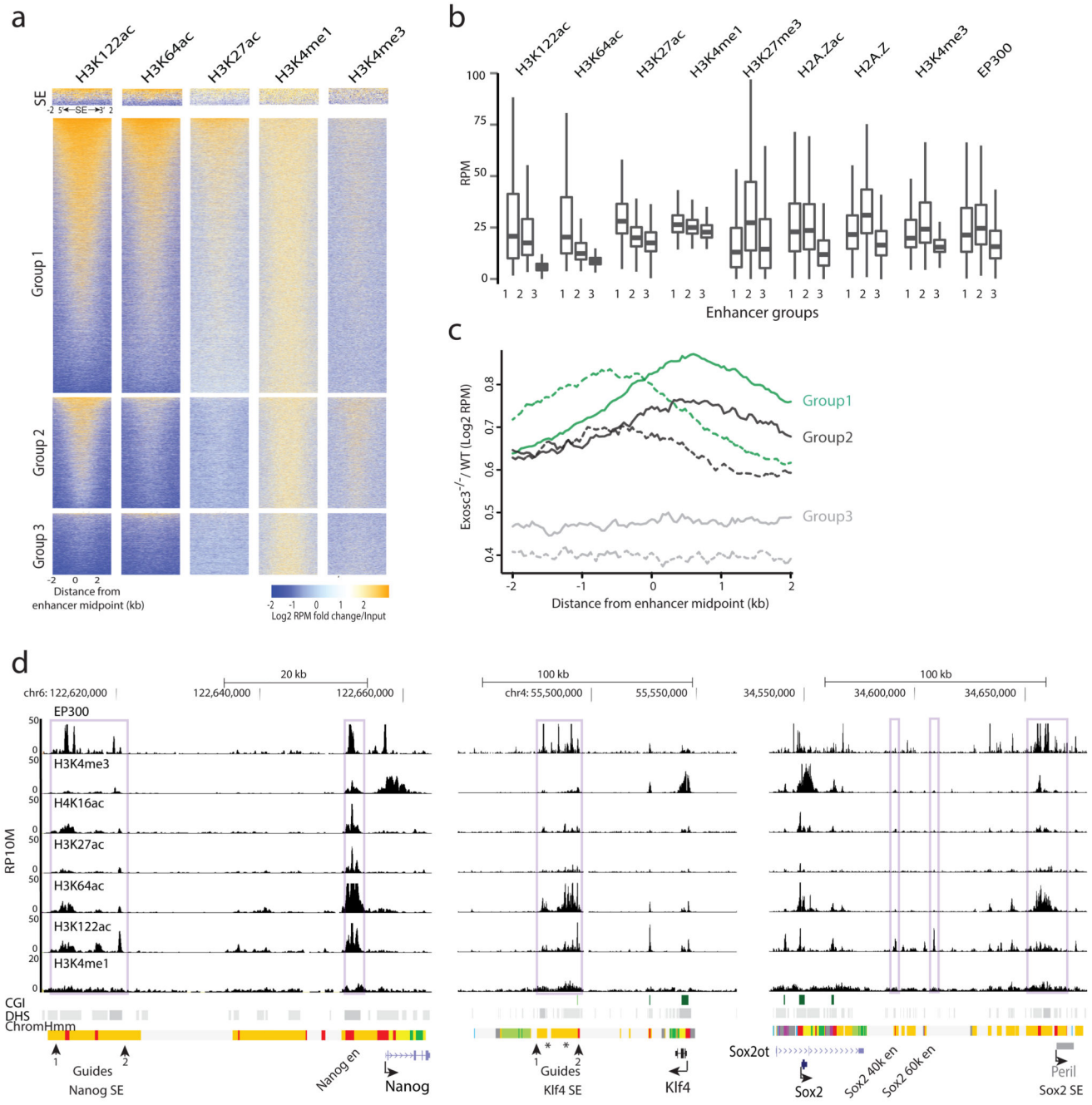

Figure 2. H3K122ac and H3K64ac marks active enhancers in ESC

a) Heatmaps of ChIP-seq (RPM) fold change/input around (± 2kb) enhancer midpoints for H3K122ac, H3K27ac H3K64ac and H3K4me3 ordered from high to low H3K122ac. Enhancers were divided into three groups; 1 - H3K27ac peaks (H3K27ac+ active enhancers, n = 23,153), 2 - H3K122ac peaks but not H3K27ac (H3K122ac+/H3K27ac– enhancers, n = 9,340), 3 - none of the above acetylation peaks (H3K27ac–/H3K122ac– inactive enhancers, n = 5,265). Similarly, heatmaps for mESC SEs16 are shown on top. Details of enhancer groups are listed in Supplementary dataset 1.

Europe PMC Funders Author Manuscripts

b) Box plots showing log2 median interquartile distributions of RPM for the three enhancer groups, for H3K122ac, H3K27ac, H3K4me1, H3K64ac, H2A.Zac, H2A.Z, H3K27me3, H4K16ac, H3K4me1, and EP300 ChIP-seq data. Pairwise significance values were calculated using Wilcoxon rank sum test (Supplementary Table 1).

c) Log2 RPM RNA seq reads from *Exosome3* knockout ES cells (*Exosc3*$^{-/-}$)/WT across the three enhancer groups in panel a. Reads from both negative (dotted) and positive (continuous) strands are shown.

d) ChIP-seq data (RP10M) for histone marks across the genetically defined *Nanog* enhancer (*Nanog* en), SEs downstream of *Klf4 (Klf4* SE), *Sox2 (Sox2* SE) and the group 2 putative enhancers downstream of *Sox2* (*Sox2* 40k en and *Sox2* 60k en). H3K27ac, H3K64ac, and H3K122ac ChIP-seq reads are averaged from two biological replicates and individual tracks for *Nanog* and *Klf4* are shown in Supplementary Figure 2. DHS and ChromHmm are shown below the tracks, colour-codes and enrichment values for histone marks across ESC ChromHmm states are in Supplementary Figure 1.

Europe PMC Funders Author Manuscripts

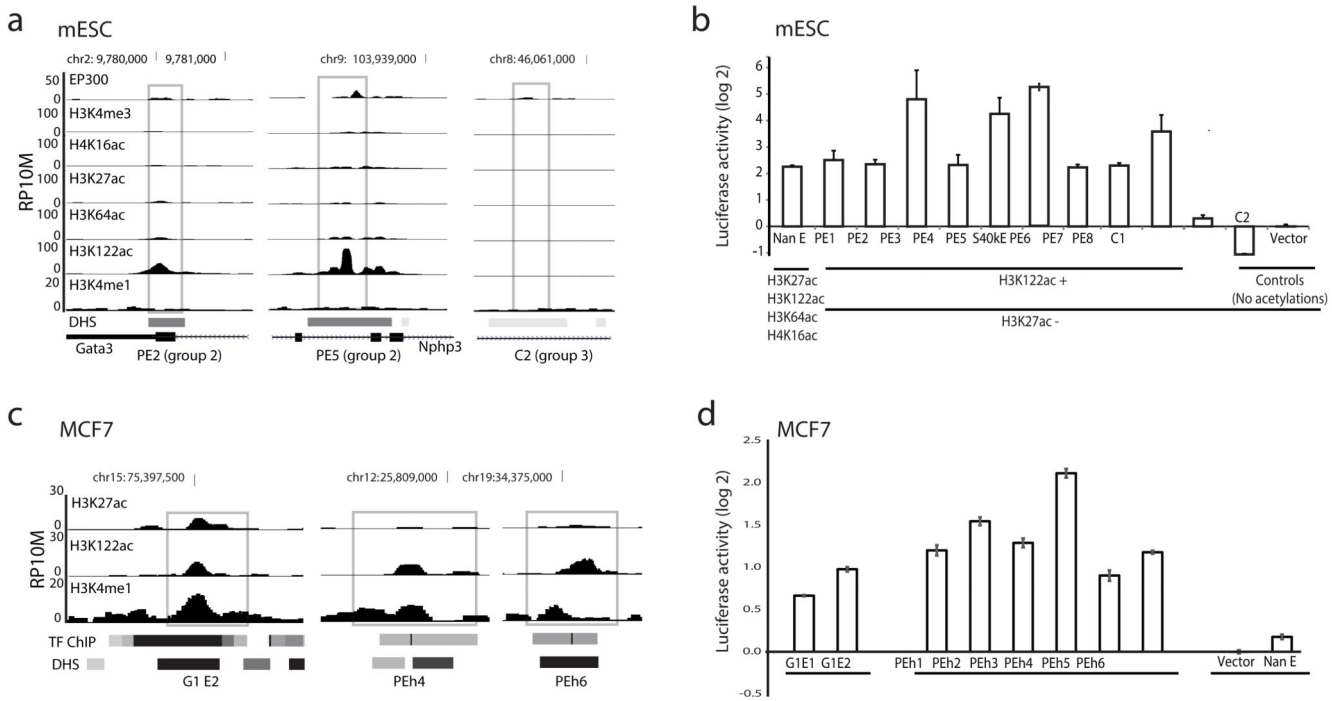

Figure 3. *In vitro* enhancer assays

a) Similar to Fig. 2d, representative H3K27ac– putative Group 2 enhancers from ESC marked with H3K122ac (PE2 and PE5), and a negative control lacking all histone acetylation marks tested (C2). Regions used for cloning into the enhancer reporter vector (pGL4.26) are indicated by grey boxes and detailed in Supplementary Table 3.

b) Luciferase reporter assays for genetically defined enhancer of *Nanog* (Fig. 2d) (Nan E), and randomly chosen H3K27ac negative putative active enhancers based on the presence of H3K122ac (PE1 – PE5, S40kE); H3K64ac (PE6, PE7); both H3K122ac and H3K64ac (PE8). *Sox2* -40k enhancer (S40 kE) region is shown in Fig. 2d. Additionally, regions with H3K4me1 but no acetylation were assayed (C1, C2), and empty vector (pGL4.26) served as negative control. Mean Log2 fold change in luciferase activity was plotted with error bars showing standard error of mean (s.e.m) from two biological and 2 technical replicates (n = 4).

c) Similar to a) but for putative enhancers from MCF7 cells9, transcription factor (TF) ChIP peaks from ENCODE are shown below. Genome co-ordinates are from the GRCh37/hg19 assembly of the human genome.

d) Similar to b) Luciferase assay done in MCF7 cells, for randomly chosen H3K27ac+ enhancers (G1E1, G1E2) and H3K122ac+/H3K27ac– putative human enhancers (PEh1 – PEh6). *Nanog* enhancer (Nan E) and vector alone served as controls. (Supplementary Table 3. Mean log2 fold change in luciferase activity was plotted with error bars showing standard error of mean from two biological and 2 technical replicates (n = 4).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

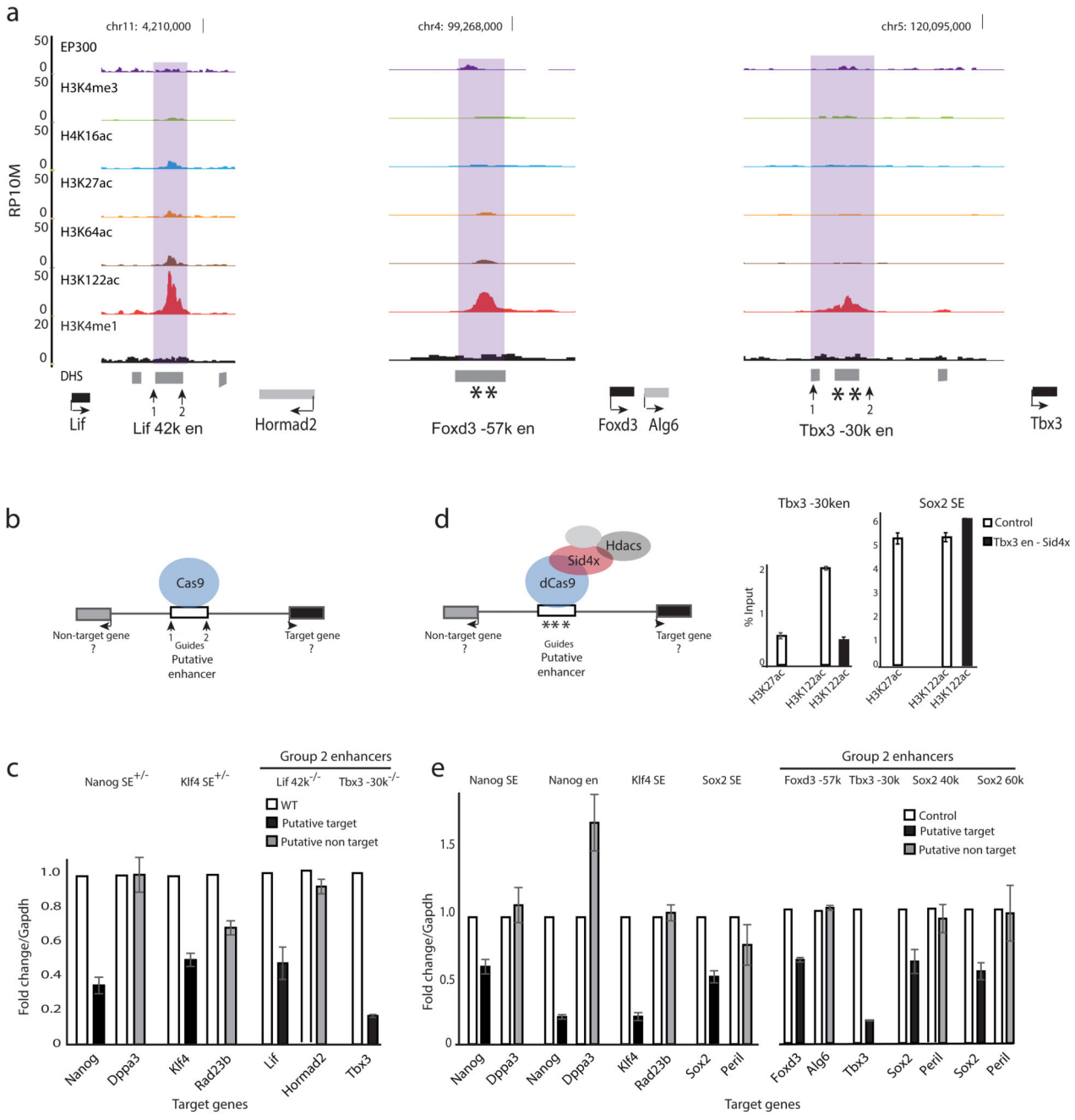

Figure 4. *In vivo* function of group 2 enhancers in gene regulation

a) RP10M, similar to Fig. 2d, but for selected candidate group 2 enhancer regions. Location of Cas9 gRNA targeting sites (arrow-heads) and dCas9-Sid4x (*) are indicated. Putative target (black) and non-target genes (grey) and the direction of transcription are indicated (arrows).

b) Schematic showing CRISPR/Cas9 mediated deletion strategy for enhancers.

c) Mean (± s.e.m) expression of putative enhancer target genes, and flanking genes, assayed by RT-qPCR, normalized to *Gapdh*, in wild-type (WT) ESCs and in ES cells with

heterozygous deletions of the *Nanog* and *Klf4* SEs or homozygous deletions of the putative Group 2 enhancers. *Lif* 42k en, *Foxd3* -20k en and *Tbx3* -30k en (*n* = 3 biological replicates). gRNAs details are given in Supplementary Table 4.

d) Schematics showing Cas9-Sid4x recruitment to enhancers (left). Right; graph showing ChIP-qPCR (mean % input ± s.e.m, n = 3 technical replicates of 2 biological replicates) for H3K27ac and H3K122ac over *Tbx3* -30k en, upon recruitment of dCas9-Sid4x to *Tbx3* -30k en. Non-targeting (control) gRNA plasmids served as control. Enrichment was compared to non-target Sox2 SE (right).

e) As for (c), RTqPCR for putative target genes *Nanog, Klf4, Sox2, Foxd3* and *Tbx3* (black) and neighbouring control genes (grey) in cells transfected with dCas9-Sid4x along with gRNA plasmids targeting *Nanog/Klf4/Sox2* SEs, *Nanog* en. *Foxd3* -20k en, *Tbx3* -30k en, *Sox2* 40k en *and Sox2* 60k en. Non-targeting (control) gRNA plasmids served as control, (*n* = 3 biological replicates). gRNAs details are given in Supplementary Table 5.

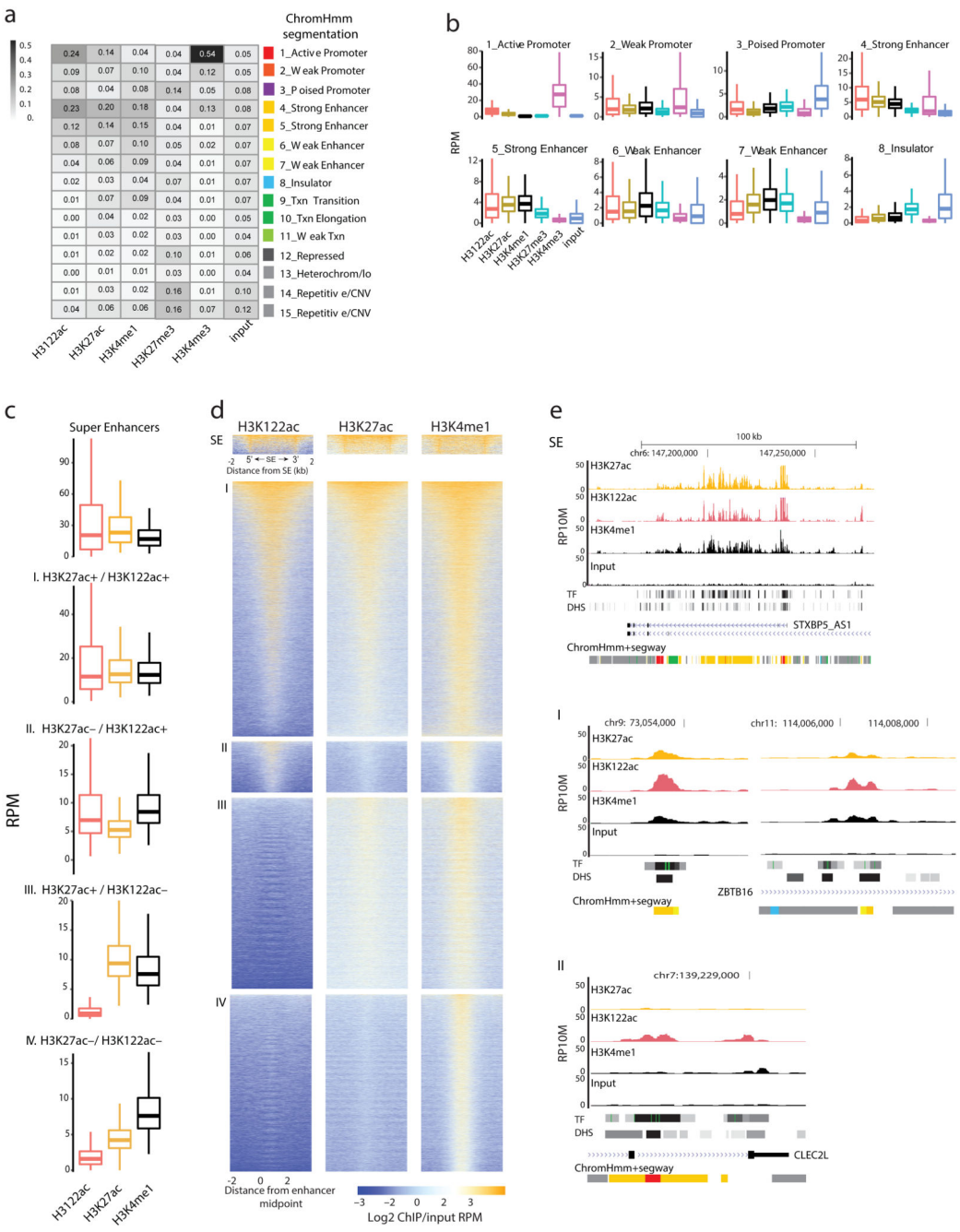

Figure 5. H3K122ac marks at K562 enhancers

a) Enrichment values for H3K122ac, H3K27ac, H3K4me1, H3K27me3, H3K4me3 ChIPs and Input reads from K562 cells across ChromHmm segmentations12.

b) Similar to panel a, boxplots showing log2 ChIP-seq RPM distributions (median value, line inside the box). The interquartile range (IQR) shows 50% of the data, the whiskers extend to 1.5 x IQR.

c and d) Heatmaps and boxplots showing enrichment (RPM) of H3K122ac (red), H3K27ac (Orange) and H3K4me1 (black) in K562 cells across five groups of enhancers – grouped

based on the acetylation patterns. Super-enhancers (SE); enhancers marked with H3K27ac and H3K122ac (I); enhancers lacking H3K27ac but are marked with H3K122ac (II); enhancers with H3K27ac but not H3K122ac (III) and enhancers lack both H3K27ac and H3K122ac (IV).

e) UCSC genome browser tracks (RP10M) showing H3K27ac, H3K122ac, H3K4me1 ChIPs and input from K562 cells for SE and group I and II enhancers. TF ChIP, DHS clusters and K562 ChromHmm+Segway tracks are shown below (colour code in Fig 5a). Genomic co-ordinates of K562 cell enhancers are listed in Supplementary dataset 2.

Europe PMC Funders Author Manuscripts