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

MRC Human Genetics Unit, IGMM, Crewe Road, Edinburgh EH4 2XU, UK

Tel: +44 131 332 2471

Fax: +44 131 467 8456

*Correspondence to:

M.M.P (email:pmadap@essex.ac.uk) or W.A.B. (email:Wendy.Bickmore@igmm.ed.ac.uk)

Keywords: Histone acetylation, H3K122ac H3K64ac, Enhancers, Polycomb

Running title: Globular histone acetylation at enhancers

Competing Interests

The authors declare no competing interests.

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 genes[1]. This includes acetylation of H3 on lysine 27 (H3K27ac), which blocks the deposition of polycomb mediated H3K27me3[2]. H3K27ac is also widely used to identify active enhancers[3,4], 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.

Covalent modifications at the globular domains of the core histones have been implicated in a variety of chromatin functions[5]. 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 structure[5]. The acetylation of H3K56 (H3K56ac) is associated with DNA unwrapping from the nucleosome and has been implicated in chromatin assembly and genome stability[6]. 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 establishment[8]. 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 templates[9] and promote nucleosome disassembly[10].

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,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 enhancers[11]. 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 TSSs[11] (Fig. 2a;). The data clustered into three groups (Supplementary dataset 1 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 status[3,4]. 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.Zac[9] and can induce transcription[14]) 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 K27[7,9,15].

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,d; 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 enhancers[17,18]. A subset of group 2 enhancers with H3K27me3 peaks are enriched for un-methylated CpG islands (CGIs) (Supplementary Figure 4b, which are located at promoters and enhancers[19]. Bidirectional transcription of enhancers correlates with enhancer activity[20], 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

100 eRNAs (Fig. 2c).

101 We tested the enhancer activity of these elements using luciferase reporter assays in
102 mESCs; a well-characterized *Nanog* enhancer[22] (Fig. 2d) served as a positive control. Group
103 2 genomic regions (H3K27ac–) with enrichment for H3K122ac (Fig. 3a,b) exhibited 4 –120
104 fold higher activity compared to negative controls and were equally, or more, active than the
105 *Nanog* enhancer. Similarly, enhancer assays performed in a human breast adenocarcinoma
106 cell line (MCF7) cells showed that H3K27ac–/H3K122ac+ enhancers[9] display higher reporter
107 activity than H3K27ac+ enhancers (Fig. 3c,d).

108 To demonstrate the *in vivo* functional importance of group 2 enhancers, we used
109 CRISPR/Cas9[23] to delete them from the ESC genome (Fig. 4a,b). As positive controls we
110 also deleted one allele of the SE located near *Nanog* and *Klf4* (Fig. 2c). This led to a
111 significant reduction in *Nanog* and *Klf4* expression, respectively (Fig. 4c), but not of *Dppa3* –
112 located 80kb upstream of *Nanog*. Expression of *Rad23b* – 180 kb downstream of *Klf4* is
113 somewhat affected by the intervening enhancer deletion. Homozygous deletion of the
114 putative group 2 enhancer 42kb downstream of *Lif* (Lif 42k en$^{-/-}$) led to reduced expression
115 of *Lif,* but not of the flanking gene *Hormad2* (Fig. 4c). Similarly, deletion of one allele of the
116 putative enhancer 30kb upstream from *Tbx3* (Tbx3 -30k en) led to down regulation of *Tbx3*.

117 To examine whether histone acetylation is important for the function of these new
118 regulatory elements we used dCas9 to recruit the Sid4x repressor complex[24] to them (Fig.
119 4d). As positive controls, recruitment of dCas9-Sid4x to the *Nanog* enhancer, and to the SE
120 of *Nanog Klf4,* and *Sox2,* led to significant reduction in expression of the respective target
121 genes but not other nearby genes (Fig. 4e). For the group 2 enhancers analysed, ChIP-qPCR
122 showed that Sid4x recruitment effectively reduced the levels of H3K122ac at the target Tbx3
123 -30k en, with no effect at the off-target control (Sox2 SE) (Fig. 4d). RT-qPCR analysis
124 showed reduced expression of putative target genes upon Sid4x recruitment to Foxd3 -57k

125 en, Tbx3 -30k en, Sox2 40k en and Sox2 60k en, but not of the control genes (Fig. 4e). Sox2
126 40k en also displayed higher activity in reporter assays (Fig. 3b).

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

136 Rather than a simple definition of active enhancers as being marked by
137 H3K4me1/H3K27ac, a more complex picture of different histone acetylation marks at
138 enhancers is emerging[25]. Our data suggests that using H3K27ac alone gives an incomplete
139 catalogue of the active enhancer repertoire, and that acetylation of H3 at the lateral surface of
140 the histone octamer can be used to identify a novel class of active enhancers that have no
141 significant H3K27ac enrichment.

142 Lysine acetyl transferases (KATs) generally have relaxed substrate specificity, with
143 the exception of KAT8, which acetylates H4K16[25,26] and is critical for the maintenance of
144 ESC pluripotency and differentiation[27,28]. H4K16ac marks active enhancers in ESCs,
145 including some that lack H3K27ac[25]. Like the globular domain acetylations of H3, H4K16ac
146 directly affects chromatin structure by perturbing inter-nucleosomal interactions *in vitro*[29],
147 but not higher-order chromatin structure[25]. The role of most histone acetylation marks at
148 enhancers is unknown, but acetylation in the histone tails can recruit reader proteins such as
149 BRD4 that are thought to be important for enhancer function[30]. This is unlikely to be the case

150 for H3K64 and H3K122 acetylation due to their location at the lateral surface of the histone
151 octamer. Rather, acetylation of these residues is believed to function by directly altering
152 nucleosomal stability and mobility, and by facilitating the binding of activators[5]. The finding
153 of H4K16ac and H3 globular domain acetylations at enhancers suggests that opening of local
154 chromatin structure might be an important facet of enhancer function and may stimulate the
155 identification of yet more regulatory histone PTMs that directly affect the physical properties
156 of the nucleosome.
157
158 **URLs**
159 GREAT: Genomic Regions Enrichment of Annotations Tool,
160 http://bejerano.stanford.edu/great/public/html; R Project for Statistical Computing,
161 https://www.r-project.org; BEDTools suite, http://bedtools.readthedocs.org/en/latest. Super-
162 enhancer database http://www.bio-bigdata.com/SEA; Gene expression omnibus (GEO)
163 server, www.ncbi.nlm.nih.gov/geo; ENCODE/Broad –K562 ChromHmm,
164 http://genome.ucsc.edu/ENCODE/downloads.html; mESC ChromHmm
165 https://github.com/gireeshkbogu/chromatin_states_chromHMM_mm9.
166 http://rsat.sb-roscoff.fr/; NGS plots, https://github.com/shenlab-sinai/ngsplot ;
167
168 **Accession codes**
169 ChIP sequencing data generated in this study have been submitted to the NCBI Gene
170 expression Omnibus (GEO) repository under accession number GSE66023. Other datasets
171 used in this study and their accession numbers are given in Supplementary Table 6.
172
173 **Acknowledgements**
174 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).

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.

References

1. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40,** 897–903 (2008).
2. Kim, T. W. *et al.* Ctbp2 Modulates NuRD-Mediated Deacetylation of H3K27 and Facilitates PRC2-Mediated H3K27me3 in Active ESC Genes During Exit From Pluripotency. *Stem Cells* **33,** 2442–55 (2015).
3. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 21931–6 (2010).
4. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470,** 279–83 (2011).
5. Tropberger, P. & Schneider, R. Scratching the (lateral) surface of chromatin regulation by histone modifications. *Nat. Struct. Mol. Biol.* **20,** 657–61 (2013).
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* **36,** 153–63 (2009).
7. Di Cerbo, V. *et al.* Acetylation of histone H3 at lysine 64 regulates nucleosome dynamics and facilitates transcription. *Elife* 1–23 (2014). 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.* **16,** 777–81 (2009).
9. Tropberger, P. *et al.* Regulation of transcription through acetylation of H3K122 on the lateral surface of the histone octamer. *Cell* **152,** 859–72 (2013).

10. Simon, M. *et al.* Histone fold modifications control nucleosome unwrapping and disassembly. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 12711–12716 (2011).
11. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39,** 311–8 (2007).
12. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–9 (2011).
13. Bogu, G. K. *et al.* Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. *Mol. Cell. Biol.* MCB.00955–15 (2015). doi:10.1128/MCB.00955-15
14. Halley, J. E., Kaplan, T., Wang, A. Y., Kobor, M. S. & Rine, J. Roles for H2A.Z and its acetylation in GAL1 transcription and gene induction, but not GAL1-transcriptional memory. *PLoS Biol.* **8,** e1000401 (2010).
15. Tie, F. *et al.* Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. *Development* **141,** 1129–39 (2014).
16. Whyte, W. a *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–19 (2013).
17. Zentner, G. E., Tesar, P. J. & Scacheri, P. C. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* **21,** 1273–83 (2011).
18. Voigt, P., Tee, W. W. & Reinberg, D. A double take on bivalent promoters. *Genes Dev.* **27,** 1318–1338 (2013).
19. Long, H. K. *et al.* Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *Elife* **2,** 1–19 (2013).
20. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507,** 455–61 (2014).
21. Pefanis, E. *et al.* RNA Exosome-Regulated Long Non-Coding RNA Transcription Controls Super-Enhancer Activity. *Cell* **161,** 774–789 (2015).
22. Jiang, J. *et al.* A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nat. Cell Biol.* **10,** 353–60 (2008).
23. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339,** 823–826 (2013).
24. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500,** 472–6 (2013).
25. Taylor, G., Eskeland, R., Hekimoglu-Balkan, B., Pradeepa, M. & Bickmore, W. A. H4K16 acetylation marks active genes and enhancers of embryonic stem cells, but does not alter chromatin compaction. *Genome Res.* **23,** 2053–2065 (2013).
26. Smith, E. R. *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.* **25,** 9175–9188 (2005).
27. Li, X. *et al.* The Histone Acetyltransferase MOF Is a Key Regulator of the Embryonic Stem Cell Core Transcriptional Network. *Cell Stem Cell* **11,** 163–178 (2012).
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,** 1–23 (2014).
29. Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311,** 844–7 (2006).
30. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153,** 320–34 (2013).

261 **Figure Legends**
262
263
264 **Figure 1. Genomic distribution of H3K122ac and H3K64ac**
265
266 a) H3K122ac and H3K64ac native ChIP-seq reads per million (RPM) around (± 2kb) the
267 transcription start (TSS) and end (TES) sites of genes, separated into quartiles according to
268 gene expression in ESCs from low to high (Q1 – Q4) (n = 2 biological replicates).
269 b) Heatmaps of H3K4me3, H3K27ac, H4K16ac, H3K64ac, H3K122ac, H2A.Zac, H2A.Z,
270 H3K27me3 ChIP-seq, CAGE tags and input chromatin RPM around (± 2kb) TSS of
271 polycomb repressed genes in ESCs.
272 c) Reads per 10 million (RP10M) for ChIP-seq of H4K16ac, H3K27ac, H3K64ac, H2A.Z,
273 H3K122ac, H3K4me3 and H3K27me3 across selected polycomb target genes *Gsc*, *Hoxa9*
274 and *Hoxa7* in ESCs. Genome co-ordinates are from the NCBI37/mm9 assembly of the mouse
275 genome, CpG islands (CGI) and ChromHmm segmentation of these coordinates are shown
276 below (purple=poised promoters; grey=heterochromatin)[13].
277 d) Sequential ChIP-qPCR over promoters of active genes – *Sox2*, *Pou5f1*, polycomb target
278 genes – *Msx1*, *Mash1*, *Hoxd1*, *Hoxa7*, *Cdx2*, *Gsc*, and non-expressed gene *Myf51*. First ChIPs
279 were performed with covalently coupled H3K4me3 (black bars) and H3K27me3 (dark grey
280 bars) antibodies, followed by a second ChIP with H3K122ac antibodies (white and light grey
281 bars for H3K4me3 and H3K27me3 first ChIP, respectively). Primer details given in
282 Supplementary Table 2. Data is a representative of one of two experiments and error bars
283 shows standard error of mean (s.e.m) from 3 technical replicates.
284 e) Heatmap showing the hierarchical clustering of ChIP-seq data for H3K27me3, H3K4me3,
285 H3K27ac, H3K4me1, H3K122ac and H3K64ac. Genome-wide Pearson's correlation
286 coefficient was calculated by dividing the genome into 10kb windows; correlation values
287 among histone modifications are shown.
288

289 **Figure 2. H3K122ac and H3K64ac marks active enhancers in ESC**
290
291 a) Heatmaps of ChIP-seq (RPM) fold change/input around (± 2kb) enhancer midpoints for
292 H3K122ac, H3K27ac H3K64ac and H3K4me3 ordered from high to low H3K122ac.
293 Enhancers were divided into three groups; 1 - H3K27ac peaks (H3K27ac+ active enhancers,
294 n = 23,153), 2 - H3K122ac peaks but not H3K27ac (H3K122ac+/H3K27ac– enhancers, n =
295 9,340), 3 - none of the above acetylation peaks (H3K27ac–/H3K122ac– inactive enhancers, n
296 = 5,265). Similarly, heatmaps for mESC SEs[16] are shown on top. Details of enhancer groups
297 are listed in Supplementary dataset 1.
298 b) Box plots showing log2 median interquartile distributions of RPM for the three enhancer
299 groups, for H3K122ac, H3K27ac, H3K4me1, H3K64ac, H2A.Zac, H2A.Z, H3K27me3,
300 H4K16ac, H3K4me1, and EP300 ChIP-seq data. Pairwise significance values were calculated
301 using Wilcoxon rank sum test (Supplementary Table 1).
302 c) Log2 RPM RNA seq reads from *Exosome3* knockout ES cells (*Exosc3*$^{-/-}$)/WT across the
303 three enhancer groups in panel a. Reads from both negative (dotted) and positive (continuous)
304 strands are shown.
305 d) ChIP-seq data (RP10M) for histone marks across the genetically defined *Nanog* enhancer
306 (*Nanog* en), SEs downstream of *Klf4 (Klf4* SE), *Sox2 (Sox2* SE) and the group 2 putative
307 enhancers downstream of *Sox2* (*Sox2* 40k en and *Sox2* 60k en). H3K27ac, H3K64ac, and
308 H3K122ac ChIP-seq reads are averaged from two biological replicates and individual tracks
309 for *Nanog* and *Klf4* are shown in Supplementary Figure 2. DHS and ChromHmm are shown
310 below the tracks, color-codes and enrichment values for histone marks across ESC
311 ChromHmm states are in Supplementary Figure 1.
312
313 **Figure 3. *In vitro* enhancer assays**

314 a) Similar to Fig. 2d, representative H3K27ac– putative Group 2 enhancers from ESC marked
315 with H3K122ac (PE2 and PE5), and a negative control lacking all histone acetylation marks
316 tested (C2). Regions used for cloning into the enhancer reporter vector (pGL4.26) are
317 indicated by grey boxes and detailed in Supplementary Table 3.
318 b) Luciferase reporter assays for genetically defined enhancer of *Nanog* (Fig. 2d) (Nan E),
319 and randomly chosen H3K27ac negative putative active enhancers based on the presence of
320 H3K122ac (PE1 – PE5, S40kE); H3K64ac (PE6, PE7); both H3K122ac and H3K64ac (PE8).
321 *Sox2* -40k enhancer (S40 kE) region is shown in Fig. 2d. Additionally, regions with
322 H3K4me1 but no acetylation were assayed (C1, C2), and empty vector (pGL4.26) served as
323 negative control. Mean Log2 fold change in luciferase activity was plotted with error bars
324 showing standard error of mean (s.e.m) from two biological and 2 technical replicates (n = 4).
325 c) Similar to a) but for putative enhancers from MCF7 cells[9], transcription factor (TF) ChIP
326 peaks from ENCODE are shown below. Genome co-ordinates are from the GRCh37/hg19
327 assembly of the human genome.
328 d) Similar to b) Luciferase assay done in MCF7 cells, for randomly chosen H3K27ac+
329 enhancers (G1E1, G1E2) and H3K122ac+/H3K27ac– putative human enhancers (PEh1 –
330 PEh6). *Nanog* enhancer (Nan E) and vector alone served as controls. (Supplementary Table
331 3. Mean log2 fold change in luciferase activity was plotted with error bars showing standard
332 error of mean from two biological and 2 technical replicates (n = 4).
333
334 **Figure 4. *In vivo* function of group 2 enhancers in gene regulation**
335 a) RP10M, similar to Fig. 2d, but for selected candidate group 2 enhancer regions. Location
336 of Cas9 gRNA targeting sites (arrow-heads) and dCas9-Sid4x (*) are indicated. Putative
337 target (black) and non-target genes (grey) and the direction of transcription are indicated
338 (arrows).

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

340 c) Mean (± s.e.m) expression of putative enhancer target genes, and flanking genes, assayed
341 by RT-qPCR, normalized to *Gapdh*, in wild-type (WT) ESCs and in ES cells with
342 heterozygous deletions of the *Nanog* and *Klf4* SEs or homozygous deletions of the putative
343 Group 2 enhancers. *Lif* 42k en, *Foxd3* -20k en and *Tbx3* -30k en (*n* = 3 biological replicates).
344 gRNAs details are given in Supplementary Table 4.

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

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

355

356 **Figure 5. H3K122ac marks at K562 enhancers**

357 a) Enrichment values for H3K122ac, H3K27ac, H3K4me1, H3K27me3, H3K4me3 ChIPs
358 and Input reads from K562 cells across ChromHmm segmentations[12].

359 b) Similar to panel a, boxplots showing log2 ChIP-seq RPM distributions (median value, line
360 inside the box). The interquartile range (IQR) shows 50% of the data, the whiskers extend to
361 1.5 x IQR.

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

364 based on the acetylation patterns. Super-enhancers (SE); enhancers marked with H3K27ac
365 and H3K122ac (I); enhancers lacking H3K27ac but are marked with H3K122ac (II);
366 enhancers with H3K27ac but not H3K122ac (III) and enhancers lack both H3K27ac and
367 H3K122ac (IV). (Whiskers are as in panel b).
368 e) UCSC genome browser tracks (RP10M) showing H3K27ac, H3K122ac, H3K4me1 ChIPs
369 and input from K562 cells for SE and group I and II enhancers. TF ChIP, DHS clusters and
370 K562 ChromHmm+Segway tracks are shown below (color code in Fig 5a). Genomic co-
371 ordinates of K562 cell enhancers are listed in Supplementary dataset 2.
372
373 **Online Methods**
374 **Cell culture**
375 46C, Sox1-GFP mouse embryonic stem cells (mESC)[31] were cultured as described
376 previously[25]. Human erythro-myeloblastoid leukemia cells (K562) were cultured in RPMI
377 1640 with L-Glutamine media containing 10% fetal bovine serum (FBS), L-glutamine,
378 penicillin and streptomycin. Cell lines were validated and Mycoplasma tested at IGMM,
379 University of Edinburgh.
380
381 **Sequential Chromatin Immunoprecipitation (ChIP)** Antibodies recognizing H3K122ac
382 and H3K64ac were previously described[7,9]. mESCs were cross-linked in 1% formaldehyde
383 for 10 min and then quenched by the addition of glycine to a final concentration of 0.125 M.
384 Chromatin was sheared using a biorupter (Diagenode) to an average fragment length of ~100
385 – 200bp. Sequential ChIP was performed as described previously[32]. Briefly, 5 µg antibodies
386 against H3K4me3 (07-473, Millipore) and H3K27me3 (07-449, Millipore) were covalently
387 coupled to Dynabeads using Invitrogen antibody coupling kit (Cat. 14311D) according to the
388 manufacturer's instructions. The first ChIP was performed using either H3K4me3 or
389 H4K27me3 antibodies, and the immunoprecipitated chromatin was then eluted with 10 mM

390 DTT, diluted 30 times with RIPA buffer (1X PBS, 1% NP-40, 0.5% Sodium Deoxycholate,
391 0.1% SDS, *Roche Protease Inhibitor Cocktail) before performing the second ChIP with anti-
392 H3K122ac[9]. Purified chromatin was quantified by qPCR using the standard curve method
393 and expressed as % of input bound. Primer details are given in Supplementary Table 2.
394
395 **Native Chromatin Immunoprecipitation**
396 10 x 10^6 mESCs and K562 cells were centrifuged at 500 *g* for 3 min, washed twice in PBS
397 and then resuspended in 200 µl of NBA buffer [85 mM NaCl, 5.5 % Sucrose, 10 mM
398 TrisHCl pH 7.5, 0.2 mM EDTA, 0.2 mM PMSF, 1 mM DTT, 1x Protease inhibitors
399 (Calbiochem, 539134-1SET)]. Cells were lysed by the addition of an equal volume of NBA +
400 0.1 % NP40 and incubated on ice for 3 min. Nuclei were pelleted at 2,000 *g* for 3 min at 4 ºC,
401 then washed with NBR buffer (85 mM NaCl, 5.5 % Sucrose, 10 mM TrisHCl pH 7.5, 3 mM
402 $MgCl_2$, 1.5 mM $CaCl_2$, 0.2 mM PMSF and 1 mM DTT) and pelleted at 2,000 *g* for 3 min at 4
403 ºC. Nuclei were resuspended (10 x10^6 nuclei/ml) in NBR supplemented with RNaseA (20
404 µg/ml) and incubated at 20 ºC for 5 min. Chromatin was fragmented for 30 min at 20 ºC
405 using 0.133 U/µl microccocal nuclease (MNase - Boehringer units; SigmaAldrich - N3755-
406 500UN; titrated to give predominantly mono-nucleosomes). Digestion was stopped with the
407 addition of an equal volume of STOP bufffer (215 mM NaCl, 10 mM TrisHCl pH 8, 20 mM
408 EDTA, 5.5 %, Sucrose, 2 % TritonX 100, 0.2 mM PMSF, 1 mM DTT, 2X Protease
409 Inhibitors) and digested nuclei left on ice overnight to release soluble, fragmented chromatin.
410 Chromatin was pre-cleared by centrifugation at 12,000 *g* for 10 min at 4 ºC and the soluble
411 chromatin (supernatant) transferred to a fresh tube. 5 % of the released chromatin was
412 retained as input and the remainder incubated for 4 h at 4 ºC on a rotating wheel with ~5 µg
413 of antibodies (H3K122ac[9]; H3K64ac[7]; H3K4me1 - Abcam ab8895, lot:GR251663-1;
414 H3K27ac - Abcam ab4729, lot:GR254707-1) pre-coupled to protein A dynabeads (Life

415 Technologies; 10002D) in PBS containing 5 mg/ml BSA and 0.1 mM PMSF. Immune
416 complexes bound to beads were washed 5x with wash buffer 1 (150 mM NaCl, 10 mM
417 TrisHCl pH 8, 2 mM EDTA, 1 % NP40 and 1 % sodium deoxycholate) on a rotating wheel
418 for 5 min each and once in room temperature TE buffer for 1 min. Chromatin was released
419 from the beads by incubation with 0.1 M $NaHCO_3$ / 1 % SDS for 30 min at 37 ºC followed by
420 the addition of proteinase K (100 ug/ml) and Tris pH 6.8 (100 mM) and incubation at 55 ºC
421 overnight. For both native and cross-linked ChIP, Dynabeads were removed using a magnetic
422 rack and the chromatin purified using Qiaquick PCR Purification columns (Qiagen)
423 according to the manufacturer's instructions.

424 **ChIP-seq library preparation and Deep Sequencing**

425 Libraries were prepared as previously described[33] with the following modifications: No
426 purification was performed between the A-tailing and ligation reactions. After A-tailing
427 reaction, enzymes were inactivated by incubation at 75 ºC for 20 min. and the ligation
428 reaction was supplemented with ligation reagents [400 U of T4 DNA ligase (NEB), 1x buffer
429 2 (NEB), 7.5 % PEG-6,000, 1 mM ATP and 13.3 nM of annealed Illumina adaptors (AU)]
430 and incubated at 16 °C overnight. Size selection following the ligation and PCR steps was
431 performed with 1x and 0.8x reaction volumes of Agencourt AMPure XP beads respectively
432 (Beckman Coulter - A63880).

433 Replicate 1 of the H3K122ac and H3K64ac ChIPs was sequenced at The Danish
434 National High-Throughput DNA sequencing Center (Copenhagen; 42 base single end reads).
435 Replicate 2 of the H3K122ac and H3K64ac ChIPs, 2 replicates of H3K27ac ChIPs and all
436 ChIP and input samples prepared from K562 cells were sequenced at Edinburgh Genomics
437 (The University of Edinburgh, 50 base single end reads).
438
439

440 **Read mapping**

441 FASTQ files were aligned using Bowtie[34] (version 0.12.8) with parameters set to retain
442 uniquely mapped reads with a maximum of two mismatches (bowtie options: -e 40 -m 1 -v
443 2). For mapping, mm9 and hg18 bowtie indexes were used for mouse (mESC) and human
444 (K562 and MCF7) datasets respectively. Mapped reads from two biological replicates of
445 H3K27ac, H3K122ac and H3K64ac were merged for further analysis.

446

447 **Peak calling**

448 Peaks were called using SICER[35]. For mESC, MNase-digested ChIP input DNA
449 (GSM1156619) was used as a background control for H3K27ac, H3K64ac and H3K122ac.
450 For H3K4me1 in ESCs (E14TG2a; GSM1003750), Input (GSM1003746) was used as a
451 background control. mESC biological replicates were merged using SAMtools (v0.1.19) prior
452 to peak calling with SICER (v1.1). SICER parameters: window size – 200 bp; fragment size
453 – 150 bp; false discovery rate – 0.01; gap size – 600 bp for H3K122ac, H3K64ac, and
454 H3K4me1 and a 200bp window size for H3K27ac.

455

456 **Generation of Bedgraphs for visualisation on UCSC genome browser**

457 Bedgraphs for each histone mark were generated from the aligned read files using the
458 HOMER software suite (v4.7)[36], at a resolution of 10 bp and with a normalized tag count of
459 10 million. Mapped reads from two biological replicates for H3K122ac, H3K64ac and
460 H3K27ac ChIPs in mESCs were combined for the generation of Bedgraphs for Figure 1 to 4.
461 UCSC tracks for individual replicates covering representative loci are shown in
462 Supplementary Figure 2. Similarly, data from single experiments for MCF7 and K562
463 ChIPseq reads were processed to generate Bedgraphs for visualization in UCSC genome
464 browser.

465

466 **Heatmaps and average profiles**

467 Heatmaps and average profile for Refseq gene transcription start sites (TSS; ± 2kb), Refseq
468 gene transcription end sites (TES; ± 2kb from), enhancer midpoints (± 2 kb from) and for
469 entire length of super-enhancers (all scaled to an equivalent length ± 2 kb), were generated
470 using ngsplot v2.61[37].

471 For Figure 1a, gene expression quartiles from high (Q4) to low (Q1) were obtained from our
472 previous study[25] and used to generate average profile plots for H3K122ac and H3K64ac
473 across TSS and TES as detailed above.

474 The Heatmap for Figure 1b was generated for TSSs (± 2kb) of genes which have been shown
475 to be repressed by polycomb complexes[38].

476 The average profile plots (Figure 2a) for enrichment of strand specific RNA-seq reads in
477 *Exosc*$^{-/-}$/WT (SRP042355)[21] for the 3 enhancer groups were generated using ngsplot[37]
478 (v2.61).

479

480 **Genome-wide correlation analysis of histone marks.**

481 Pearson's correlation coefficients were calculated between datasets using bamCorrelate tool[39]
482 (version 1.5.9, removing duplicate reads and a using a resolution of 10 kb). The correlation
483 matrix was hierarchically clustered and visualized using the Bioconductor package pheatmap.

484

485 **ChromHMM analysis**

486 To calculate the distribution of histone marks against different chromatin states the
487 bamCorrelate tool was used to count reads within chromHMM segments for K562[12] and
488 mESCs[13]. Datasets were normalized to read per million (RPM).

489

490 **Enhancer analysis**

491 Enhancers were defined as H3K4me1 peaks, with gene TSSs (RefSeq TSS ± 2kb) and
492 genome blacklist[40] regions removed. Active enhancer regions (group 1) were defined as
493 genomic intervals overlapping both H3K4me1 and H3K27ac peaks. Inactive enhancers,
494 defined as peaks of H3K4me1 with no associated H3K27ac peak, were stratified into group 2
495 and 3 representing those with and without an associated H3K122ac peak, respectively
496 (Supplementary datasets 1 and 2). Peak intersections were performed using the BEDtools[41]
497 (v2.23.0) intersect function. Super-enhancer co-ordinates for K562 cells and mESC were
498 obtained from super-enhancer archive. H3K27me3 peak regions were called using MACS2[42]
499 (v2.1.0, broadpeak with no input control).

500

501 **TF motif enrichment analysis**

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

506

507 **Gene ontology (GO) enrichment analysis**

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

513

514 **Enrichment analysis of DNaseI hypersensitivity sites (DHS) and un-methylated CpG**
515 **islands (CGIs)**[19]
516 To determine the enrichment of DHS (GSM1014154) and CGIs at subgroups of enhancers; a
517 Fisher's exact test was performed using BEDtools fisher (default options) [41]. Un-methylated
518 CGI for mESCs were obtained from GSE43512.
519
520 **Dual luciferase enhancer assays**
521 Putative enhancer regions were PCR amplified from mouse (E14TG2a ESC) or human
522 (HepG2) genomic DNA, cloned into pGL4.26 plasmid and sequence verified. Details of
523 enhancers and PCR products used in this assay are given in the Supplementary Table 3.
524 Putative enhancers from mESCs were assayed in E14TG2a mESCs and putative MCF7 cell
525 enhancers from were assayed in MCF7 cells. Forty-eight hours post-transfection, a luciferase
526 assay was performed using the Dual-luciferase Reporter assay (Promega) as per the
527 manufacturer's instructions. Firefly luciferase activity was normalized to transfection
528 efficiency with Renilla luciferase activity using pRL-TK. All values are shown as log2 ratios
529 of enhancer activity vs. empty vector.
530
531 **Enhancer deletions**
532 Pairs of gRNAs (Supplementary Table 4) designed to direct Cas9 to regions flanking putative
533 enhancers, were cloned into SpCas9-2A-GFP (PX458, Addgene number 48138) and
534 transfected using lipofectamine 2000 (Invitrogen) into 46C ESCs[44]. 24 hours after
535 transfection, transfected cells were FACS sorted for GFP + and are seeded at the 5,000
536 cells/100mm dish. Surviving colonies were isolated and screened for deletion by PCR and
537 homozygous clones were verified by Sanger sequencing. RT-qPCR was performed as

538 described previously[32], altered gene expression upon deletion of enhancer elements was
539 measured v/s wild type control.
540
541 **Sid4x recruitment to enhancers**
542 The repressive mSin3 Interaction Domain (Sid4x) was cloned C-terminal of dCas9 (pAC-
543 Sid4x) by replacing VP160 from dCas9VP160-2A-puro (pAC94)[45]. 2-3 guides per enhancer,
544 or 5 – 7 guides per super-enhancer (Supplementary Table 5), were designed and oligos were
545 synthesized from Sigma or IDT and cloned into pSLQ sgRNA expression plasmid as
546 described[46]. All clones were verified by Sanger sequencing. Equal ratios of guideRNA pools
547 and dCas9-Sid4x plasmids were co-transfected into mESCs using Lipofectamine 2000. 24
548 hours after transfection puromycin (2µg/ml) was added to the media. Surviving transfected
549 cells were harvested 48 hrs post transfection and RT-qPCR was performed as described[32] and
550 native ChIP was performed for H3K122ac and H3K27ac. ChIP enrichment was calculated as
551 the percentage input bound by the standard curve method. As a control pAC-Sid4x was
552 transfected along with non-targeting pSLQ sgRNA plasmid.
553
554 **Online References**
555
556 31. Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic
557 stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.*
558 **21,** 183–186 (2003).
559 32. Pradeepa, M. M., Grimes, G. R., Taylor, G. C. a, Sutherland, H. G. & Bickmore, W. A.
560 Psip1/Ledgf p75 restrains Hox gene expression by recruiting both trithorax and
561 polycomb group proteins. *Nucleic Acids Res.* **42,** 9021–32 (2014).
562 33. Bowman, S. K. *et al.* Multiplexed Illumina sequencing libraries from picogram
563 quantities of DNA. *BMC Genomics* **14,** (2013).
564 34. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient
565 alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25
566 (2009).
567 35. Zang, C. *et al.* A clustering approach for identification of enriched domains from
568 histone modification ChIP-Seq data. *Bioinformatics* **25,** 1952–8 (2009).
569 36. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors
570 Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol.*

571 *Cell* **38,** 576–589 (2010).
572 37. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of
573 next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15,**
574 284 (2014).
575 38. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two
576 classes of bivalent domains. *PLoS Genet.* **4,** e1000242 (2008).
577 39. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible
578 platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42,** W187–91 (2014).
579 40. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome.
580 *Nature* **489,** 57–74 (2012).
581 41. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
582 genomic features. *Bioinformatics* **26,** 841–842 (2010).
583 42. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137
584 (2008).
585 43. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory
586 regions. *Nat. Biotechnol.* **28,** 495–501 (2010).
587 44. Ran, F., Hsu, P., Wright, J. & Agarwala, V. Genome engineering using the CRISPR-
588 Cas9 system. *Nat. Protoc.* **8,** 2281–308 (2013).
589 45. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an
590 RNA-guided transcriptional activator system. *Cell Res.* **23,** 1163–1171 (2013).
591 46. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an
592 optimized CRISPR/Cas system. *Cell* **155,** 1479–91 (2013).
593

a

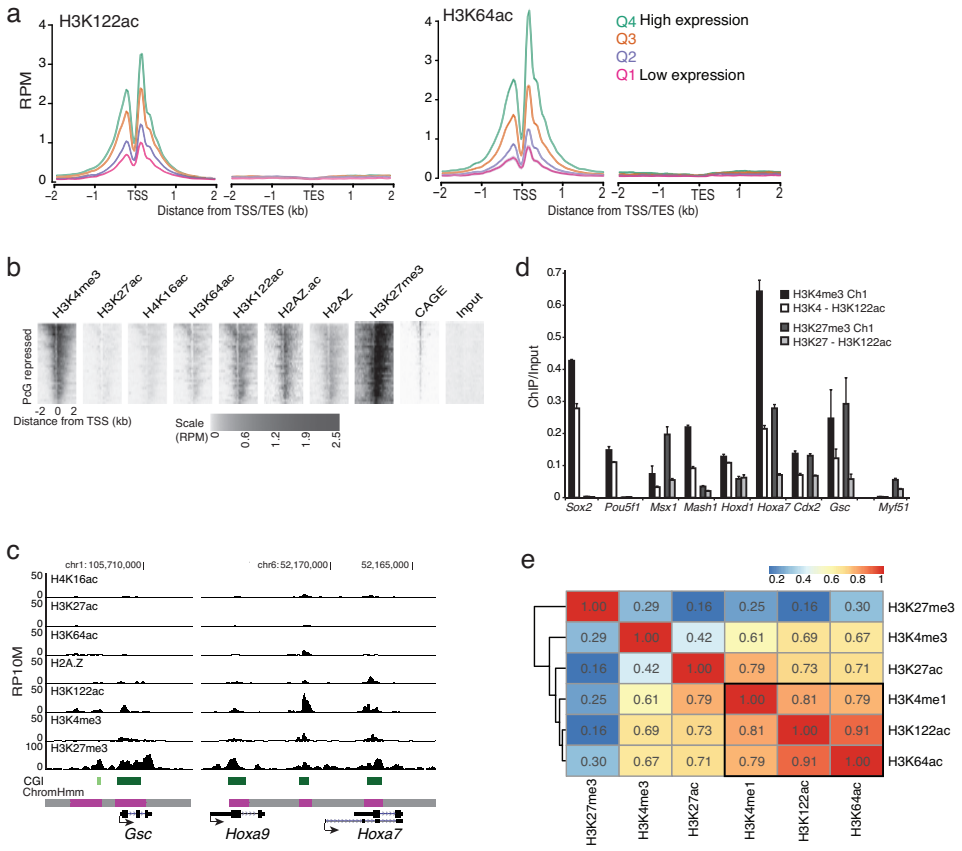
a
H3K122ac
H3K64ac
RPM
Q4 High expression
Q3
Q2
Q1 Low expression
TSS
TES
Distance from TSS/TES (kb)
b
H3K4me3
H3K27ac
H4K16ac
H3K64ac
H3K122ac
H2AZ.ac
H2AZ
H3K27me3
CAGE
Input
P-G repressed
Distance from TSS (kb)
Scale (RPM)
c
chr1: 105,710,000
chr6: 52,170,000
52,165,000
RP10M
H4K16ac
H3K27ac
H3K64ac
H2AZ
H3K122ac
H3K4me3
H3K27me3
CGI
ChromHmm
Gsc
Hoxa9
Hoxa7
d
ChIP/Input
H3K4me3 Ch1
H3K4 - H3K122ac
H3K27me3 Ch1
H3K27 - H3K122ac
Sox2
Pou5f1
Msx1
Mash1
Hoxd1
Hoxa7
Cdx2
Gsc
Myf51
e
H3K27me3
H3K4me3
H3K27ac
H3K4me1
H3K122ac
H3K64ac

Figure 1

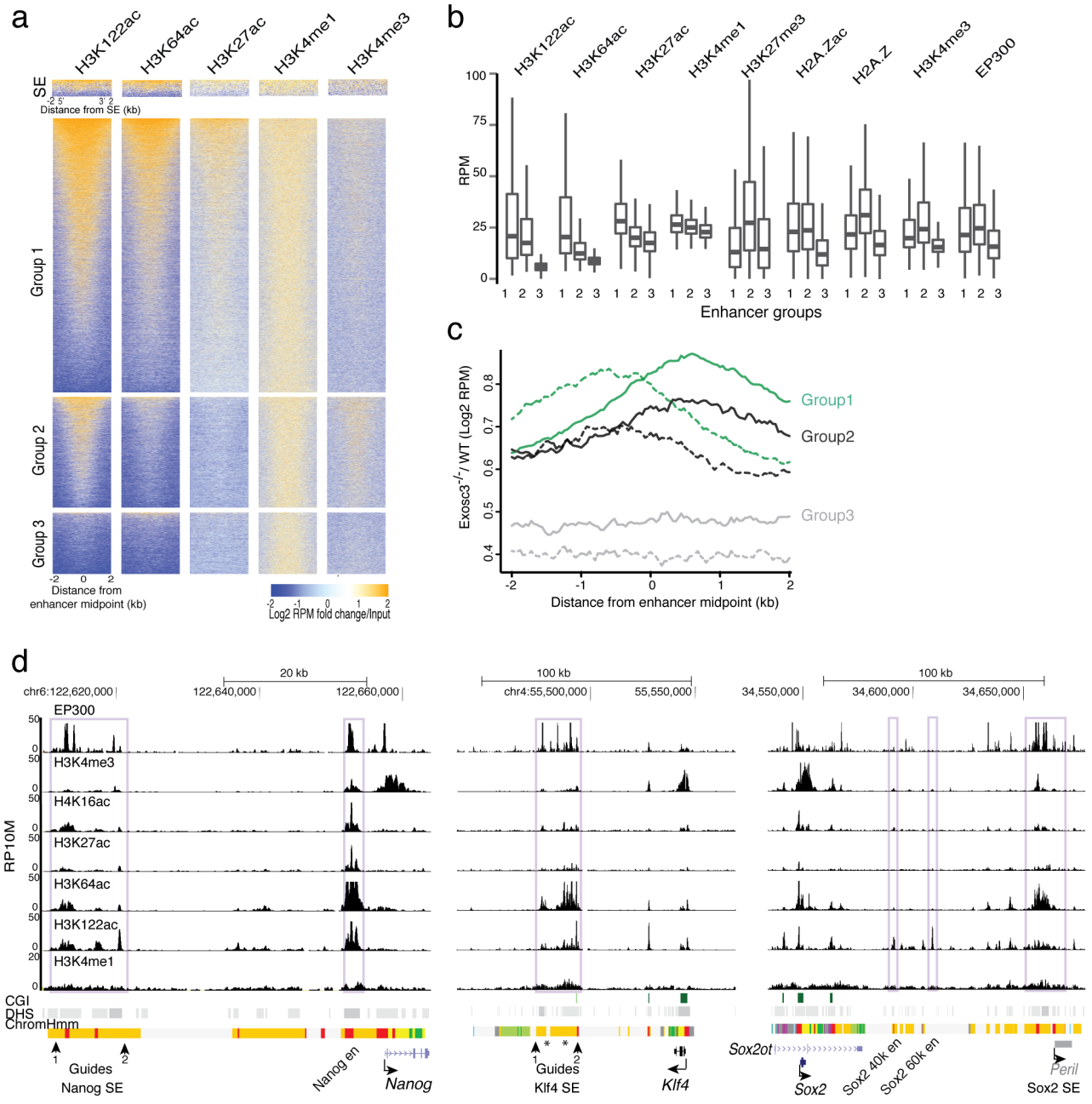
a
H3K122ac
H3K64ac
H3K27ac
H3K4me1
H3K4me3
SE
Distance from SE (kb)
Group 1
Group 2
Group 3
Distance from enhancer midpoint (kb)
Log2 RPM fold change/Input
b
H3K122ac
H3K64ac
H3K27ac
H3K4me1
H3K27me3
H2A.Zac
H2A.Z
H3K4me3
EP300
RPM
Enhancer groups
c
Exosc3-/- / WT (Log2 RPM)
Group1
Group2
Group3
Distance from enhancer midpoint (kb)
d
20 kb
chr6:122,620,000
122,640,000
122,660,000
100 kb
chr4:55,500,000
55,550,000
100 kb
34,550,000
34,600,000
34,650,000
EP300
H3K4me3
H4K16ac
H3K27ac
H3K64ac
H3K122ac
H3K4me1
RP10M
CGI
DHS
ChromHmm
Guides
Nanog SE
Nanog en
Nanog
Guides
Klf4 SE
Klf4
Sox2ot
Sox2
Sox2 40k en
Sox2 60k en
Peril
Sox2 SE

Figure 2

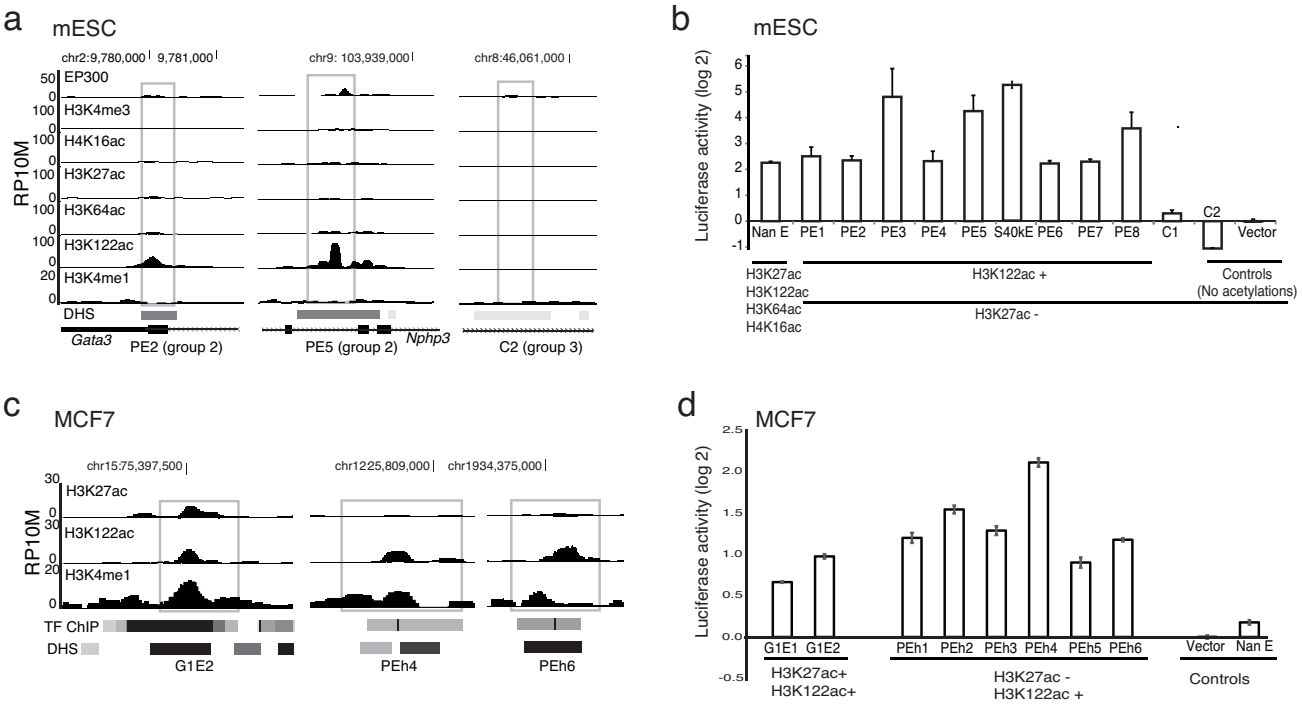
a
mESC
chr2:9,780,000
9,781,000
chr9: 103,939,000
chr8:46,061,000
EP300
H3K4me3
H4K16ac
H3K27ac
H3K64ac
H3K122ac
H3K4me1
RP10M
DHS
Gata3
PE2 (group 2)
PE5 (group 2)
Nphp3
C2 (group 3)
b
mESC
Luciferase activity (log 2)
Nan E
PE1
PE2
PE3
PE4
PE5
S40kE
PE6
PE7
PE8
C1
C2
Vector
H3K27ac
H3K122ac
H3K64ac
H4K16ac
H3K122ac +
H3K27ac -
Controls
(No acetylations)
c
MCF7
chr15:75,397,500
chr1225,809,000
chr1934,375,000
H3K27ac
H3K122ac
H3K4me1
RP10M
TF ChIP
DHS
G1E2
PEh4
PEh6
d
MCF7
Luciferase activity (log 2)
G1E1
G1E2
PEh1
PEh2
PEh3
PEh4
PEh5
PEh6
Vector
Nan E
H3K27ac+
H3K122ac+
H3K27ac -
H3K122ac +
Controls

Figure 3

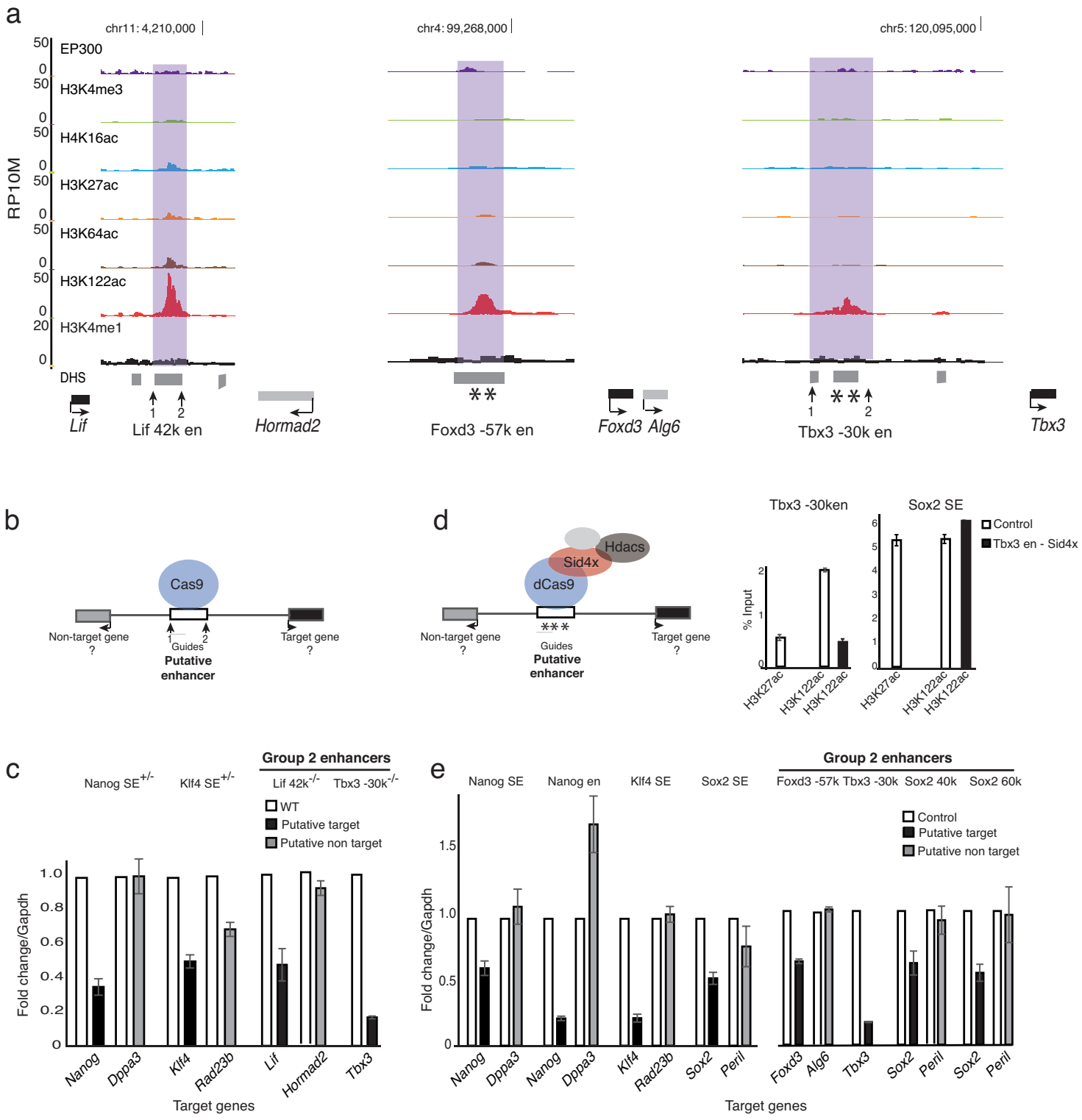
a
chr11: 4,210,000
chr4: 99,268,000
chr5: 120,095,000
EP300
H3K4me3
H4K16ac
H3K27ac
H3K64ac
H3K122ac
H3K4me1
DHS
RP10M
Lif
Lif 42k en
Hormad2
Foxd3 -57k en
Foxd3
Alg6
Tbx3 -30k en
Tbx3
b
Cas9
Non-target gene
Guides
Target gene
Putative enhancer
d
Hdacs
Sid4x
dCas9
Tbx3 -30ken
Sox2 SE
Control
Tbx3 en - Sid4x
% Input
H3K27ac
H3K122ac
c
Group 2 enhancers
Nanog SE+/-
Klf4 SE+/-
Lif 42k-/-
Tbx3 -30k-/-
WT
Putative target
Putative non target
Fold change/Gapdh
Nanog
Dppa3
Klf4
Rad23b
Lif
Hormad2
Tbx3
Target genes
e
Nanog SE
Nanog en
Klf4 SE
Sox2 SE
Foxd3 -57k
Tbx3 -30k
Sox2 40k
Sox2 60k
Control
Putative target
Putative non target
Foxd3
Alg6
Sox2
Peril

Figure 4

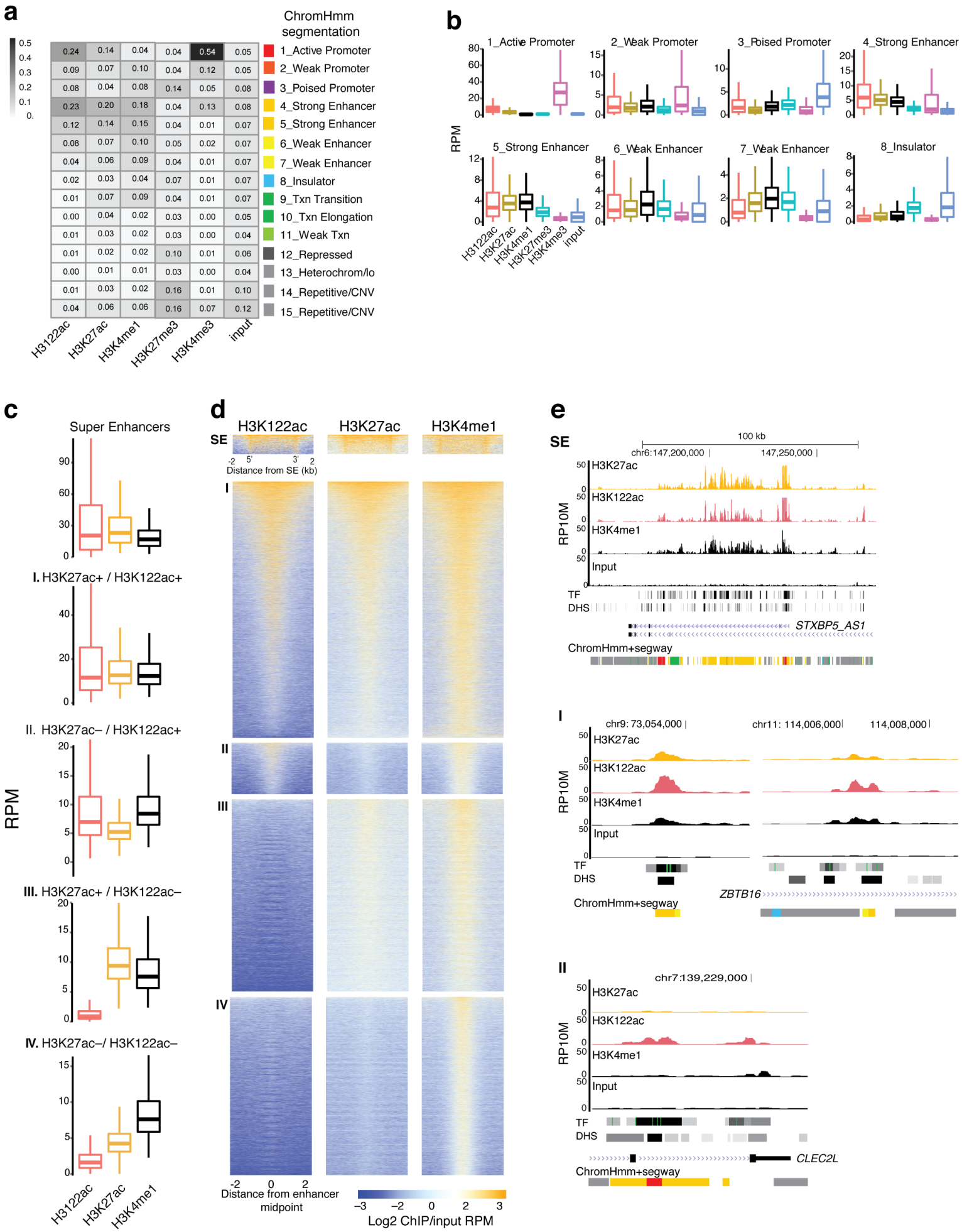
a
ChromHmm segmentation
1_Active Promoter
2_Weak Promoter
3_Poised Promoter
4_Strong Enhancer
5_Strong Enhancer
6_Weak Enhancer
7_Weak Enhancer
8_Insulator
9_Txn Transition
10_Txn Elongation
11_Weak Txn
12_Repressed
13_Heterochrom/lo
14_Repetitive/CNV
15_Repetitive/CNV
H3122ac
H3K27ac
H3K4me1
H3K27me3
H3K4me3
input
b
RPM
c
Super Enhancers
I. H3K27ac+ / H3K122ac+
II. H3K27ac– / H3K122ac+
III. H3K27ac+ / H3K122ac–
IV. H3K27ac–/ H3K122ac–
d
H3K122ac
H3K27ac
H3K4me1
SE
Distance from SE (kb)
Distance from enhancer midpoint
Log2 ChIP/input RPM
e
100 kb
chr6:147,200,000
147,250,000
RP10M
TF
DHS
STXBP5_AS1
ChromHmm+segway
chr9: 73,054,000
chr11: 114,006,000
114,008,000
ZBTB16
chr7:139,229,000
CLEC2L

Figure 5