



HAL
open science

The histone variant macroH2A1.1 regulates gene expression by direct association with their transcription start site

Ludmila Recoules, Alexandre Heurteau, Flavien Raynal, Fatima Moutahir, Fabienne Bejjani, Isabelle Jariel-Encontre, Olivier Cuvier, Anne-Claire Lavigne, Kerstin Bystricky

► To cite this version:

Ludmila Recoules, Alexandre Heurteau, Flavien Raynal, Fatima Moutahir, Fabienne Bejjani, et al.. The histone variant macroH2A1.1 regulates gene expression by direct association with their transcription start site. 2020. hal-03037462

HAL Id: hal-03037462

<https://hal.science/hal-03037462v1>

Preprint submitted on 3 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

The histone variant macroH2A1.1 regulates gene expression by direct association with their transcription start site.

Ludmila Recoules¹, Alexandre Heurteau¹, Flavien Raynal¹, Fatima Moutahir¹,
Fabienne Bejjani³, Isabelle Jariel-Encontre³, Olivier Cuvier¹, Anne-Claire Lavigne^{1*}
and Kerstin Bystricky^{1,2*}.

¹ Laboratoire de Biologie Moléculaire Eucaryote (LBME), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, UPS, F-31062 Toulouse, France.

² Institut Universitaire de France (IUF).

³ Institut de Génétique Moléculaire de Montpellier, CNRS, UMR5535, Equipe Labellisée Ligue Nationale contre le Cancer, F-34293, France.

*Corresponding authors.

E-mails : kerstin.bystricky@ibcg.biotoul.fr ; anne-claire.lavigne@ibcg.biotoul.fr

25 **Abstract**

26 The histone variant macroH2A1 (mH2A1) is involved in cellular growth,
27 differentiation and reprogramming, but the underlying molecular mechanisms are a
28 matter of debate. Different roles of mH2A1 in gene expression may relate to
29 functional differences of its two splicing isoforms, mH2A1.1 and mH2A1.2. Here, we
30 map for the first time genome-wide localization of endogenous mH2A1.1 and link the
31 distribution of mH2A1.1 to control of gene expression in human breast cancer cells.
32 In addition to localization shared with mH2A1.2 to facultative heterochromatin,
33 mH2A1.1 specifically associates with regulatory elements required for gene
34 activation, super-enhancers and promoters of highly expressed genes. Depending on
35 the recruitment profile of mH2A1.1 to these elements, selective depletion of mH2A1.1
36 up- or downregulates its target genes. mH2A1.1 represses transcription when its
37 binding is spread over the entire gene and promoter, and activates transcription
38 when its binding is strictly confined to the transcription start site (TSS). Notably, RNA
39 Polymerase II was frequently in pause at mH2A1.1-activated genes. Functionally,
40 mH2A1.1-dependent regulation of a subset of paused genes impedes mammary
41 tumor cell migration. Molecular mechanisms of mH2A1.1 function at the TSS
42 uncovered by our study define an intriguing new mode of transcription regulation in
43 cancer cells.

44

45 **Author Summary**

46

47 Control of gene expression driving cellular functions from differentiation to
48 epistasis and causing, when dysfunctional, uncountable diseases, relies on
49 modifications of chromatin structure. One key element enabling chromatin plasticity is

50 the replacement of canonical histones by histone variants. Among histone variants
51 macroH2A1 (mH2A) is an extraordinary H2A variant possessing a large non-histone
52 domain placed outside of the nucleosome. Two splicing isoforms, mH2A1.1 and
53 mH2A1.2, are produced, but these are rarely studied separately because they only
54 differ in a 30 amino acid region and are difficult to distinguish experimentally, which
55 likely explains contradictory functions reported in the literature. Here, we take
56 advantage of a mH2A1.1 specific antibody to generate the first genome-wide
57 chromatin-associated map of this histone variant in the invasive breast cancer cells
58 line MDA-MB231. We confirm that mH2A1.1, like mH2A1.2, is enriched at facultative
59 heterochromatin in agreement with its reported role as a repressor. However, we
60 discovered that unlike its splicing isoform, mH2A1.1 specifically binds to super-
61 enhancers and the transcription start site of highly transcribed genes. mH2A1.1 is
62 necessary for regulating transcription of these genes. At the cellular level, we
63 demonstrate that mH2A1.1 inhibits migration capacity of highly metastatic breast
64 cancer cells. Our study characterizes for the first time binding profiles of mH2A1.1
65 that are linked to regulation of gene expression, thereby providing a new molecular
66 mechanisms which govern the plasticity of human tumor cells.

67

68 **Introduction**

69 Compaction of DNA into chromatin modulates DNA accessibility, thereby
70 regulating key molecular processes such as transcription [1,2]. Histone post-
71 translational modifications, chromatin-remodeling enzymes, DNA-binding factors and
72 architectural proteins fine-tune genome organization and dynamics [1,2]. In addition,
73 histone variants replace canonical histones in a locus-specific manner, which endows

74 chromatin with additional properties required to regulate DNA accessibility and
75 functions [3].

76 Among the histone variants, macroH2A1 (mH2A1) is a vertebrate-specific [4,5]
77 histone H2A variant composed of an N-terminal H2A-like domain (64 % identical to
78 H2A) and a C-terminal 25 kDa ‘macro’ domain. These two domains are joined by an
79 unstructured 41 amino acid long linker that positions the macro domain outside of the
80 nucleosome [6]. Expression of the highly conserved *H2AFY* gene produces two
81 splicing isoforms, mH2A1.1 and mH2A1.2, whose sequences differ in a 30 amino-
82 acid region within the macro domain [6].

83 mH2A1 was originally found to be enriched on the transcriptionally silent X
84 chromosome [7]. mH2A1 is also present at autosomes, forming large domains in
85 association with histone marks associated with heterochromatin, such as H3K27me3
86 and H3K9me3 [8–10]. *In vitro* studies have demonstrated that nucleosomal mH2A1
87 interferes with binding of the transcription factor NFκB, and inhibits nucleosome
88 sliding by the remodeling complex SWI/SNF and initiation of RNA polymerase II (Pol
89 II) transcription [11,12]. Therefore, mH2A1 is believed to play a role in transcriptional
90 repression. However, in a few cases the presence of mH2A1 has been seen to
91 correlate with active transcription of a subset of genes involved in a variety of
92 processes such as cell differentiation, lipid metabolism and cell-cell signaling [8,13–
93 17]. Thus, the roles of mH2A1 in regulating gene expression are seemingly
94 contradictory. Studying mH2A1 isoforms separately may help to better understand
95 their roles.

96 The two mH2A1 splice variants exhibit tissue- and cell-specific expression
97 patterns[18]. In normal cells, the mH2A1.2 isoform appears ubiquitously expressed
98 [19–21]. Mainly incorporated into heterochromatin (X-inactive chromosome and

99 autosomic heterochromatin) [7–9,22], this splicing isoform is generally associated
100 with gene repression. However, a few recent studies discovered that mH2A1.2 was
101 required for gene expression. Indeed, in mouse muscle cells [13], mH2A1.2 binds
102 muscle-specific enhancers necessary for the activation of the myogenic regulator
103 network. In human embryonic stem cells (hESCs) [23], binding of mH2A1.2 to gene
104 promoters seems to control gene expression both positively and negatively.
105 However, the binding profile of mH2A1.2 in hESC was not conserved in mesoderm-
106 derived human fibroblasts [23].

107 In contrast to mH2A1.2, mH2A1.1 is only expressed in differentiated cells with
108 low proliferation rates [19–21]. In tumors, expression of the mH2A1.1 isoform is
109 frequently reduced, as compared to normal tissues, suggesting that this isoform is a
110 tumor suppressor [20,21,24]. In highly metastatic cancers such as triple-negative
111 breast cancers however, expression levels of mH2A1.1 are increased and correlate
112 with poor prognosis [24].

113 Only mH2A1.1 can bind NAD⁺ metabolites through its macro domain [25] and
114 interact with the DNA-damage repair and chromatin remodeling factor PARP1 (Poly-
115 (ADP-Ribose) Polymerase 1) [14,26–28]. Interaction between mH2A1.1 and PARP1
116 seems to be key for the capacity of mH2A1.1 to regulate DNA damage responses
117 [29,30], mitochondrial respiration [28] and gene transcription [14,27,31]. In particular,
118 it was proposed that mH2A1.1 and PARP1 recruit CBP (CREB-binding protein) to
119 mediate acetylation of H2BK12/K120 and to regulate, in part, mH2A1-target gene
120 expression in IMR90 cells, without really demonstrated the specific action of
121 mH2A1.1 with respect to mH2A1.2 [14]. In other hand, recruitment of mH2A1.1 to
122 repressed signal-inducible genes is required for their stress response [31,32].
123 Despite this knowledge about functions of mH2A1.1, neither its chromatin

124 association, independently of the one of mH2A1.2, nor its specific mechanisms of
125 action during gene activation, was clearly analyzed.

126 To gain a better understanding of the roles of mH2A1.1 in regulating gene
127 expression, we determined 1) the genomic localization of mH2A1.1 and 2) its effect
128 on gene expression. To that end, we generated a ChIP-grade mH2A1.1-specific
129 antibody. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) of
130 mH2A1.1, of total mH2A1 and of a variety of histone marks, shows that both mH2A1
131 isoforms bind to heterochromatin and to enhancers. Surprisingly, only the mH2A1.1
132 isoform binds TSSs (Transcription Start Site) of active genes. Interestingly, we show
133 that mH2A1.1-activated genes are mostly Pol II paused genes. Frequently, these
134 genes are negatively implied in cell migration in mammary tumor cells. We
135 demonstrate that mH2A1.1 impedes cell migration whereas mH2A1.2 promotes it.
136 Our work describes a novel gene activation pathway in cancer cells dependent on
137 the selective recruitment of mH2A1.1 to the TSS of paused genes.

138

139 **Results**

140

141 **mH2A1.1 and mH2A1.2 preferentially associate with facultative**
142 **heterochromatin.** We determined the genomic localization of mH2A1.1 in the
143 claudin-low breast cancer cell line MDA-MB231 which expresses this histone variant
144 at a high level compared to other types of breast cancer cell lines (S1A, S1B Fig)
145 [24]. To that end, we developed a ChIP-grade polyclonal rabbit antibody that
146 exclusively recognizes mH2A1.1 (Ab α mH2A1.1) (S1C-S1F Fig) By ChIP-seq, we
147 determined the distribution of mH2A1.1 and of total mH2A1 using Ab α mH2A1.1 and

148 a commercially available ChIP-grade antibody (Ab37264 (Ab α mH2A1)) (S1, S2
149 Tables and S1G-S1I Fig), respectively.

150 We identified 29,112 peaks for mH2A1.1 (Ab α mH2A1.1) and 22,789 peaks for
151 total mH2A1 (Ab α mH2A1), covering combined 13.6 % of the genome. Genome-
152 wide, ChIP-seq results obtained with these two antibodies were highly similar with a
153 significant co-occurrence of peaks (Fisher exact test (FET): p-value $< 2.2 \times 10^{-16}$ and
154 Odd ratio = 27.40) and a Pearson coefficient correlation (PCC) of 0.92 (S2A Fig).

155 Analysis of peaks detected with Ab α mH2A1 and Ab α mH2A1.1 shows that
156 regions occupied by mH2A1.1 and mH2A1.2 do not overlap completely (Fig 1A)
157 suggesting that some regions are preferentially bound by either one of the variants.
158 Indeed, a significant number of sites (18,838) is preferentially recognized by Ab
159 α mH2A1.1 and not by Ab α mH2A1 (Fig 1A). Of note, Ab α mH2A1 has lower efficacy
160 than Ab α mH2A1.1 in detecting mH2A1.1 and may, when reducing detection
161 thresholds, recognize a fraction of these sites (S1B, S1G, S1H Fig). Genomic regions
162 common to both antibodies correspond to one or more peaks exclusively recognized
163 as mH2A1.1 or both isoforms in unknown relative proportions. For further analysis,
164 we defined three types of genomic regions: regions incorporating preferentially only
165 mH2A1.1 (mH2A1.1 only), regions incorporating preferentially only mH2A1.2
166 (mH2A1.2 only) and regions incorporating either mH2A1.1 or both isoforms
167 (mH2A1s) (Fig 1A).

168 These ChIP-seq results raise the possibility that preferential genomic
169 localization of the mH2A1.1 variant relate to its function. To test this possibility, we
170 first characterized the distribution of mH2A1.1 ChIP-seq peaks with respect to
171 selected genomic features (Fig 1B; see Materials and Methods). Nearly one third of

172 mH2A1.1 peaks were associated with distal intergenic regions and another third of
173 mH2A1.1 peaks with promoters (TSS \pm 1kb) (Fig 1B).

174 Next, we analyzed mH2A1.1 genomic localization with respect to the
175 chromatin environment of the identified regions. We first integrated ENCODE ChIP-
176 seq of heterochromatin histone marks (H3K9me3 and H3K27me3) [33]. We found
177 that genome-wide, mH2A1.1 ChIP-seq peaks were significantly enriched with the
178 H3K27me3 repressive histone mark, in a manner comparable to the one observed
179 using Ab α mH2A1 (Figs 1C and S2A). We confirmed the enrichment of mH2A1.1 on
180 heterochromatin domains in independent ChIP-qPCR experiments on WT and
181 mH2A1.1-deficient cells using two different siRNAs directed against mH2A1.1
182 isoform (S3A-S3E and S4 Figs). Remarkably, even if the related PCCs were low
183 between H3K9me3 and mH2A1 isoforms (S2A Fig), in particular for mH2A1.1, we
184 detected that both mH2A1 isoforms overlapped significantly with H3K9me3 (Fig 1C).
185 At heterochromatin domains (i.e. marked by H3K27me3 and/or H3K9me3), both
186 mH2A1 ChIP-seq peaks were highly similar with a PCC of 0.94 (S2B Fig).
187 Surprisingly, 80% of H3K27me3 overlapped with H3K9me3 (Fig 1C), although often
188 presented as non-overlapping heterochromatin histone marks [34]. However, within
189 peaks common to H3K27me3 and H3K9me3, the contribution of each mark was
190 inversely proportional (Figs 1D and S5A). In addition, a high H3K27me3 to H3K9me3
191 ratio was more associated with genes whereas low H3K27me3 to H3K9me3 ratio
192 was more associated with gene-poor genomic regions (Fig 1D and S5B). Hence, in
193 this cell line, levels of relative enrichment between H3K27me3 and H3K9me3 could
194 be used to distinguish “facultative-like” heterochromatin (with high levels of
195 H3K27me3 and low levels of H3K9me3), from “constitutive-like” heterochromatin
196 (with low levels of H3K27me3 and high levels of H3K9me3). In this context, we

197 observed that the genomic distribution of both mH2A1 isoforms positively correlates
198 with regions rich in H3K27me3 and poor in H3K9me3 levels, hence corresponding to
199 “facultative-like” heterochromatin (Fig 1E).

200

201 **mH2A1.1 binds to super-enhancers.** Because a large number of mH2A1.1 sites
202 are associated with promoter regions (Fig 1B), we performed ChIP-seq experiments
203 for a panoply of histone marks usually associated with active transcription
204 (H3K4me1, H3K4me3, H3K27ac and H3K36me3) (*Bejjani et al, in preparation*). We
205 found that mH2A1.1 correlated positively with H3K4me1 and H3K27ac, two
206 chromatin modifications which characterize active enhancer regions [35] (Figs 2A
207 and S2A). We found that more than 40% of mH2A1.1 and total mH2A1 overlapped
208 significantly with “putative” enhancers outside of TSSs (Fig 2B; see Materials and
209 Methods). 17% of “putative” enhancers were bound by mH2A1 isoforms. Among
210 them, some were common to mH2A1.1 and total mH2A1 (48%) but many were only
211 bound by mH2A1.1 (33%) or by mH2A1.2 (19%). At “putative” enhancers, both
212 mH2A1 ChIP-seq results were still similar with a PCC of 0.80 (S2B Fig). Interestingly,
213 mH2A1-bound regions frequently formed large domains comprising a group of
214 enhancers marked with H3K27ac (Fig 2C), which could correspond to super-
215 enhancers (SE) [36,37]. Using the ROSE package to detect SEs marked by the
216 H3K27ac signal [36,37], we identified “putative” SEs in MDA-MB231 cells (Materials
217 and Methods). We show that 85% of these SEs were bound by at least one mH2A1
218 isoform, 19% of them were associated with mH2A1.1 compared to only 3% with
219 mH2A1.2 only (Fig 2D). The enrichment of mH2A1.1 at super-enhancers was
220 confirmed in independent ChIP-qPCR experiments in WT and mH2A1.1-deficient

221 cells using two different siRNAs directed against the mH2A1.1 isoform (S3A-S3E and
222 S4 Figs).

223

224 **mH2A1.1 binds to the transcription start site of active genes.** A striking genomic
225 feature identified for mH2A1.1 was its binding to TSSs (+/- 1kb) (hereafter referred to
226 as “promoter regions”) (Fig 1B) confirmed by independent ChIP-qPCR experiments
227 in WT and mH2A1.1-deficient cells by two different siRNAs directed against
228 mH2A1.1 isoform (S3A-S3E and S4 Figs). Therefore, we characterized the
229 association of mH2A1 isoforms at these genomic sites in more detail. Our ChIP-seq
230 data showed that mH2A1.1 was more frequently associated with promoter regions
231 than mH2A1 (41% vs 35%, respectively) (Fig 3A). Interestingly, among promoter
232 regions bound by mH2A1 isoforms, 45% of them were occupied by mH2A1.1 only,
233 whereas 14% were occupied by mH2A1.2 only (Fig 3A). We thus wondered whether
234 the chromatin environment of promoter regions correlates with specific binding of
235 either isoform. At promoter regions, we found that mH2A1.1 positively correlated with
236 active marks (H3K4me3 and Pol II) whereas mH2A1 positively correlated with
237 repressive histone marks (H3K27me3 and H3K9me3) (S6A Fig). Interestingly, this
238 difference in localization identified by the two antibodies was maintained when we
239 considered the two genomic populations specific to each antibody, mH2A1.1 only
240 and mH2A1.2 only (S6B, S6C Fig). These results strongly suggest that mH2A1.2
241 preferentially binds to promoters in a closed chromatin state whereas mH2A1.1 is
242 mainly recruited to promoters in an open chromatin state. In agreement, at promoter
243 regions, both mH2A1 ChIP-seq experiments differed from one another with a PCC of
244 0.41 (in comparison to a PCC of 0.92 genome-wide) (S2A, S2B Fig). To assess
245 whether mH2A1.1 binding at the TSS correlates with transcription levels, we

246 generated RNA-seq data from MDA-MB231 cells. We then stratified genes into four
247 equal categories according to their expression levels (from silent to highly expressed
248 genes). We determined the distribution of both mH2A1 isoforms at the TSS (+/- 2 kb)
249 and over the gene body (region from the TSS to the Transcription End Site (TES)) for
250 those four groups of genes (see Materials and Methods). We found that the amount
251 of mH2A1.1 bound to TSSs increased with expression levels while relative amounts
252 recruited onto gene bodies decreased (Figs 3B, 3C and S7A-S7D). On the contrary,
253 mH2A1 bound uniformly to TSSs and over the gene body of silent genes and its
254 presence was restricted to the proximal promoter regions (around 1kb upstream
255 TSSs) and downstream of TESs of expressed genes (Figs 3B, 3C and S7A-S7D). In
256 agreement, mH2A1.1 peaks were mainly present at the TSS +/- 1kb of highly
257 transcribed genes (Fig 3D), while mH2A1 peaks were present on the TSSs +/- 1kb of
258 silent genes (Fig 3D). Similar observations were seen for active and repressive
259 histone marks (S7E Fig). Here again, this difference in localization identified by the
260 two antibodies was even stronger when we considered the two genomic populations
261 specific to each antibody, mH2A1.1 only and mH2A1.2 only (Fig 3D). Notably, we
262 observed that mH2A1.1 binding at TSSs was distributed uniformly around TSSs,
263 peaking at maximum Pol II binding (Fig 3E). In agreement, at TSSs, 70% of
264 mH2A1.1 peaks (two third of them being mH2A1.1 only) overlapped significantly with
265 Pol II (Fig 3F), with a PCC of 0.48 (in comparison to a PCC of 0.07 genome-wide)
266 (S2A and S7F Figs). Coherent with these results, mH2A1.2 only peaks were hardly
267 detectable (3%) at Pol II TSS-bound sites (Fig 3F), and at TSSs, both mH2A1 ChIP-
268 seq are anti-correlated with a PCC of -0.07 (S2B Fig). These results highlight highly
269 specific association of the mH2A1.1 variant with the TSSs of transcribed genes.
270

271 **mH2A1.1 regulates gene expression.** Binding of mH2A1.1 to TSSs of transcribed
272 genes prompted us to test whether mH2A1.1 was required for modulating gene
273 expression. We generated RNA-seq data from MDA-MB231 cells in which mH2A1.1
274 protein expression was reduced by around 90 % by siRNA without significantly
275 affecting expression of the mH2A1.2 isoform (S3A-S3C Fig). 533 genes (56.3%)
276 were down-regulated (mH2A1.1-activated genes) and 412 genes (43.7%) were up-
277 regulated (mH2A1.1-repressed genes) in mH2A1.1 knock-down (mH2A1.1 KD)
278 conditions compared to WT conditions (Fig 4A, S8A and S3 Table). Altered gene
279 expression was confirmed by RT-qPCR on a subset of genes using two different
280 siRNAs directed against the mH2A1.1 isoform (S3 Fig). We found that all mH2A1.1
281 regulated genes, activated- as well as repressed- genes, were active or even highly
282 expressed genes in WT conditions (S8B and S8C Fig). Strikingly, we revealed that
283 mH2A1.1 and mH2A1 binding differed between mH2A1.1-activated and repressed
284 genes (Figs 4B-4D and S9A). Indeed, mH2A1.1 was mainly bound upstream of the
285 TSS and within the gene body of mH2A1.1-repressed genes. At mH2A1.1-activated
286 genes, however, mH2A1.1 exclusively associated with the TSSs and upstream (up to
287 1kb) (Figs 4B-4D and S9A). mH2A1 was bound upstream and downstream of the
288 TSS of mH2A1.1-repressed genes, whereas its level was at its lowest at the TSS. In
289 contrast, its relative enrichment was very low at mH2A1.1-activated genes, with its
290 highest levels upstream TSS (up to 1 kb) (Figs 4B-4D and S9A). Overall, these
291 results suggest that mH2A1.1 promotes transcription when its binding is restricted to
292 TSSs and represses transcription when its binding is spread over genes.

293

294 **mH2A1.1 regulates expression of paused genes.** We noted that Pol II was bound
295 to the TSSs of mH2A1.1-activated genes but was barely detected on the

296 corresponding gene bodies (Figs 4B-4D and S9A). This pattern of Pol II distribution
297 raised the possibility that mH2A1.1-activated genes may be in pause. To confirm this,
298 we calculated the Pol II pausing index (PI) for transcribed genes using Pol II ChIP-
299 seq data as described in [38] (see Materials and Methods). To see if the binding of
300 mH2A1.1 at TSSs was related to the level of Pol II pausing, we first plotted the
301 mH2A1.1 ChIP-seq signal around the TSS +/- 10 kb for each gene ranked by their PI
302 (Fig 5A). We observed that confinement of mH2A1.1 to the TSS and its absence
303 from the gene body is a characteristic of genes with a high PI. Genes where
304 mH2A1.1 was spread along the entire gene body were characterized by low PIs (Fig
305 5A). In agreement, we found that mH2A1.1 peaks, as well as Pol II ones, were
306 significantly sharper at the TSS (TSS +/- 1kb) of genes with a high PI as opposed to
307 genes with a low PI (Figs 5B, 5C and S9B). Moreover, 40% of genes bound by
308 mH2A1.1 only at the TSS have a PI > 2 and 31% of paused genes (defined as genes
309 with PI>2, n=6,821) have mH2A1.1 only peaks at their TSS (Fig 5D). Furthermore,
310 the correlation between mH2A1.1 binding and activation of paused genes appeared
311 to be functionally relevant because the majority of mH2A1.1-activated genes (68%)
312 have a pausing index greater than 2 and their PI are significantly higher than that of
313 any other gene category tested (Fig 5E and 5F). In contrast, the majority of
314 mH2A1.1-repressed genes (69%) was characterized by a PI < 2. These results
315 suggest that mH2A1.1 spreading over the gene body represses genes whose
316 expression is independent of Pol II pausing, whereas mH2A1.1 restricted to the TSS
317 activates genes whose expression is dependent of Pol II pausing. Altogether, our
318 results provide the first evidence of a functional link between the binding of mH2A1.1
319 at TSSs and Pol II pausing at promoter proximal regions. We propose that the

320 association of mH2A1.1 with the TSSs of a subset of paused genes enhances their
321 transcription.

322

323 **mH2A1.1 inhibits cell migration in MDA-MB231 cells.** We found that genes de-
324 regulated by the loss of mH2A1.1 were involved in four main processes: cell cycle,
325 DNA repair, cytoskeleton organization and cell adhesion (S10A, S10B Fig and S4
326 Table). The two first processes were expected based on earlier studies [21,29,30,39].
327 However, the relationship between mH2A1.1 and expression of genes involved in
328 cytoskeleton organization or cell adhesion was undocumented. Using a standard
329 wide-field microscope, we observed that cells became more elongated after
330 transfecting two different siRNA against mH2A1.1 (Fig 6A and 6B). We used a siRNA
331 against mH2A1.2 (S3A-S3C Fig) to test whether this effect was specific of mH2A1.1.
332 Interestingly, we observed that mH2A1.2 KD cells became rounder compared to
333 control cells (Fig 6A-6B). Importantly, we noticed that numerous mH2A1.1 KD de-
334 regulated genes modulate cell migration (examples of anti-migratory genes:
335 ARRDC3 [40], SOCS4 [41], HACE1 [42] and FBXL4 [43] - and of pro-migratory
336 genes: EIF6 [44], MT1E [45], JUND [46] and DAPK3 [47]). Thus, we investigated the
337 effect of mH2A1.1 and mH2A1.2 depletion on the migratory capacity of MDA-MB231
338 cells using a Boyden Chamber assay (see Materials and Methods). Upon depletion
339 of mH2A1.1, the migratory capacity of MDA-MB231 cells was significantly increased
340 compared to control cells. In contrast, depleting the mH2A1.2 isoform led to a
341 decrease in cell migration (Fig 6C and 6D). Strikingly, mH2A1.1-activated genes
342 involved in cytoskeleton organization and cell adhesion were also amongst genes
343 with a high Pol II pausing index (Fig 6E; see Materials and Methods). Taken together,

344 these results suggest that mH2A1.1 inhibits cell migration by in part enhancing the
345 expression of paused genes involved in cytoskeleton organization and cell adhesion.

346

347 **Discussion**

348

349 In this study, we present the first map of the histone variant mH2A1.1 genomic
350 distribution over the entire human genome in breast cancer cells. Integration of this
351 map with chromatin features confirms that mH2A1.1, together with its splice variant
352 mH2A1.2, localizes to “facultative-like” heterochromatin domains and enhancers.
353 Importantly, for the first time in cancer cells, we identify a direct link between
354 mH2A1.1 recruitment of transcribed genes and their transcriptional regulation. The
355 impact of this recruitment on the transcriptional rate is functionally dichotomous, with
356 positive or negative effects. This bivalence correlates with a differential distribution of
357 mH2A1.1 around the TSS as well as to the level of dependency of the transcribed
358 gene to the Pol II pausing.

359

360 The results of our mH2A1 ChIP-seq analysis are overall consistent with
361 previously published reports identifying this histone variant as a widespread
362 chromatin feature on autosomes, covering large domains, with a preferential
363 recruitment to H3K27me3-decorated “facultative-like” heterochromatin (Fig 1C)
364 [8,9,14,48]. Our mH2A1.1 ChIP-seq analysis allows us to show that even if mH2A1.1
365 distribution is mainly reminiscent of that of mH2A1.2, each variant has also specific
366 heterochromatin localizations. Genome mapping of mH2A1.2 is not yet possible due
367 to the poor specificity or lack of precipitation efficacy of commercial antibodies.

368 We further demonstrate that mH2A1.1-bound chromatin co-localizes
369 significantly with the H3K9me3 histone mark (Fig 1C). A fraction of these sites are
370 devoid of H3K27me3 and could correspond to the recently identified macroH2A
371 localization at constitutive heterochromatin⁸. However, the vast majority of mH2A1.1-
372 bound H3K9me3-decorated chromatin contained also tri-methylated H3K27 (Fig 1C).
373 This difference may be a feature of the MDA-MB231 cell line, a high migratory
374 capacity cancer cell line in which there is an abnormal expansion of frequently
375 overlapping H3K9me3 and H3K27me3 marks along the genome [49,50].

376 Moreover, we noted that even though mH2A1.1 associates with
377 heterochromatin domains, its partial loss alone is not sufficient to reactivate silenced
378 genes present in these domains (S8 Fig). Similarly, even if mH2A1 binding was
379 shown to overlap with H3K27me3 modified chromatin in primary human cells, no
380 enrichment of H3K27me3 at mH2A1-regulated genes was observed [14].
381 Furthermore, mH2A1.2-occupied and repressed target genes are not reactivated
382 upon mH2A1.2 knock-down [13]. These observations suggest that enrichment of
383 both mH2A1 isoforms in heterochromatin domains may serve as a lock to conserve
384 heterochromatin stability and architecture [9].

385

386 In parallel, we show that the depletion of mH2A1.1 alone is sufficient to alter
387 the expression of many genes present in chromatin domains devoid of H3K27me3
388 (Figs 4A and S8C). Previous work identified a role for mH2A1.1 in the transcriptional
389 rate of genes encompassed in H2B acetylated regions but this mechanism was
390 described as specific of primary cells and absent from cancer cells [14]. Here, we
391 demonstrate that the effect of mH2A1.1 depletion correlates with the distribution of
392 mH2A1.1 along transcribed genes. Indeed, mH2A1.1 binding is restricted to the

393 TSSs of mH2A1.1-activated genes while mH2A1.1 binds both upstream and on the
394 gene body of mH2A1.1-repressed genes (Fig 4B). Our findings reinforce the notion
395 that mH2A1.1 has a dual role in gene regulation and this seems to depend on its
396 distribution on mH2A1.1-regulated genes [14]. Molecular mechanisms determining
397 mH2A1.1 differential localization and the consequent effects on gene regulation
398 should now to be determined. Perhaps, mH2A1.1 post translational modifications
399 favor binding to the TSS of active genes. Indeed, mH2A1-S137 phosphorylation
400 excluded mH2A1 from the inactive X chromosome [51]. Specific protein partners
401 could also be involved in the recruitment of mH2A1.1 to the TSSs of transcribed
402 genes, for instance PARP1. Indeed, association of mH2A1.1 and PARP1 was shown
403 to be involved in gene transcription, DNA repair and global cellular metabolism
404 [14,18,29].

405 The vast majority of mH2A1.1 narrow peaks at TSSs was only recognized by
406 our home-made antibody Ab α mH2A1.1 (Fig 3). We think that this is certainly due to
407 the high affinity of Ab α mH2A1.1 towards mH2A1.1 compared to Ab α mH2A1 (S1F-
408 S1H Fig). The quantity of mH2A1.1-restricted to the TSSs is significantly smaller than
409 in the large domains, such as heterochromatin and SEs (Figs 1D,2C and 3C).
410 Therefore, Ab α mH2A1 is certainly less efficient in immunoprecipitating mH2A1.1 in
411 this context. Having a high affinity antibody led us to investigate new localizations
412 and functions of the mH2A1.1 isoform. Now, it will be interesting to have a high
413 affinity antibody against mH2A1.2. Indeed, through native-ChIP-seq experiments
414 using a mH2A1.2 custom-made specific antibody, Dimitris Thanos and his
415 collaborators recently showed that mH2A1.2 binds to the TSSs of transcribed genes
416 in human embryonic stem cells [23]. Thus, adapted sequencing techniques, isoform-
417 specific antibodies and appropriate extractions of mH2A1 chromatin sub-complexes

418 will enable studying mH2A1 splicing isoforms in different chromatin- and cell-
419 dependend contexts.

420 Interestingly, we remarked that mH2A1.1 TSS-binding is distributed in a
421 uniform manner around the TSS placing the maximum binding to the nucleosome
422 free region (NRF) (Fig 3E). Its localization around the TSSs was also observed for
423 mH2A1.2 using Native ChIP-seq [23], eliminating the hypothesis of a technical bias.
424 The long- and unstructured- linker domain places the macro domain of mH2A1
425 outside nucleosomes [6]. Because antibodies against mH2A1 frequently recognize
426 the linker or the macro domain (as Ab α mH2A1.1 and Ab α mH2A1, respectively), we
427 hypothesize that mH2A1.1 is incorporated at the two adjacent nucleosomes of the
428 TSSs but their macro domain could be joined at the NFR, which could explain the
429 higher signal at the NFR compared to the two adjacent nucleosomes. Future analysis
430 will be necessary to confirm this hypothesis.

431

432 At the cellular level, we observed that the silencing of mH2A1.1 promotes cell
433 migration, whereas mH2A1.2 silencing increases migration in the MDA-MB231
434 breast cancer cell line. mH2A1 has been shown to be involved in cell migration in
435 mouse [52,53] and human [52–55] models. The opposite roles of both splice variants
436 on the migratory capacities that we see are in agreement with observations in
437 previous studies using gastric cancer cells [55] and MDA-MB231 cells [52]. At the
438 molecular level, the selective incorporation of mH2A1.2 along the SOD3 gene
439 (playing a part in cell migration) has been shown to be directly linked to gene
440 repression in mouse metastatic cells [52]. Here, we show that the migratory
441 capacities of human metastatic breast cancer cells are directly regulated by selective
442 recruitment of mH2A1.1. Among mH2A1.1-activated genes were genes responsible

443 for inhibiting cell migration. Thus, it is tempting to speculate that the negative effect of
444 mH2A1.1 on cell migration is due to mH2A1.1-dependent activation of genes
445 inhibiting migration. Interestingly, even if mH2A1.2 is very similar to mH2A1.1 in
446 terms of amino-acid sequence, it has an opposite function to mH2A1.1 in the
447 regulation of cell migration. Possibly the NAD⁺ metabolite binding pocket of mH2A1.1
448 plays a role in directing function. Here, we decipher in part the molecular mechanism
449 by which mH2A1.1 could inhibit cell migration. Taken together, our data reveals
450 antagonistic cellular functions of both mH2A1 isoforms and highlight the need to
451 distinguish these isoforms when studying the role of mH2A1.

452

453 We demonstrate that mH2A1.1 favors expression of paused genes (70% of
454 mH2A1.1-activated genes) (Fig 5E). It will be now interesting to dissect the molecular
455 mechanisms that allow mH2A1.1 to promote transcription of paused genes. Perhaps,
456 mH2A1.1 binding to the TSS could stimulate the recruitment of P-TEFb (Positive
457 Transcription Elongation Factor) and consequently allow Pol II release [38]. PARP1,
458 well known partner of mH2A1.1, is also involved in Pol II pausing release by
459 mediating ADP-ribosylation of NELF (Negative elongation factor) [56] and could be
460 recruited by mH2A1.1. This hypothesis could also explain why we observed
461 mH2A1.1 at the TSS of highly transcribed genes (Fig 3B and 3D). The presence of
462 mH2A1.1 at SEs could also induce Pol II release (Fig 2C and 2D). SEs are involved
463 in cellular identity through the regulation of key genes involved in cellular identity
464 [36,57]. Furthermore, SEs are known to play an important part in many diseases,
465 including several cancers⁵¹⁻⁵³ in which they drive expression of oncogenes [37,58].
466 To note, several clinical trials already utilize SE blockers (bromodomain and extra-
467 terminal motif (BET) inhibitor and CDK7i) [59]. Among them, BRD4, one of the BET

468 protein family members, was targeted. BRD4 binds acetylated histones at TSSs and
469 SEs, brings them together, and mediates transcriptional activation and elongation by
470 RNA Pol II [37,58]. Interestingly, we observed that 74% of SEs are bound by
471 mH2A1.1 and BRD4 in MDA-MB231 cells (*data not shown*). Could the presence of
472 mH2A1.1 at SEs and TSSs be involved in the recruitment of BRD4 to enable
473 elongation of RNA Pol II on paused mH2A1.1-activated genes ?

474

475 Overall, we demonstrate for the first time that mH2A1.1 activates transcription
476 of paused genes by its selective recruitment to their TSSs. It remains to demonstrate
477 whether mH2A1.1 is directly involved in the recruitment of factors acting on Pol II
478 pausing events and/or on the establishment of 3D-chromatin structures (e.g
479 chromatin looping mechanisms), facilitating Pol II pausing release.

480

481 **Materials and Methods**

482

483 **Cell culture.** MDA-MB231, HEK-293T and MCF7 cell lines were purchased from
484 ATCC, and were maintained and amplified in Dulbecco's Modified Eagle's (DMEM)
485 for HEK-93T and MDA-MB231 cells, and in DMEM-F12 for MCF7 cells,
486 supplemented with gentamycin (50 µg/ml) (Gibco), fetal bovine serum (10%, Gibco)
487 and sodium pyruvate (100 mM, Sigma). Cells were maintained in a humidified
488 incubator at 37°C with 5% CO₂. Cells lines were regularly tested for mycoplasma
489 infection (MycoAlert, Lonza). In Montpellier, MDA-MB231 cells were cultured in
490 DMEM supplemented with 10% fetal calf serum and penicillin/streptomycin (100
491 µg/ml each) and regularly tested for mycoplasma infection.

492 **Transfection of siRNAs and plasmids.** At 30-50% cell confluence, transfection of
493 siRNA (11nM) was performed using INTERFERin (Polyplus-Ozyme) according to the
494 manufacturer's protocol. Cells in control condition were transfected with INTERFERin
495 without any siRNA Transfection of plasmid (1µg) was done with FuGene HD
496 (Promega) according to the manufacturer's protocol. siRNA and plasmid sequences
497 are available in S5 Table. Two and three days post plasmid and siRNA transfection
498 respectively, cells were recovered for experiments.

499 **Western blotting.** Cells were lysed and subjected to western blot analysis as
500 previously described [60]. Briefly, proteins extracts were separated in 10%
501 polyacrylamide (1:125 bisacrylamide:acrylamide) SDS gels, transferred onto
502 nitrocellulose membrane (Bio-Rad) and blocked with PBS-Tween 0.4% - Milk 5% for
503 1h at RT with rotation. Membranes were then incubated with primary antibodies
504 overnight (O/N) at 4°C in PBS-Tween 0.4% - Milk 5% with rotation (or 1h30 at RT).
505 Primary antibodies are described in the S1 Table Membranes were next incubated
506 with secondary antibody in PBS-Tween 0.4% - Milk 5% 1h at RT with rotation and the
507 signal was detected using chemiluminescence. Secondary antibodies are described
508 in the S1 Table. Signal quantifications were carried out with Image Lab software (Bio-
509 Rad).

510 **RNA extraction, reverse transcription and quantitative real-PCR (qRT-PCR).**
511 Total RNA was isolated using the RNAeasy midi kit (Qiagen). Purified RNA was
512 reversed transcribed to cDNA using Maxima H Minus first Strand cDNA synthesis kit
513 (Promega). The sequences of the primers used are available in S6 Table. RT-PCR
514 was performed using iTAq Universal SYBR Green (Bio-Rad) according to
515 manufacturer's instructions. At least two independent experiments were performed

516 for each condition. The relative expression levels of mRNA were normalized to
517 RPLP0 mRNA expression and evaluated according to the $2^{-\Delta\Delta Ct}$ method [61].

518 **Fluorescence microscopy.** Two or three days post-transfection, cells were fixed
519 with 4 % paraformaldehyde for 15 min for MDA-MB231 cells and 10 min for HEK-
520 293T at RT. Cells permeabilization was carried out using 0.1 % Triton X-100 in PBS
521 for 10 min at RT. Cells were then blocked with 5 % BSA-0.15% Tween in PBS for 1h
522 at RT. Next, cells were incubated with primary antibody O/N at 4°C. Cells were then
523 incubated with Alexa conjugated secondary antibody for 1h at RT. Actin was labelled
524 using cytoPainter Phalloidin iFluor diluted 1:1000 with secondary antibody according
525 to the manufacturer's protocol (Abcam; see S1 Table). Antibodies references and
526 dilutions are provided in S1 Table. The coverslips were finally incubated with Hoechst
527 (Invitrogen, 33342) for 30 min and then mounting with mounting media (Vectashield).
528 Images were acquired with Zeiss LSM 710 big confocal microscope using an x63 PL
529 APO oil DIC On 1.4 objective for all experiments. Images were taken in Z-stacks with
530 a voxel size of 300 nm. A Z-stack or Standard deviation intensity projections of Z-
531 stacks are shown.

532 **Chromatin immunoprecipitation and library preparation.** Cells were cross-linked
533 in DMEM containing 1.2 % of paraformaldehyde at RT for 10 min with rotation.
534 Cross-link was stopped by the addition of glycine to a final concentration of 0.125M
535 for 5 min. Cell were harvested and lysed in cell lysis buffer (10 mM Tris-HCl pH 7.4,
536 15 mM NaCl, 60 mM KCL, 1 mM EDTA, 0.1 mM EGTA, 0.2% NP-40, 5% sucrose).
537 After 10 min in ice, cell lysis was amplified with a 2mL dounce (Kimble Chase) to
538 enhance the nuclei separation from cytoplasm. Cell lysis buffer containing lysed cells
539 was deposit up to a pillow buffer (10 mM Tris-HCl pH 7.4, 15 mM NaCl, 60 mM KCL,
540 1 mM EDTA, 0.1 mM EGTA, 0.2% NP-40, 10% sucrose). Nuclei were then pelleted

541 by centrifugation and wash with washing buffer (10 mM Tris-HCl pH 7.4, 15 mM
542 NaCl, 60 mM KCL). Nuclei were then resuspended in sonication buffer (50 mM Tris-
543 HCl pH 7.5, 150 mM KCl, 5 mM EDTA, 1% NP-40, 0.1% SDS, 0.5 % Sodium
544 deoxycholate, Protease Inhibitor (Roche)). Chromatin was sheared using a Bioruptor
545 (Diagenode) (30 cycles, 30 sec ON/ 30 sec OFF) in order to obtain chromatin
546 fragments with an average size of 300-500 bp. Quality and size of chromatin
547 fragments was monitored by ethidium-bromide stained agarose gel electrophoresis
548 after DNA purification. Then, 100 µg of DNA was incubated with antibody O/N at 4°C
549 on a rotation wheel. Antibodies are described in the S1 Table. 3 mg of protein A
550 magnetic dynabeads (Sigma) were added for 3h at 4°C on a rotation wheel.
551 Immunoprecipitates were then exposed to serial washes for 5 min each on a rotation
552 wheel at 4°C in the following buffers (two times/buffer) : WB_I: 2 mM EDTA, 20 mM
553 Tris pH 8.1, 1 % Triton 100X, 150 mM NaCl, WB_{II}: 2 mM EDTA, 20 mM Tris pH 8.1, 1
554 % Triton X100, 500 mM NaCl, WB_{III}: 1 mM EDTA, 10 mM Tris pH 8.1, 250 mM LiCl,
555 1 % Sodium deoxycholate, 1 % NP-40 and WB_{IV}: 1 M EDTA, 10 mM Tris pH 8.1.
556 Chromatin was eluted from the magnetic beads with DNA isolation buffer (2% SDS,
557 0.1 M NaHCO₃) for 1h at 65°C under agitation. Extracts were reverse-crosslinked
558 with SDS O/N at 65°C. RNAs were degraded with RNase A and proteins were finally
559 degraded with proteinase K. Same procedure was performed for input (10 µg of
560 DNA). DNA was finally extracted with a phenol-chloroform extraction. Quantity and
561 quality of DNA was tested with a nanodrop (NanoDrop2000, Thermo). Samples were
562 sequenced with the GeT core facility, Toulouse, France (<http://get.genotoul.fr>).
563 Sequencing was done HiSeq3000-HWI-J00115 according to the manufacturer's
564 protocol. Same procedure was done for ChIPqPCR, expected that 20 µg of DNA was
565 used with 1µg of antibody). The sequences of the primers used are available in

566 Supplementary Table 7. For western blot analysis, extracts (Input (10% IP), No
567 immunoprecipitated (NoIP) fraction and IP fraction were processed as ChIP extract
568 but not incubated with the proteinase K and RNAase A. Extracts were then subjected
569 to western blot analysis as previously described in the western blot paragraph. To
570 compare different extracts, we loaded 2 % of Input, 0.5 % of Input, 0.5 % of NoIP
571 fraction and 20% of IP fraction. Percentages are relative to the DNA quantity used for
572 ChIP (100 µg).

573 ChIP-seq of H3K27ac, H3K4me1, H3K4me3, H3K36me3 and Pol II were done
574 essentially as previously described [62]. Briefly, after cell fixation with 1 % of
575 paraformaldehyde at RT for 5 min, cells were incubated in cell lysis buffer (PIPES 5
576 mM, KCL 85 mM, NP40 0.5%, Na Butyrate 10 mM, protease inhibitors) for 10 min on
577 ice. After mild centrifugation, nuclei were lysed in Nuclei Lysis Buffer (Tris-HCL 50
578 mM pH 7.5, SDS 0.125%, EDTA 10 mM, Na Butyrate 10 mM, protease inhibitors) at
579 4°C for 2h and, then, sonicated for 10 cycles at 4°C using BioruptorPico device from
580 Diagenode. For immunoprecipitation of H3K4me1, H3K4me3 and H3K27ac, 150 µl of
581 chromatin (equivalent to $4 \cdot 10^6$ cells) and 4.5 µg of the corresponding antibodies were
582 used. For Pol II, 850 µl of chromatin (equivalent to $22 \cdot 10^6$ cells) and 20 µg of the
583 corresponding antibody were used. Two independent replicates for each ChIP were
584 sequenced by the MGX genomic platform (Montpellier) using the Hi-seq2500 Illumina
585 sequencer.

586 **Strand-specific total RNA library preparation.** Total RNA was isolated using the
587 RNAeasy midi kit (Qiagen). RNA-seq quality and quantity control were performed
588 using a Nanodrop (NanoDrop2000, Thermo) and BioAnalyser. Library preparation
589 and sequencing was done by GeT core facility, Toulouse, France
590 (<http://get.genotoul.fr>) with the kit TruSeq Stranded total RNA according to

591 manufacturer's institutions. Sequencing was done HiSeq3000-HWI-J00115 according
592 to the manufacturer's protocol.

593 **ChIP-seq data processing.** The quality of the reads was estimated with FastQC
594 (Illumina, 1.0.0). Published ChIP-seq data of H3K9me3 (GSM2258862), H3K27me3
595 (GSM2258850) and corresponding input (GSM2258864) in MDA-MB231 cells were
596 downloaded from GEODATASETS (<https://www.ncbi.nlm.nih.gov/geo/>, GEO
597 accession number : GSE85158) [33], and reanalyzed as subsequently described.
598 Published ChIP-seq data of BRD4 in MDA-MB231 cells were downloaded
599 (GSM2862187) and corresponding input (GSM2862178) from GEODATASETS
600 (<https://www.ncbi.nlm.nih.gov/geo/>, GEO accession number : GSE107176) [63], and
601 reanalyzed as subsequently described. Sequenced reads were aligned to the human
602 genome Assembly GRCh38 using STAR (2.5.1) algorithm with defaults
603 parameters[64]. Details are supplied in Supplementary Table 2. Low quality reads
604 were then filtered out using Samtools (Samtools, options `-q 10 -view`) [65].
605 Conversion of BAM files to bigWig files was performed with Bamcompare tool
606 (DeepTools utilities v3.1.3) [66]. Corresponding ChIP-seq data generated from
607 genomic DNA (Input) were used as control for every bigWig files normalization
608 (options: `--normalizeUsing RPKM --operation subtract --binSize 50 bp --`
609 `smoothLength 150 bp`). Peaks were determined with the enrichR function of NormR
610 package [67] (v3.8, Options: `fdr = 5e-2, binsize = 500 bp` for H3K27ac, H3K4me1,
611 H3K4me3, BRD4 and `binsize =3000 bp` for H3K36me3, PolII, H3K9me3, H3K27me3,
612 Ab α H2A1.1 and Ab α H2A1). Peaks of Ab α H2A1.1 (and Ab α H2A1) spaced
613 less than 3000 bp on the linear genome were merged. All downstream analyses were
614 mainly performed with R studio. ChIP-seq signal and peaks positions visualization
615 were obtained with IGV [68,69]. Boxplots were done with ggplot2 [70]. Distributions of

616 mH2A1 isoforms and H3K27me3/H3K9me3 common peaks identified at specific
617 genomic features were calculated using ChIPseeker package with default
618 parameters (Figs 1B and S5B) [71]. Statistical analyses are presented in Statistics
619 and Reproducibility paragraph.

620 **Identification of “putative” enhancers and super-enhancers.** All putative
621 enhancers correspond to the union of H3K4me1 and H3K27ac outside TSS (+/- 2 kb)
622 to avoid TSS bias (Fig 2B) [72]. TSS annotation is based on
623 TxDb.Hsapiens.UCSC.hg38.knownGene release. Super-enhancers were determined
624 with ROSE utility tools based on H3K27ac signal (options : stitching_distance = 12.5
625 kb and TSS_exclusion_zone_size : 2500 bp) (Fig 2C and 2D) [36,37].

626 **Pol II pausing index calculation.** Pausing index (PI) was defined as previously[73],
627 which is the ratio of Pol II (total) density in the promoter-proximal region ([-30;300] bp
628 centered on TSS) to the total Pol II density in the transcribed regions (TSS + 300 bp
629 to TES). Pausing index was only calculated for actively expressed genes (n=10,198,
630 see RNA-seq analysis). Paused genes were defined as genes that have a PI upper
631 to 2 (n=6,821).

632 **Venn diagrams.** Intersection of peaks were determined with the function
633 findOverlaps() from GenomicRanges package. To note that for two ChIP-seq peaks
634 intersections, only number of overlaps is contabilised and not the number of each
635 peaks contained per overlap. This particularity explained why number of peaks
636 changes between venn diagrams for a same ChIP-seq (i.e Fig 1A Ab α mH2A1.1
637 (n=29,112) and Fig 1C α mH2A1.1 (n=19,867)). The area-proportional Venn diagrams
638 were drawn based on images generated by Vennable package. For Venn diagrams
639 in Fig 3, 5 and S6, intersections were performed at TSS (1-bp or +/- 1kb centered on

640 TSS). Enrichment tests associated to Venn diagrams are explained in Statistics and
641 Reproducibility paragraph.

642 **Correlation heatmaps.** Correlation heatmaps were done with multiBigwigSummary
643 (options: -bins for whole genome (S2A Fig) or -BED-file for Heterochromatin,
644 enhancer, TSS +/-1kp; TSS 1-bp, Figs S2B, S6A and S7B) and plotCorrelation
645 (option: -spearman correlation heatmap) from DeepTools utilities (3.1.3) [66].

646 **Metagenes profils.** Metagene analysis profiles were performed with R Seqplot
647 package using bigWig files (function getPlotSetArray and plotAverage) [74].
648 Heatmaps profiles of Pol II, Ab α H2A1.1 and Ab α H2A1 around TSS +/- 10 kb
649 were ranked by mH2A1.1 KD de-regulated genes (mH2A1.1-activated and repressed
650 genes, not ranked by the deregulation level) and done with Seqplot package
651 (function getPlotSetArray and plotHeatmap) (Fig 4B). Heatmaps profiles of Pol II, Ab
652 α H2A1.1 and Ab α H2A1 around TSS +/- 10 kb were ranked by pausing index and
653 carried out with Seqplot package (function getPlotSetArray and plotHeatmap)(Fig
654 5A).

655 **RNA-seq analysis.** The quality of the reads was estimated with FastQC (Illumina,
656 1.0.0). The reads were mapped to the human reference genome GRCh38 using the
657 default parameters of STAR (2.5.1) [64]. Details are supplied in S1 Table . Low
658 quality reads and duplicates were then filtered out using SAMtools (Samtools,
659 options -q 10 -view ; -rmdup)[65]. Unstranded normalized Bedgraph files in Read Per
660 Millions (RPM) were obtained with STAR using Output Wiggle parameters (options : -
661 -outWigTypebedgraph; --outWigStrandUnstranded; --outWigNorm RPM) [64]. Gene
662 counts were performed with htseq-count utilities with default parameters (0.8.0) [74].
663 FPKM for all genes were calculated with the formula : $FPKM = (RC_g \times 10^6) / (RC_p \times L)$
664 where RC_g corresponds to the number of reads mapped to the gene, RC_p to the

665 number of reads mapped to all protein-coding genes and L, the Length of the gene in
666 base pairs. Differential expression analysis was performed with DESeq2 package
667 [75] with cutoff $|FC| > 1.5$ and $padj < 0.1$. Corresponding volcano plot was done with
668 EnhancedVolcano package [76] (Fig 4A). The mH2A1.1 KD de-regulated genes are
669 listed in S3 Table.

670 **GO analysis.** GO analysis was performed with LIMMA package (--function goana)
671 (3.8)[77] and corresponding GO terms are supplied in the S4 Table. Selection of
672 genes related to their functions was done with biomaRt package (function getBM())
673 [78,79]. We took genes related to: cytoskeleton (GO:0005856), cell adhesion
674 (GO:0007155), Cilium (GO:0005929) and cell junction (GO:0030054) (n=6,821).

675 **Transwell migration assay.** Transwell migration assays were performed using
676 Transwell plates with 0.8 μm pore polycarbonate membranes (CorningTranswell,
677 Sigma). Three days post siRNA transfection, MDA-MB231 cells (10^5) were seeded in
678 the upper chamber without FBS and allowed to invade to the reverse side of the
679 chamber under chemoattractant condition with 10% FBS medium in the lower
680 chamber. Following incubation for 16h at 37°C, the cells were fixed with 3.7%
681 formaldehyde for 2 min at RT. Cells permeabilization was carried out with methanol
682 incubation for 20 min at RT. Cells were then stained with Giesma for 15 min at RT.
683 Same final total cell number between conditions was always checked by wide field
684 microscope to avoid proliferation bias during number of migratory cell comparison.
685 Not migrated cells were finally removed from the upper chamber by using a cotton
686 swab. Migrated cells adhering to the underside of the chamber were photographed
687 using a light microscope at x200 magnification (Invitrogen EVOS Digital Color
688 Fluorescence Microscope). Cell counting was done with ImageJ in ten different fields

689 per condition[80]. Three independent experiments were performed for each
690 condition.

691 **Statistics and reproducibility.** All western blot, RTqPCR and Boyden Chamber
692 assay experiments were repeated at least twice as independent biological replicates
693 and results are presented as mean +/- sd. All statistical analyses were done with R.
694 For Western blot, RTqPCR and Boyden Chamber, Wilcoxon tests were used to
695 compare mean values between conditions. p-values were considered as significant
696 when * ≤ 0.05 and highly significant when ** ≤ 0.01 ; *** ≤ 0.001 ; **** ≤ 0.0001 . All
697 enrichment tests were performed on a base set made up of genomic bins. Each
698 genomic bin is defined by merging ranges of ChIP-seq data (reduce function
699 GenomicRanges R package) ChIP-seq list : Ab α H2A1.1, Ab α H2A1, Pol II,
700 H3K4me3, H3K4me1, H3K36me3, H3K27ac, H3K27me3, H3K9me3, “putative”
701 enhancers and “putative” Super-enhancers, on whole genome or TSS +/- 1kb and
702 TSS 1-bp.

703

704 **Data availability.** ChIP-seq and RNA-seq data have been deposited to GEO under
705 accession number GSE140022. Additional data are available upon reasonable
706 request.

707

708 **Acknowledgements**

709 We thank M. Buschbeck from JCLR Institute at Barcelona for kindly providing Flag-
710 mH2A1.1 and Flag-mH2A1.2 expression plasmids. ChIP-seq data against mH2A1.1
711 and mH2A1 isoforms as well as RNA-seq data were performed in collaboration with
712 the GeT core facility, Toulouse, France (<http://get.genotoul.fr>), and were supported
713 by France Génomique National infrastructure, funded as part of “Investissement

714 d’avenir” program managed by Agence Nationale pour la Recherche (contract ANR-
715 10-INBS-09) and by the Fondation Recherche Medical (DEQ43940 to O.C team
716 including A.H). We thank Marc Piechaszky for critical reading of the manuscript. We
717 acknowledge support from the light imaging Toulouse CBI platform. The work was
718 generously funded by the Institut National du Cancer (INCA PL-BIO-16-269) to KB.

719

720 **Author contributions**

721 L.R, A-C.L and K.B conceived this study. F.M validated custom Ab α mH2A1.1
722 antibody specificity against mH2A1.1. L.R performed ChIP-seq against mH2A1
723 isoforms and RNA-seq. I.E-J. and F.B performed ChIP-seq against active histone
724 marks. L.R, A.H and F.R realized bioinformatic analysis of all ChIP-seq and RNA-seq
725 data. Statistical analyses were done by A.H and L.R. L.R performed all other
726 experimental data. L.R, A-C.L and K.B designed experiments and interpreted results.
727 L.R, A-C.L and K.B wrote the manuscript with input from all other authors.

728

729 **Competing interests**

730

731 The authors declare no competing interests.

732

733 **References**

- 734 1. Venkatesh S, Workman JL. Histone exchange, chromatin structure and the
735 regulation of transcription. Nat Rev Mol Cell Biol [Internet]. 2015 Mar 4 [cited
736 2019 Nov 6];16(3):178–89. Available from:
737 <http://www.ncbi.nlm.nih.gov/pubmed/25650798>
- 738 2. Luger K, Dechassa ML, Tremethick DJ. New insights into nucleosome and

- 739 chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell*
740 *Biol* [Internet]. 2012 Jul 22 [cited 2019 Jul 23];13(7):436–47. Available from:
741 <http://www.ncbi.nlm.nih.gov/pubmed/22722606>
- 742 3. Buschbeck M, Hake SB. Variants of core histones and their roles in cell fate
743 decisions, development and cancer. *Nat Rev Mol Cell Biol* [Internet]. 2017 Feb
744 1 [cited 2017 May 10];18(5):299–314. Available from:
745 <http://www.ncbi.nlm.nih.gov/pubmed/28144029>
- 746 4. Pehrson JR, Fuji RN. Evolutionary conservation of histone macroH2A subtypes
747 and domains. *Nucleic Acids Res* [Internet]. 1998 Jun 15 [cited 2019 Jul
748 23];26(12):2837–42. Available from: [https://academic.oup.com/nar/article-](https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/26.12.2837)
749 [lookup/doi/10.1093/nar/26.12.2837](https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/26.12.2837)
- 750 5. Rivera-Casas C, Gonzalez-Romero R, Cheema MS, Ausió J, Eirín-López JM.
751 The characterization of macroH2A beyond vertebrates supports an ancestral
752 origin and conserved role for histone variants in chromatin. *Epigenetics*
753 [Internet]. 2016 Jun 2 [cited 2019 Jul 23];11(6):415–25. Available from:
754 <http://www.tandfonline.com/doi/full/10.1080/15592294.2016.1172161>
- 755 6. Gamble MJ, Kraus WL. Multiple facets of the unique histone variant
756 macroH2A: From genomics to cell biology. *Cell Cycle*. 2010;9(13):2568–74.
- 757 7. Costanzi C, Pehrson JR. Histone macroH2A1 is concentrated in the inactive X
758 chromosome of female mammals. *Nature*. 1998;628(1997):1997–9.
- 759 8. Gamble MJ, Frizzell KM, Yang C, Krishnakumar R, Kraus WL. The histone
760 variant macroH2A1 marks repressed autosomal chromatin, but protects a
761 subset of its target genes from silencing. *Genes Dev* [Internet]. 2010 Jan 1
762 [cited 2019 Jul 23];24(1):21–32. Available from:
763 <http://genesdev.cshlp.org/cgi/doi/10.1101/gad.1876110>

- 764 9. Douet J, Corujo D, Malinverni R, Renauld J, Sansoni V, Posavec Marjanović
765 M, et al. MacroH2A histone variants maintain nuclear organization and
766 heterochromatin architecture. *J Cell Sci.* 2017;130(9):1570–82.
- 767 10. Sun Z, Filipescu D, Andrade J, Gaspar-Maia A, Ueberheide B, Bernstein E.
768 Transcription-associated histone pruning demarcates macroH2A chromatin
769 domains. *Nat Struct Mol Biol [Internet].* 2018 Oct 5 [cited 2019 Jul
770 23];25(10):958–70. Available from: [http://www.nature.com/articles/s41594-018-](http://www.nature.com/articles/s41594-018-0134-5)
771 0134-5
- 772 11. Doyen C-M, An W, Angelov D, Bondarenko V, Mietton F, Studitsky VM, et al.
773 Mechanism of polymerase II transcription repression by the histone variant
774 macroH2A. *Mol Cell Biol [Internet].* 2006 Feb 1 [cited 2019 Jul 23];26(3):1156–
775 64. Available from: [http://mcb.asm.org/cgi/doi/10.1128/MCB.26.3.1156-](http://mcb.asm.org/cgi/doi/10.1128/MCB.26.3.1156-1164.2006)
776 1164.2006
- 777 12. Angelov D, Molla A, Perche P-Y, Hans F, Côté J, Khochbin S, et al. The
778 histone variant macroH2A interferes with transcription factor binding and
779 SWI/SNF nucleosome remodeling. *Mol Cell [Internet].* 2003 Apr [cited 2019 Jul
780 23];11(4):1033–41. Available from:
781 <http://www.ncbi.nlm.nih.gov/pubmed/12718888>
- 782 13. Dell'Orso S, Wang AH, Shih H-Y, Saso K, Berghella L, Gutierrez-Cruz G, et al.
783 The Histone Variant MacroH2A1.2 Is Necessary for the Activation of Muscle
784 Enhancers and Recruitment of the Transcription Factor Pbx1. *Cell Rep*
785 [Internet]. 2016 Feb 9 [cited 2019 Jul 23];14(5):1156–68. Available from:
786 <https://linkinghub.elsevier.com/retrieve/pii/S2211124716000036>
- 787 14. Chen H, Ruiz PD, Novikov L, Casill AD, Park JW, Gamble MJ. MacroH2A1.1
788 and PARP-1 cooperate to regulate transcription by promoting CBP-mediated

- 789 H2B acetylation. *Nat Struct Mol Biol* [Internet]. 2014 Oct 12 [cited 2016 Nov
790 19];21(11):981–9. Available from:
791 <http://www.nature.com/doi/10.1038/nsmb.2903>
- 792 15. Podrini C, Koffas A, Chokshi S, Vinciguerra M, Lelliott CJ, White JK, et al.
793 MacroH2A1 isoforms are associated with epigenetic markers for activation of
794 lipogenic genes in fat-induced steatosis. *FASEB J* [Internet]. 2015 May [cited
795 2019 Jul 23];29(5):1676–87. Available from:
796 <http://www.fasebj.org/doi/10.1096/fj.14-262717>
- 797 16. Wan D, Liu C, Sun Y, Wang W, Huang K, Zheng L. MacroH2A1.1 cooperates
798 with EZH2 to promote adipogenesis by regulating Wnt signaling. *J Mol Cell Biol*
799 [Internet]. 2017 Aug 1 [cited 2019 Jul 23];9(4):325–37. Available from:
800 [http://academic.oup.com/jmcb/article/9/4/325/4067698/MacroH2A11-](http://academic.oup.com/jmcb/article/9/4/325/4067698/MacroH2A11-cooperates-with-EZH2-to-promote)
801 [cooperates-with-EZH2-to-promote](http://academic.oup.com/jmcb/article/9/4/325/4067698/MacroH2A11-cooperates-with-EZH2-to-promote)
- 802 17. Changolkar LN, Singh G, Cui K, Berletch JB, Zhao K, Distèche CM, et al.
803 Genome-wide distribution of macroH2A1 histone variants in mouse liver
804 chromatin. *Mol Cell Biol* [Internet]. 2010 Dec 1 [cited 2019 Jul 23];30(23):5473–
805 83. Available from: <http://mcb.asm.org/cgi/doi/10.1128/MCB.00518-10>
- 806 18. Posavec Marjanović M, Hurtado-Bagès S, Lassi M, Valero V, Malinverni R,
807 Delage H, et al. MacroH2A1.1 regulates mitochondrial respiration by limiting
808 nuclear NAD⁺ consumption. *Nat Struct Mol Biol* [Internet]. 2017 Nov 9 [cited
809 2019 Aug 7];24(11):902–10. Available from:
810 <http://www.nature.com/articles/nsmb.3481>
- 811 19. Sporn JC, Kustatscher G, Hothorn T, Collado M, Serrano M, Muley T, et al.
812 Histone macroH2A isoforms predict the risk of lung cancer recurrence.
813 *Oncogene* [Internet]. 2009 Sep 3 [cited 2019 Jul 23];28(38):3423–8. Available

- 814 from: <http://www.ncbi.nlm.nih.gov/pubmed/19648962>
- 815 20. Cantariño N, Douet J, Buschbeck M. MacroH2A – An epigenetic regulator of
816 cancer. *Cancer Lett* [Internet]. 2013 Aug [cited 2019 Jul 23];336(2):247–52.
817 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0304383513002504>
- 818 21. Sporn JC, Jung B. Differential regulation and predictive potential of MacroH2A1
819 isoforms in colon cancer. *Am J Pathol* [Internet]. 2012 Jun [cited 2019 Jul
820 23];180(6):2516–26. Available from:
821 <http://www.ncbi.nlm.nih.gov/pubmed/22542848>
- 822 22. Mermoud JE, Costanzi C, Pehrson JR, Brockdorff N. Histone MacroH2A1 . 2
823 Relocates to the Inactive X Chromosome after Initiation and Propagation of X-
824 Inactivation. *J Cell Biol*. 1999;147(7):1399–408.
- 825 23. Pliatska M, Kapasa M, Kokkalis A, Polyzos A, Thanos D. The Histone Variant
826 MacroH2A Blocks Cellular Reprogramming by Inhibiting Mesenchymal-to-
827 Epithelial Transition. *Mol Cell Biol* [Internet]. 2018 Feb 15 [cited 2019 Jul
828 23];38(10):e00669-17. Available from:
829 <http://mcb.asm.org/lookup/doi/10.1128/MCB.00669-17>
- 830 24. Lavigne A-C, Castells M, Mermet J, Kocanova S, Dalvai M, Bystricky K.
831 Increased macroH2A1.1 expression correlates with poor survival of triple-
832 negative breast cancer patients. Haibe-Kains B, editor. *PLoS One* [Internet].
833 2014 Jun 9 [cited 2019 Jul 23];9(6):e98930. Available from:
834 <https://dx.plos.org/10.1371/journal.pone.0098930>
- 835 25. Kustatscher G, Hothorn M, Pugieux C, Scheffzek K, Ladurner AG. Splicing
836 regulates NAD metabolite binding to histone macroH2A. *Nat Struct Mol Biol*
837 [Internet]. 2005 Jul 19 [cited 2019 Jul 23];12(7):624–5. Available from:
838 <http://www.ncbi.nlm.nih.gov/pubmed/15965484>

- 839 26. Ray Chaudhuri A, Nussenzweig A. The multifaceted roles of PARP1 in DNA
840 repair and chromatin remodelling. *Nat Rev Mol Cell Biol* [Internet]. 2017 Oct 5
841 [cited 2019 Jul 23];18(10):610–21. Available from:
842 <http://www.ncbi.nlm.nih.gov/pubmed/28676700>
- 843 27. Ouararhni K, Hadj-Slimane R, Ait-Si-Ali S, Robin P, Mietton F, Harel-Bellan A,
844 et al. The histone variant mH2A1.1 interferes with transcription by down-
845 regulating PARP-1 enzymatic activity. *Genes Dev* [Internet]. 2006 Dec 1 [cited
846 2019 Jul 23];20(23):3324–36. Available from:
847 <http://www.genesdev.org/cgi/doi/10.1101/gad.396106>
- 848 28. Marjanović MP, Hurtado-bagès S, Lassi M, Valero V, Malinverni R, Delage H,
849 et al. MacroH2A1 . 1 regulates mitochondrial respiration by limiting nuclear
850 NAD + consumption. 2018;24(11):902–10.
- 851 29. Xu C, Xu Y, Gursoy-Yuzugullu O, Price BD. The histone variant macroH2A1.1
852 is recruited to DSBs through a mechanism involving PARP1. *FEBS Lett*
853 [Internet]. 2012 Nov 2 [cited 2019 Jul 23];586(21):3920–5. Available from:
854 <http://doi.wiley.com/10.1016/j.febslet.2012.09.030>
- 855 30. Kim J, Oberdoerffer P, Khurana S. The histone variant macroH2A1 is a
856 splicing-modulated caretaker of genome integrity and tumor growth. *Mol Cell*
857 *Oncol* [Internet]. 2018 Mar 7 [cited 2019 Jul 23];5(3):e1441629. Available from:
858 <https://www.tandfonline.com/doi/full/10.1080/23723556.2018.1441629>
- 859 31. Chen H, Ruiz PD, McKimpson WM, Novikov L, Kitsis RN, Gamble MJ.
860 MacroH2A1 and ATM Play Opposing Roles in Paracrine Senescence and the
861 Senescence-Associated Secretory Phenotype. *Mol Cell* [Internet]. 2015 Sep
862 [cited 2019 Dec 11];59(5):719–31. Available from:
863 <https://linkinghub.elsevier.com/retrieve/pii/S1097276515005699>

- 864 32. Ouararhni K, Hadj-Slimane R, Ait-Si-Ali S, Robin P, Mietton F, Harel-Bellan A,
865 et al. The histone variant mH2A1.1 interferes with transcription by down-
866 regulating PARP-1 enzymatic activity. *Genes Dev* [Internet]. 2006 Dec 1 [cited
867 2019 Dec 11];20(23):3324–36. Available from:
868 <http://www.genesdev.org/cgi/doi/10.1101/gad.396106>
- 869 33. Franco HL, Nagari A, Malladi VS, Li W, Xi Y, Richardson D, et al. Enhancer
870 transcription reveals subtype-specific gene expression programs controlling
871 breast cancer pathogenesis. *Genome Res* [Internet]. 2018 Feb [cited 2019 Jul
872 23];28(2):159–70. Available from:
873 <http://genome.cshlp.org/lookup/doi/10.1101/gr.226019.117>
- 874 34. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and
875 function. *Nat Rev Mol Cell Biol* [Internet]. 2018 Apr 13 [cited 2019 Jul
876 26];19(4):229–44. Available from: <http://www.nature.com/articles/nrm.2017.119>
- 877 35. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et
878 al. Histone H3K27ac separates active from poised enhancers and predicts
879 developmental state. *Proc Natl Acad Sci* [Internet]. 2010 Dec 14 [cited 2019 Jul
880 23];107(50):21931–6. Available from:
881 <http://www.pnas.org/cgi/doi/10.1073/pnas.1016071107>
- 882 36. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al.
883 Master Transcription Factors and Mediator Establish Super-Enhancers at Key
884 Cell Identity Genes. *Cell* [Internet]. 2013 Apr [cited 2019 Jul 23];153(2):307–19.
885 Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867413003929>
- 886 37. Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective
887 inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* [Internet].
888 2013 Apr 11 [cited 2019 Jul 23];153(2):320–34. Available from:

- 889 <https://linkinghub.elsevier.com/retrieve/pii/S0092867413003930>
- 890 38. Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II:
891 emerging roles in metazoans. *Nat Rev Genet* [Internet]. 2012 Oct 18 [cited
892 2019 Jul 23];13(10):720–31. Available from:
893 <http://www.ncbi.nlm.nih.gov/pubmed/22986266>
- 894 39. Novikov L, Park JW, Chen H, Klerman H, Jalloh AS, Gamble MJ. QKI-
895 mediated alternative splicing of the histone variant MacroH2A1 regulates
896 cancer cell proliferation. *Mol Cell Biol* [Internet]. 2011 Oct 15 [cited 2019 Jul
897 23];31(20):4244–55. Available from:
898 <http://mcb.asm.org/cgi/doi/10.1128/MCB.05244-11>
- 899 40. Draheim KM, Chen H-B, Tao Q, Moore N, Roche M, Lyle S. ARRDC3
900 suppresses breast cancer progression by negatively regulating integrin beta4.
901 *Oncogene* [Internet]. 2010 Sep 9 [cited 2019 Jul 23];29(36):5032–47. Available
902 from: <http://www.nature.com/articles/onc2010250>
- 903 41. Mei Z, Chen S, Chen C, Xiao B, Li F, Wang Y, et al. Interleukin-23 Facilitates
904 Thyroid Cancer Cell Migration and Invasion by Inhibiting SOCS4 Expression
905 via MicroRNA-25. Zhang Z, editor. *PLoS One* [Internet]. 2015 Oct 5 [cited 2019
906 Jul 23];10(10):e0139456. Available from:
907 <https://dx.plos.org/10.1371/journal.pone.0139456>
- 908 42. Castillo-Lluva S, Tan C-T, Daugaard M, Sorensen PHB, Malliri A. The tumour
909 suppressor HACE1 controls cell migration by regulating Rac1 degradation.
910 *Oncogene* [Internet]. 2013 Mar 28 [cited 2019 Jul 23];32(13):1735–42.
911 Available from: <http://www.nature.com/articles/onc2012189>
- 912 43. Stankiewicz E, Mao X, Mangham DC, Xu L, Yeste-Velasco M, Fisher G, et al.
913 Identification of FBXL4 as a Metastasis Associated Gene in Prostate Cancer.

- 914 Sci Rep [Internet]. 2017 Dec 11 [cited 2019 Jul 23];7(1):5124. Available from:
915 <http://www.nature.com/articles/s41598-017-05209-z>
- 916 44. Pinzaglia M, Montaldo C, Polinari D, Simone M, La Teana A, Tripodi M, et al.
917 EIF6 over-expression increases the motility and invasiveness of cancer cells by
918 modulating the expression of a critical subset of membrane-bound proteins.
919 BMC Cancer [Internet]. 2015 Mar 15 [cited 2019 Jul 23];15(1):131. Available
920 from: <http://bmccancer.biomedcentral.com/articles/10.1186/s12885-015-1106-3>
- 921 45. Ryu H-H, Jung S, Jung T-Y, Moon K-S, Kim I-Y, Jeong Y-I, et al. Role of
922 metallothionein 1E in the migration and invasion of human glioma cell lines. Int
923 J Oncol [Internet]. 2012 Oct [cited 2019 Jul 23];41(4):1305–13. Available from:
924 <https://www.spandidos-publications.com/10.3892/ijo.2012.1570>
- 925 46. Selvaraj N, Budka JA, Ferris MW, Plotnik JP, Hollenhorst PC. Extracellular
926 signal-regulated kinase signaling regulates the opposing roles of JUN family
927 transcription factors at ETS/AP-1 sites and in cell migration. Mol Cell Biol
928 [Internet]. 2015 Jan 1 [cited 2019 Jul 23];35(1):88–100. Available from:
929 <http://mcb.asm.org/lookup/doi/10.1128/MCB.00982-14>
- 930 47. Kake S, Usui T, Ohama T, Yamawaki H, Sato K. Death-associated protein
931 kinase 3 controls the tumor progression of A549 cells through ERK MAPK/c-
932 Myc signaling. Oncol Rep [Internet]. 2017 Feb [cited 2019 Jul 23];37(2):1100–
933 6. Available from: [https://www.spandidos-](https://www.spandidos-publications.com/10.3892/or.2017.5359)
934 [publications.com/10.3892/or.2017.5359](https://www.spandidos-publications.com/10.3892/or.2017.5359)
- 935 48. Lavigne MD, Vatsellas G, Polyzos A, Mantouvalou E, Sianidis G, Maraziotis I,
936 et al. Composite macroH2A/NRF-1 Nucleosomes Suppress Noise and
937 Generate Robustness in Gene Expression. Cell Rep [Internet]. 2015 May [cited
938 2019 Jul 23];11(7):1090–101. Available from:

- 939 <https://linkinghub.elsevier.com/retrieve/pii/S221112471500409X>
- 940 49. Segal T, Salmon-Divon M, Gerlitz G. The Heterochromatin Landscape in
941 Migrating Cells and the Importance of H3K27me3 for Associated
942 Transcriptome Alterations. *Cells* [Internet]. 2018 Nov 9 [cited 2019 Jul
943 23];7(11):205. Available from: <http://www.mdpi.com/2073-4409/7/11/205>
- 944 50. Yokoyama Y, Hieda M, Nishioka Y, Matsumoto A, Higashi S, Kimura H, et al.
945 Cancer-associated upregulation of histone H3 lysine 9 trimethylation promotes
946 cell motility in vitro and drives tumor formation in vivo. *Cancer Sci* [Internet].
947 2013 Jul [cited 2019 Aug 7];104(7):889–95. Available from:
948 <http://doi.wiley.com/10.1111/cas.12166>
- 949 51. Bernstein E, Muratore-Schroeder TL, Diaz RL, Chow JC, Changolkar LN,
950 Shabanowitz J, et al. A phosphorylated subpopulation of the histone variant
951 macroH2A1 is excluded from the inactive X chromosome and enriched during
952 mitosis. *Proc Natl Acad Sci U S A* [Internet]. 2008 Feb 5 [cited 2019 Jul
953 23];105(5):1533–8. Available from:
954 <http://www.pnas.org/cgi/doi/10.1073/pnas.0711632105>
- 955 52. Dardenne E, Pierredon S, Driouch K, Gratadou L, Lacroix-Triki M, Espinoza
956 MP, et al. Splicing switch of an epigenetic regulator by RNA helicases
957 promotes tumor-cell invasiveness. *Nat Struct Mol Biol* [Internet]. 2012 Nov 30
958 [cited 2019 Jul 23];19(11):1139–46. Available from:
959 <http://www.nature.com/articles/nsmb.2390>
- 960 53. Kapoor A, Goldberg MS, Cumberland LK, Ratnakumar K, Segura MF, Emanuel
961 PO, et al. The histone variant macroH2A suppresses melanoma progression
962 through regulation of CDK8. 2011;468(7327):1105–9.
- 963 54. Borghesan M, Fusilli C, Rappa F, Panebianco C, Rizzo G, Oben JA, et al. DNA

- 964 Hypomethylation and Histone Variant macroH2A1 Synergistically Attenuate
965 Chemotherapy-Induced Senescence to Promote Hepatocellular Carcinoma
966 Progression. 2017;76(3):594–606.
- 967 55. Li F, Yi P, Pi J, Li L, Hui J, Wang F, et al. QKI5-mediated alternative splicing of
968 the histone variant macroH2A1 regulates gastric carcinogenesis. *Oncotarget*
969 [Internet]. 2016 May 31 [cited 2019 Jul 23];7(22):32821–34. Available from:
970 <http://www.oncotarget.com/fulltext/8739>
- 971 56. Gibson BA, Zhang Y, Jiang H, Hussey KM, Shrimp JH, Lin H, et al. Chemical
972 genetic discovery of PARP targets reveals a role for PARP-1 in transcription
973 elongation. *Science (80-)* [Internet]. 2016 Jul 1 [cited 2019 Aug
974 7];353(6294):45–50. Available from:
975 <http://www.ncbi.nlm.nih.gov/pubmed/27256882>
- 976 57. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, et al. Super-
977 Enhancers in the Control of Cell Identity and Disease. *Cell* [Internet]. 2013 Nov
978 7 [cited 2019 Aug 6];155(4):934–47. Available from:
979 <http://www.ncbi.nlm.nih.gov/pubmed/24119843>
- 980 58. Donati B, Lorenzini E, Ciarrocchi A. BRD4 and Cancer: going beyond
981 transcriptional regulation. *Mol Cancer* [Internet]. 2018 Dec 22 [cited 2019 Aug
982 6];17(1):164. Available from: [https://molecular-](https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-018-0915-9)
983 [cancer.biomedcentral.com/articles/10.1186/s12943-018-0915-9](https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-018-0915-9)
- 984 59. He Y, Long W, Liu Q. Targeting Super-Enhancers as a Therapeutic Strategy
985 for Cancer Treatment. *Front Pharmacol* [Internet]. 2019 Apr 11 [cited 2019 Aug
986 6];10:361. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31105558>
- 987 60. Yang P-C, Liu Z-Q, Mahmood T. Western blot: Technique, theory and trouble
988 shooting. *N Am J Med Sci* [Internet]. 2014 [cited 2019 Jul 23];6(3):160.

- 989 Available from: <http://www.najms.org/text.asp?2014/6/3/160/128482>
- 990 61. Rao X, Huang X, Zhou Z, Lin X. An improvement of the 2^{-delta delta CT}
- 991 method for quantitative real-time polymerase chain reaction data analysis.
- 992 *Biostat Bioinforma Biomath* [Internet]. 2013 Aug [cited 2019 Jul 23];3(3):71–85.
- 993 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25558171>
- 994 62. Tolza C, Bejjani F, Evanno E, Mahfoud S, Moquet-Torcy G, Gostan T, et al.
- 995 AP-1 Signaling by Fra-1 Directly Regulates HMGA1 Oncogene Transcription in
- 996 Triple-Negative Breast Cancers. *Mol Cancer Res* [Internet]. 2019 Oct [cited
- 997 2019 Oct 29];17(10):1999–2014. Available from:
- 998 <http://mcr.aacrjournals.org/lookup/doi/10.1158/1541-7786.MCR-19-0036>
- 999 63. Chan HL, Beckedorff F, Zhang Y, Garcia-Huidobro J, Jiang H, Colaprico A, et
- 1000 al. Polycomb complexes associate with enhancers and promote oncogenic
- 1001 transcriptional programs in cancer through multiple mechanisms. *Nat Commun*
- 1002 [Internet]. 2018 [cited 2019 Aug 7];9(1):3377. Available from:
- 1003 <http://www.ncbi.nlm.nih.gov/pubmed/30139998>
- 1004 64. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR:
- 1005 ultrafast universal RNA-seq aligner. *Bioinformatics* [Internet]. 2013 Jan 1 [cited
- 1006 2019 Jul 23];29(1):15–21. Available from:
- 1007 [https://academic.oup.com/bioinformatics/article-](https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts635)
- 1008 [lookup/doi/10.1093/bioinformatics/bts635](https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bts635)
- 1009 65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The
- 1010 Sequence Alignment/Map format and SAMtools. *Bioinformatics* [Internet]. 2009
- 1011 Aug 15 [cited 2019 Jul 23];25(16):2078–9. Available from:
- 1012 <http://www.ncbi.nlm.nih.gov/pubmed/19505943>
- 1013 66. Ram F, Ryan DP, Bhardwaj V, Kilpert F, Richter AS, Heyne S, et al.

- 1014 deepTools2 : a next generation web server for deep-sequencing data analysis.
1015 2016;44(April):160–5.
- 1016 67. Helmuth J CH. normr: Normalization and difference calling in ChIP-seq data . R
1017 Packag version 180, <https://github.com/your-highness/normR>. 2018;2018.
- 1018 68. James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman,
1019 Eric S. Lander, Gad Getz JPM. Integrative genomics viewer. Nat Biotechnol.
1020 2000;29(220):23–4.
- 1021 69. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer
1022 (IGV): high-performance genomics data visualization and exploration. Brief
1023 Bioinform [Internet]. 2013 Mar 1 [cited 2019 Jul 23];14(2):178–92. Available
1024 from: <https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbs017>
- 1025 70. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag
1026 New York [Internet]. 2016;174(1):245–6. Available from:
1027 http://doi.wiley.com/10.1111/j.1467-985X.2010.00676_9.x
- 1028 71. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP
1029 peak annotation, comparison and visualization. Bioinformatics [Internet]. 2015
1030 Jul 15 [cited 2019 Jul 23];31(14):2382–3. Available from:
1031 [https://academic.oup.com/bioinformatics/article-](https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv145)
1032 [lookup/doi/10.1093/bioinformatics/btv145](https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv145)
- 1033 72. Blinka S, Reimer MH, Pulakanti K, Pinello L, Yuan G-C, Rao S. Identification of
1034 Transcribed Enhancers by Genome-Wide Chromatin Immunoprecipitation
1035 Sequencing. Methods Mol Biol [Internet]. 2017 [cited 2019 Jul 23];1468:91–
1036 109. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27662872>
- 1037 73. Zhang X, Chiang H-C, Wang Y, Zhang C, Smith S, Zhao X, et al. Attenuation of
1038 RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation

- 1039 and tumorigenesis. *Nat Commun* [Internet]. 2017 Aug 26 [cited 2019 Jul
1040 23];8(1):15908. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28649985>
- 1041 74. Stempor P, Ahringer J. SeqPlots - Interactive software for exploratory data
1042 analyses, pattern discovery and visualization in genomics. *Wellcome Open*
1043 *Res* [Internet]. 2016;1(0):14. Available from:
1044 <https://wellcomeopenresearch.org/articles/1-14/v1>
- 1045 75. Love MI, Huber W, Anders S. Moderated estimation of fold change and
1046 dispersion for RNA-seq data with DESeq2. *Genome Biol* [Internet]. 2014 Dec 5
1047 [cited 2019 Jul 23];15(12):550. Available from:
1048 <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8>
- 1049 76. K B. EnhancedVolcano: Publication-ready volcano plots with enhanced
1050 colouring and labeling . <https://github.com/kevinblighe/EnhancedVolcano>
1051 [Internet]. 2018;R package:2018. Available from:
1052 <https://github.com/kevinblighe/EnhancedVolcano>.
- 1053 77. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers
1054 differential expression analyses for RNA-sequencing and microarray studies.
1055 *Nucleic Acids Res*. 2015;43(7):e47.
- 1056 78. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al.
1057 BioMart and Bioconductor: a powerful link between biological databases and
1058 microarray data analysis. *Bioinformatics* [Internet]. 2005 Aug 15 [cited 2019 Jul
1059 23];21(16):3439–40. Available from:
1060 [https://academic.oup.com/bioinformatics/article-](https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bti525)
1061 [lookup/doi/10.1093/bioinformatics/bti525](https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/bti525)
- 1062 79. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the
1063 integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat*

1064 Protoc [Internet]. 2009 Aug 23 [cited 2019 Jul 23];4(8):1184–91. Available
1065 from: <http://www.nature.com/articles/nprot.2009.97>
1066 80. Caroline A Schneider WSR& KWE. NIH Image to ImageJ: 25 years of image
1067 analysis. INMATEH - Agric Eng [Internet]. 2012;9(7):671–5. Available from:
1068 <http://dx.doi.org/10.1038/nmeth.2089>

1069

1070 **Figure legends**

1071

1072 **Fig 1. Histone mH2A1 isoforms are incorporated to “facultative-like”**
1073 **heterochromatin.** (A) Overlap of “mH2A1.1” (Ab α mH2A1.1) peaks with “mH2A1”
1074 (Ab α mH2A1). Peaks detected exclusively by Ab α mH2A1.1 are called “mH2A1.1
1075 only”, peaks detected exclusively by Ab α mH2A1 are called “mH2A1.2 only” and
1076 peaks detected by both correspond to mH2A1.1 only or mH2A1.1 and mH2A1.2 in
1077 unknown proportions, called “mH2A1s”. (B) Proportions of mH2A1 isoforms peaks at
1078 different genomic features (Methods). Arrows highlights differences between the two
1079 ChIP-seq. Promoters are defined around TSS. (C) Overlap of heterochromatin
1080 histone marks (H3K27me3 and H3K9me3) with mH2A1.1 peaks (left) and mH2A1
1081 peaks (right). Percentages of overlaps are also given. Results of fisher exact tests
1082 are shown (p-value (p) and Odd ratio) (Methods). (D) Genome browser view of ChIP-
1083 seq profiles illustrating occupancy of mH2A1 isoforms with heterochromatin histone
1084 marks (H3K27me3 and H3K9me3). At the top, region with high level of H3K27me3
1085 and at the bottom, region with high level of H3K9me3. Peaks and unstranded RNA-
1086 seq signal are shown in parallel. The black arrows show the direction of transcription.
1087 (E) Heatmap showing enrichment of mH2A1 isoforms sites (see Fig 1A) with
1088 H3K27me3 ; from left to right, 5 equal size categories as a function of differences in

1089 H3K27me3 and H3K9me3 signals. Stars indicate significance, red and blue
1090 highlights positive and negative correlations, respectively.

1091

1092 **Fig 2. The mH2A1.1 isoform is enriched at super-enhancers.** (A) Genome
1093 browser view of ChIP-seq profiles illustrating occupancy of mH2A1 isoforms with
1094 H3K27ac and H3K4me1 histone marks on genomic regions excluding TSS. (B)
1095 Overlap of “putative” enhancers (Methods) with mH2A1 isoforms. (C) Genome
1096 browser view of ChIP-seq profiles illustrating occupancy of mH2A1 isoforms (Ab
1097 α mH2A1.1 and Ab α mH2A1) at “putative” super-enhancers (SEs). SEs were defined
1098 using the ROSE package (Methods). (D) Overlap of SEs with mH2A1 isoforms. On
1099 genome browser views, peaks and unstranded RNA-seq signal are shown in parallel.
1100 The black arrows show the direction of transcription. On venn diagrams, percentages
1101 of overlaps are given. Results of fisher exact tests are shown (p-value (p) and Odd
1102 ratio) (see Materials and Methods).

1103

1104 **Fig 3. mH2A1.1 is recruited to the transcription start site of active genes.** (A)
1105 Overlap of mH2A1 isoforms with TSSs (+/- 1kb). (B) Metagene plot of average (+/-
1106 s.d) of mH2A1 isoforms enrichment at TSSs (+/- 2 kb) categorized by gene
1107 expression levels (Methods). Colors denote gene classes as indicated. (C) Genome
1108 browser view of ChIP-seq profiles illustrating the binding of mH2A1.1 to the TSS of a
1109 transcribed gene in an open chromatin state. Peaks and unstranded RNA-seq signal
1110 are shown in parallel. The black arrows show the direction of transcription. (D)
1111 Heatmap showing peak enrichment of mH2A1 isoforms sites (see Fig 1A) at TSSs
1112 (+/- 1kb), divided into 5 equal size categories as a function of expression level
1113 (Methods). Stars indicate significance. Red and blue highlights positive and negative

1114 correlations, respectively.(E) Metagene plot of average (+/- s.d) of mH2A1.1,
1115 H3K27ac and PolII enrichment at TSS (+/- 2kb). (F) Overlap of mH2A1 isoforms with
1116 Pol II at TSSs. On venn diagrams, percentages of overlaps are given. Results of
1117 fisher exact tests are shown (p-value (p) and Odd ratio) (see Materials and Methods).
1118

1119 **Fig 4. mH2A1.1 activates gene expression when it binds only on their TSSs.** (A)
1120 Volcano plot showing fold change of gene expression in mH2A1.1 KD compared to
1121 WT MDA-MB231 cells. Red dots represent significantly de-regulated genes with a
1122 foldchange > 1.5 and padj < 0.1. Total mH2A1.1 regulated-genes are shown. (B)
1123 Heatmap showing mH2A1 isoforms and Pol II relative enrichment around the TSS
1124 (+/- 10 kb) of mH2A1.1 KD de-regulated genes (see Fig 4A). Colour intensity reflects
1125 level of ChIP-seq enrichment. (C) Genome browser view of ChIP-seq profiles
1126 illustrating mH2A1.1 localization and genomic environment of a mH2A1.1 repressed-
1127 gene. (D) As in (C) but for a mH2A1.1 activated- gene. On genome browser views,
1128 peaks and unstranded RNA-seq signal are shown in parallel. The black arrows show
1129 the direction of transcription.

1130

1131 **Fig 5. mH2A1.1 binds a subset of TSSs of paused genes and promotes their**
1132 **transcription.** (A) Heatmap showing Pol II and mH2A1 isoforms enrichment around
1133 the TSS (+/- 10 kb) of transcribed genes (n=10,198) ranked by their pausing index
1134 (Methods). Colour intensity reflects level of ChIP-seq enrichment. (B) Heatmap
1135 showing correlation between mH2A1.1 peak width at TSSs (+/-1 kb) and pausing
1136 index. Genes were divided in 3 equal size categories according to their pausing
1137 index. mH2A1.1 peaks were divided in 5 categories according to their width. Stars
1138 indicate significance. Red and blue highlights positive and negative correlations,

1139 respectively. (C) Genome browser view of ChIP-seq profiles illustrating mH2A1.1
1140 localization on a paused gene. Peaks and unstranded RNA-seq signal are shown in
1141 parallel. The black arrows show the direction of transcription. (D) Overlap of mH2A1
1142 isoforms sites (mH2A1.1 only and mH2A1.2 only, see Fig 1A) with paused genes
1143 (Pausing index , PI > 2, n=6,821) at their TSSs. (E) Overlap of mH2A1.1 regulated-
1144 genes with paused genes (PI > 2, n=6,821). (F) Boxplot showing pausing index of
1145 five different groups of genes, as indicated. “****” = p-value < 2.2x10⁻¹⁶. NS, not
1146 significant. On venn diagrams, percentages of overlaps are given. Results of fisher
1147 exact tests are shown (p-value (p) and Odd ratio) (see Materials and Methods).

1148

1149 **Fig 6. mH2A1.1 inhibits cell migration by in part favouring expression of**
1150 **paused genes involved in cytoskeleton and cell adhesion in MDA-MB231 cells.**

1151 (A) Representative DIC microscopy images of WT, mH2A1.1 KD (two different
1152 siRNA) and mH2A1.2 KD MDA-MB231 cells. Scale bar = 100 μM. (B)
1153 Immunofluorescence of Actin (up), Tubulin-α (middle) and Vimentin (down) in WT,
1154 mH2A1.1 KD and mH2A1.2 KD MDA-MB231 cells. Nuclei are stained with Hoechst.
1155 Scale bar = 20 μM. (C) Boyden chamber assay representatives images of WT,
1156 mH2A1.1 KD and mH2A1.2 KD MDA-MB231 cells. Only migrated cells are labelled in
1157 purple. Scale bar = 150 μM. (D) Quantification of Boyden chamber assay presented
1158 in (C). Error bar represents s.d from n=3 independent experiments as illustrated in
1159 (C). “*” = p-value (p) < 0.05, **, p < 0.01. (E) Overlap of paused genes (n= 6,821)
1160 with mH2A1.1 regulated genes related to cytoskeleton and cell adhesion (see
1161 Materials and Methods). On venn diagrams, percentages of overlaps are given.
1162 Results of fisher exact tests are shown (p-value (p) and Odd ratio) (see Materials and
1163 Methods).

1164

1165 **S1 Fig. The antibody Ab α H2A1.1 recognizes unequivocally only the isoform**
1166 **mH2A1.1.** (A) RTqPCR on MDA-MB231 and MCF7 cells showing their relative
1167 mH2A1 isoforms mRNA expression level. Relative expressions are normalized by
1168 RPLP0 mRNA. (B) Western blot on whole cell extracts of MDA-MB231 and MCF7
1169 cells showing the better affinity of Ab α H2A1.1 to recognize mH2A1.1 compared to
1170 Ab#37264 (Ab α H2A1). Ab#61427 is specific to mH2A1.2. GAPDH is used as a
1171 loading control. (C) Western blot showing specific recognition of mH2A1.1 isoform by
1172 Ab α H2A1.1 antibody. For that, HEK-293T cells were transfected with plasmids
1173 coding for Flag-mH2A1.1 (Flag1.1) or Flag-mH2A1.2 (Flag1.2) fusion overexpressed-
1174 protein. Western blot was then done with Ab α H2A1.1, Ab#Flag and Ab#E215 (that
1175 preferentially recognizes mH2A1.2) antibodies on whole cell extracts. GAPDH is
1176 used as a loading control. (D) Immunofluorescence in HEK-293T cells showing
1177 specific recognition of mH2A1.1 isoform by Ab α H2A1.1. (E) Western blot on ChIP
1178 extracts from HEK-293T cells overexpressing Flag1.1 or Flag1.2 showing that Ab
1179 α H2A1.1 immunoprecipitates only mH2A1.1 isoform (Methods). Different extracts
1180 were loaded : Input fraction (Input), Non immunoprecipitated fraction (NoIP) and
1181 immunoprecipitated fraction (IP). Percentages represent percentages loaded on
1182 western blot compared to quantity used for ChIP. (F) Western blot showing that Ab
1183 α H2A1.1 is also working in ChIP in MDA-MB231 cells on the endogenous protein.
1184 (G) As in (C), but for Ab α H2A1 (Ab#37264) antibody showing that its antibody
1185 recognizes both isoforms but it less affine for Flag1.1 than Ab α H2A1.1. (H) As in
1186 (G), but in immunofluorescence. (I), As in (E), but for Ab α H2A1 (Ab#37264)
1187 antibody.

1188 **S2 Fig.** (A) Whole genome spearman correlation heatmap of ChIP-seq data.
1189 Pearson coefficient correlations (PCC, r) are given. Red and blue colours denote
1190 high correlation (r close to 1) and anti-correlation (r close to -1), respectively. (B)
1191 Table showing PCCs obtained between mH2A1.1 and mH2A1 at different genomic
1192 region, as presented.

1193 **S3 Fig. Specific partial depletion of mH2A1 isoforms in MDA-MB231 cells.** (A)
1194 RTqPCR showing specific partial depletion of mH2A1 isoforms mRNA by specific
1195 siRNA three days post-transfection (see Materials and Methods). (B) Western blot
1196 showing specific depletion of mH2A1 isoforms protein by specific siRNA three days
1197 post-transfection. H3 is used as a loading control. (C) Immunofluorescence showing
1198 specific partial depletion of mH2A1 isoforms by specific siRNA three days post-
1199 transfection. DNA is labelled with Hoechst. Scale bar = 10 μ m. (D) As in (A) but with
1200 a second siRNA against mH2A1.1. (E) As in (B) but with a second siRNA against
1201 mH2A1.1. GAPDH is used as a loading control. (F) RTqPCR analysis of a subset of
1202 RNAseq-defined mH2A1.1 regulated-genes. Genes are divided in three groups, as
1203 indicated. On RTqPCR, mRNA expressions are relatives to WT condition and
1204 normalized by RPLP0 mRNA. Error bars represent s.d from independent
1205 experiments ($n \geq 2$). “*” : $p < 0.05$, “***” : $p < 0.001$, NS, not significant. On western
1206 blot, mH2A1.1 and mH2A1.2 proteins are detected with Ab α mH2A1.1 and
1207 Ab#61427, respectively. Quantifications are showed, normalized by protein loading
1208 control.

1209 **S4 Fig.** (A) mH2A1 and mH2A1.1 binding at genomic regions selected based on
1210 ChIP-sequencing. Localisation of primers used for ChIP-qPCR are shown in red. (B)
1211 Occupancy of mH2A1 isoforms (left part : Ab α mH2A1.1 ; right part : Ab α mH2A1)
1212 analysed by ChIP-qPCR in control cells (WT) and cells partially deficient for

1213 mH2A1.1 using two different siRNA (mH2A1.1 KD #1 and mH2A1.1 KD #2). Error
1214 bars represent +s.d from independent experiments (n>=2).

1215 **S5 Fig. Levels of H3K27me3 and H3K9me3 are anti-correlated on common**
1216 **H3K27me3/H3K9me3 peaks.** (A) Boxplots showing H3K27me3 and H3K9me3
1217 enrichment levels on H3K27me3/H3K9me3 common peaks. Peaks were divided into
1218 5 equal size categories according to the level of H3K27me3, as mentioned. (B)
1219 Histogram showing proportions of heterochromatin genomic localization, as
1220 mentioned. Peaks were divided into 5 equal size categories according to ratio
1221 between H3K27me3 and H3K9me3, as mentioned.

1222 **S6 Fig. The mH2A1.1 isoform binds to TSSs in open chromatin state whereas**
1223 **mH2A1.2 binds to TSSs in closed chromatin state.** (A) TSS (+/- 1kb) centred
1224 spearman correlation heatmap of ChIP-seq data. Pearson coefficient correlations
1225 (PCC, r) are given. Red and blue colours denote high correlation (r close to 1) and
1226 anti-correlation (r close to -1), respectively. (B) Overlap of mH2A1 isoforms sites
1227 (mH2A1.1 only and mH2A1.2 only, see Fig. 1a) with H3K4me3 and Pol II at TSS (+/-
1228 1 kb). (C) As in (B), but with H3K9me3 and H3K27me3 histone marks. Results of
1229 fisher exact tests are shown (p-value (p) and Odd ratio) (see Materials and Methods).

1230 **S7 Fig. The mH2A1.1 isoform binds the TSSs of transcribed genes.** (A) Boxplots
1231 of mH2A1 isoforms levels and Pol II at TSS (+/- 1kb) categorized by gene expression
1232 levels (see Materials and Methods). Colors denote gene classes, as indicated. “****”
1233 p-value < 2.2x10⁻¹⁶ (see Materials and Methods). (B) Metagene plot of average (+/-
1234 s.d) of mH2A1 isoforms enrichment from TSS to TES (+/- 2 kb) categorized by gene
1235 expression levels (Methods). Colors denote gene classes, as indicated. (C) Genome
1236 browser view of ChIP-seq profiles illustrating occupancy of mH2A1 isoforms on a not
1237 expressed gene. (D) As in (C) but for a highly expressed gene. (E) Heatmap showing

1238 peak enrichment of Pol II and histone marks at TSS (+/- 1kb) divided into 5 equal
1239 size categories as a function of expression level (Methods). Stars indicate
1240 significance. Red highlights positive and blue indicates negative correlations. (F)
1241 TSS-centred spearman correlation heatmap of ChIP-seq data. Pearson coefficient
1242 correlations (PCC, r) are given. Red and blue colours denote high correlation (r close
1243 to 1) and anti-correlation (r close to -1), respectively. On genome browser views,
1244 peaks and unstranded RNA-seq signal are shown in parallel. The black arrows show
1245 the direction of transcription.

1246

1247 **S8 Fig. The mH2A1.1 isoform regulates expression of active genes.** (A) Boxplot
1248 comparing gene expression of three groups of genes (as indicated) between WT and
1249 mH2A1.1 KD cells. (B) Pie chart showing proportion of mH2A1.1 regulated-genes
1250 categorized by gene expression levels in WT cells (see Materials and Methods).
1251 Colors denote gene classes, as indicated. Results of fisher exact tests are shown (p-
1252 value (p) and Odd ratio) (see Materials and Methods). (C) Boxplot showing mH2A1
1253 isoforms occupancy levels, Pol II and histone marks at TSS on mH2A1.1 KD
1254 regulated genes and not affected genes, as indicated. On boxplots, p-value are
1255 shown, “****” = p-value < 2.2×10^{-16} ; NS, not significant (see Materials and Methods).

1256

1257 **S9 Fig. The recruitment of mH2A1.1 differs between mH2A1.1-activated and**
1258 **repressed genes.** (A) Metagene plot of average (+/- sd) of mH2A1 isoforms levels
1259 and Pol II from TSS to TES (+/- 2 kb) on mH2A1.1 regulated-genes. (B) Heatmap
1260 showing correlation between Pol II peak width at TSS (+/-1 kb) and pausing index.
1261 Genes were divided in 3 equal size categories according to their pausing index. Pol II

1262 peaks were divided in 5 categories according to their width. Stars indicate
1263 significance. Red and blue highlights positive and negative correlations, respectively.

1264

1265 **S10 Fig. The isoform mH2A1.1 modulates expression of genes involved in cell**
1266 **cycle, DNA repair and cell shape.** (A) List of gene ontology (GO) terms for
1267 mH2A1.1 activated-genes. The most significantly regulated ontologies were
1268 determined by adjusted p-value and are shown in three different classes, Biological
1269 Process (upper graph), Molecular function (middle graph) and Cellular Component
1270 (lower graph). A full list of enriched GO terms is provided in S4 Table. (B) As in (A)
1271 but for mH2A1.1 repressed-genes.

1272

1273 **Supporting information legends**

1274

1275 S1 Table. List of antibodies used for western blot, immunofluorescence, and ChIP.

1276 S2 Table. Summary of ChIP-seq and RNA-seq.

1277 S3 Table. List of mH2A1.1-regulated genes.

1278 S4 Table. List of GO terms related to mH2A1.1-regulated genes.

1279 S5 Table. Sequences used for mH2A1.1 specific antibody production and siRNAs.

1280 S6 Table. List of primers used for RTqPCR.

1281 S7 Table. List of ChIPqPCR primers used for ChIPqPCR.

1282

Fig 1

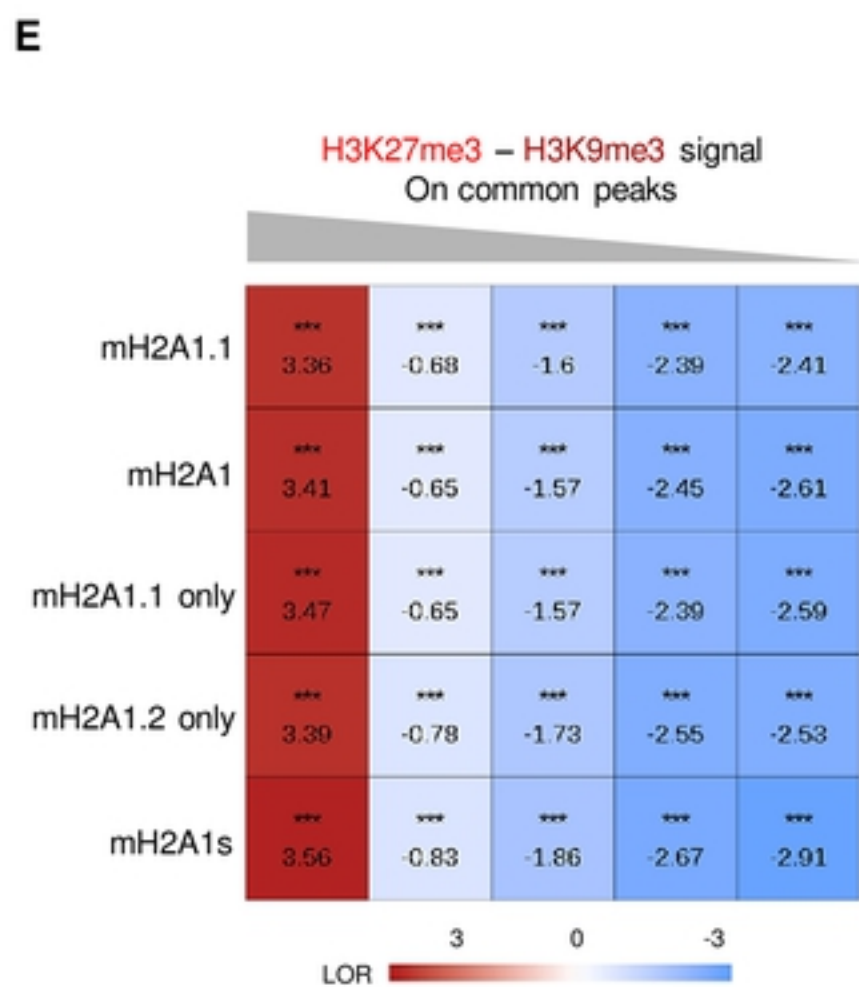
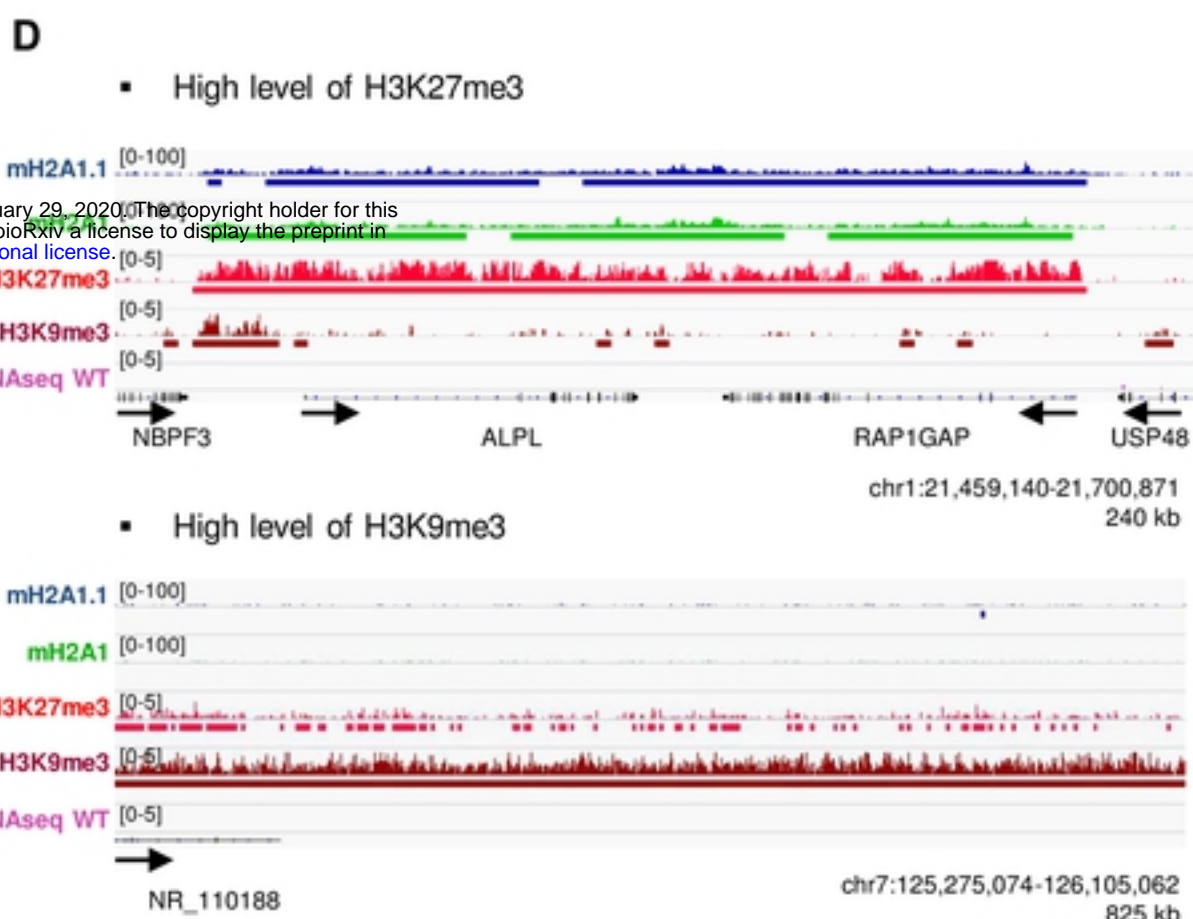
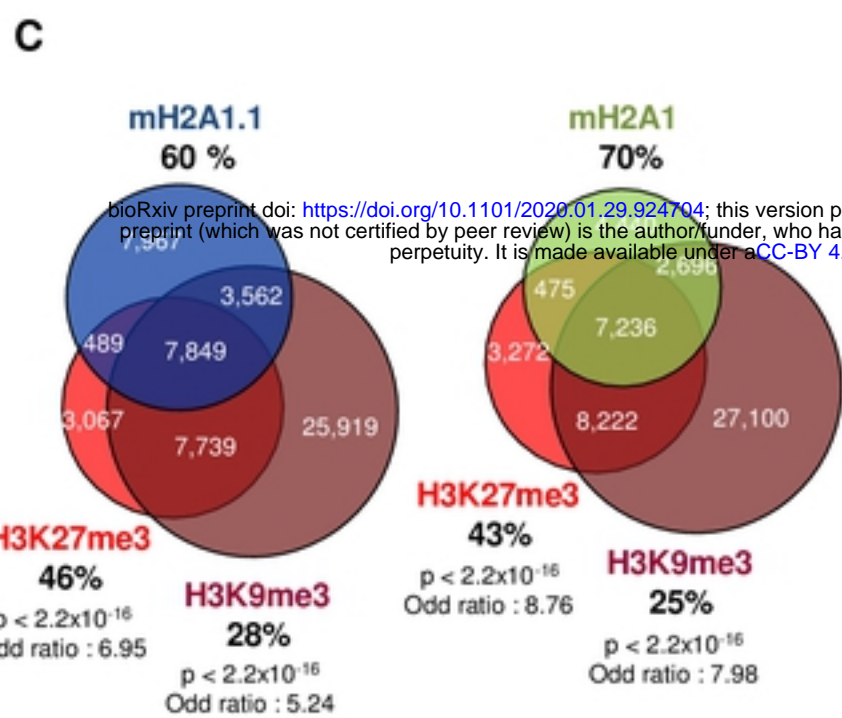
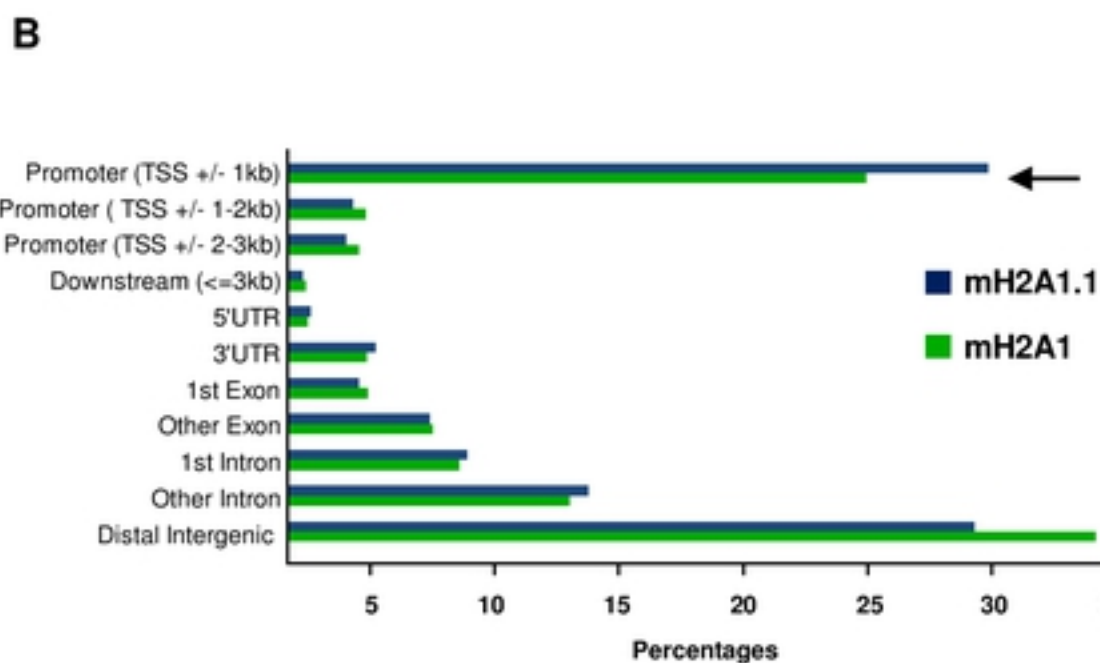
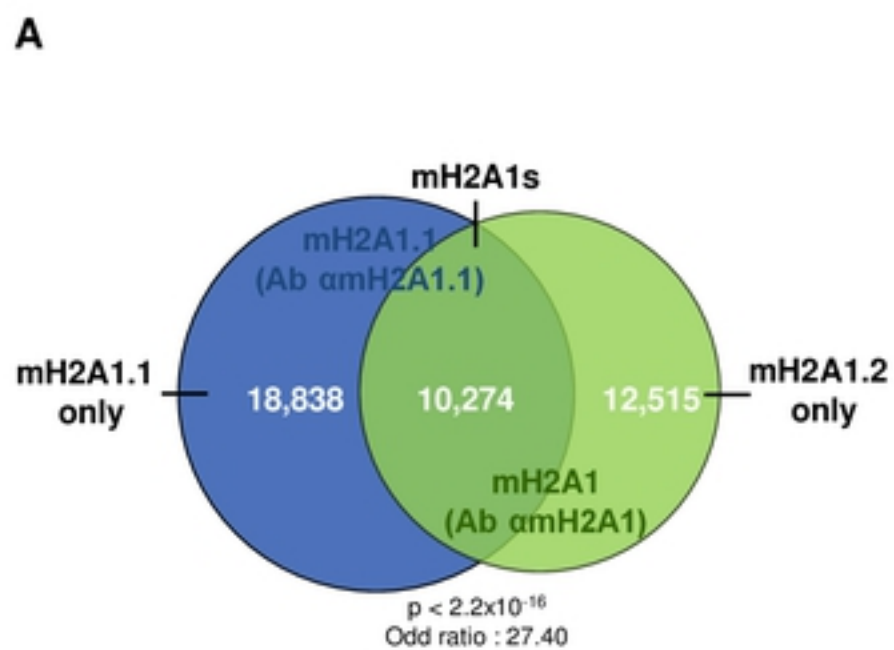
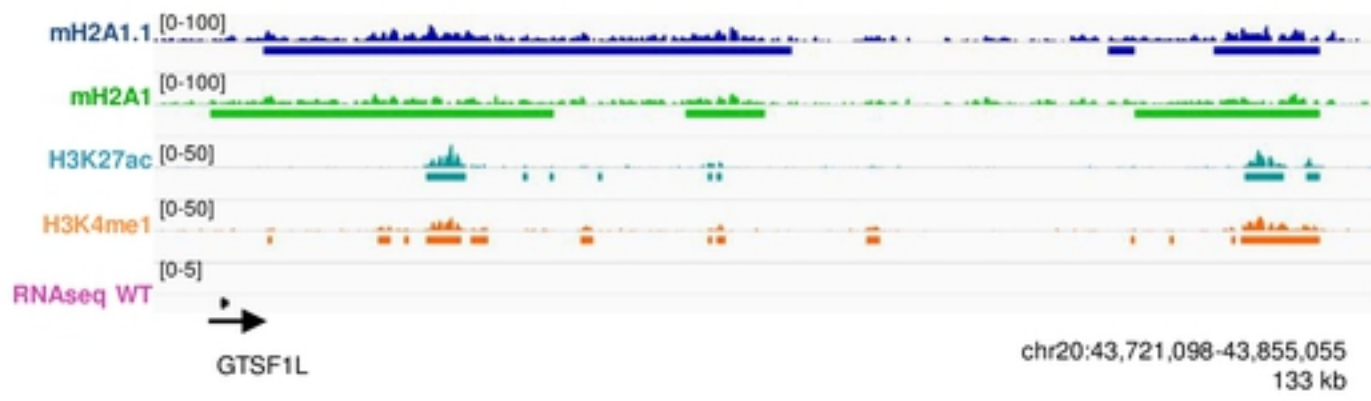
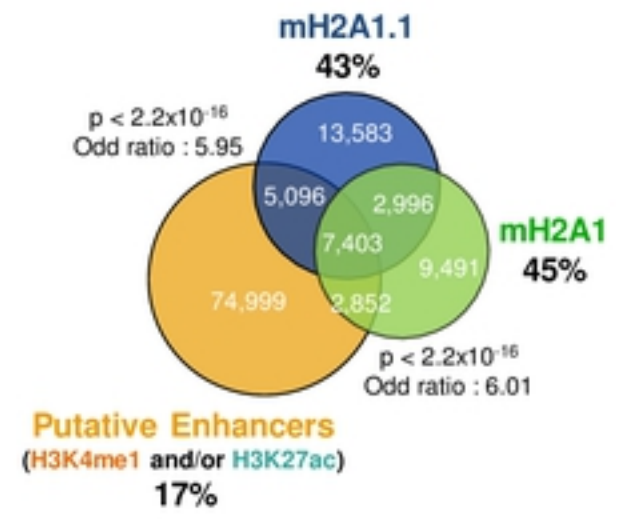


Fig 2

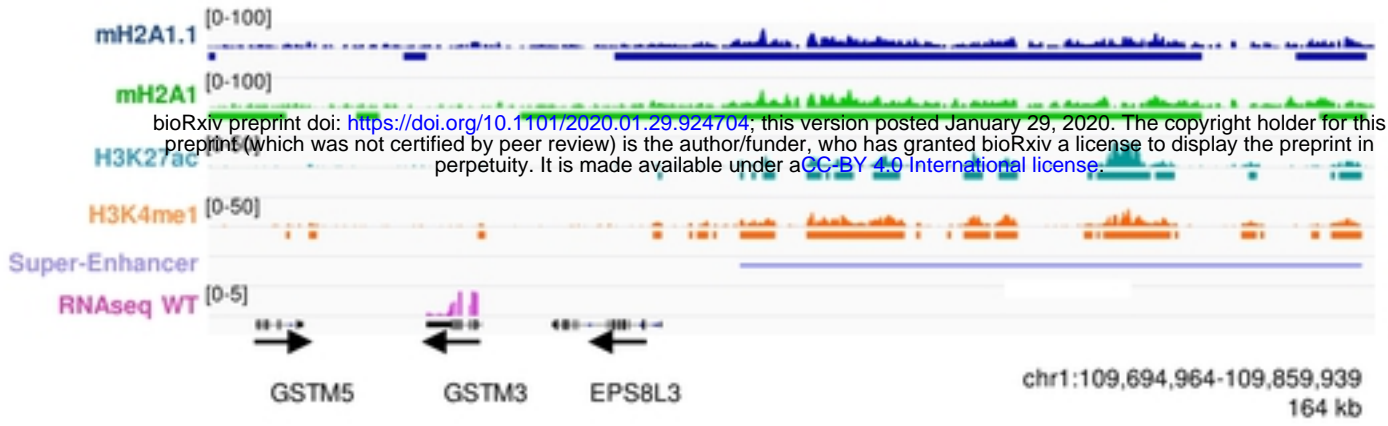
A



B



C



D

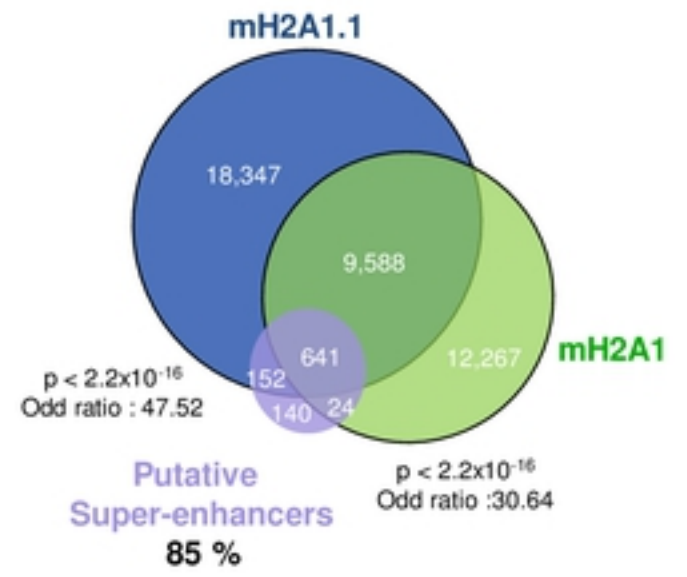
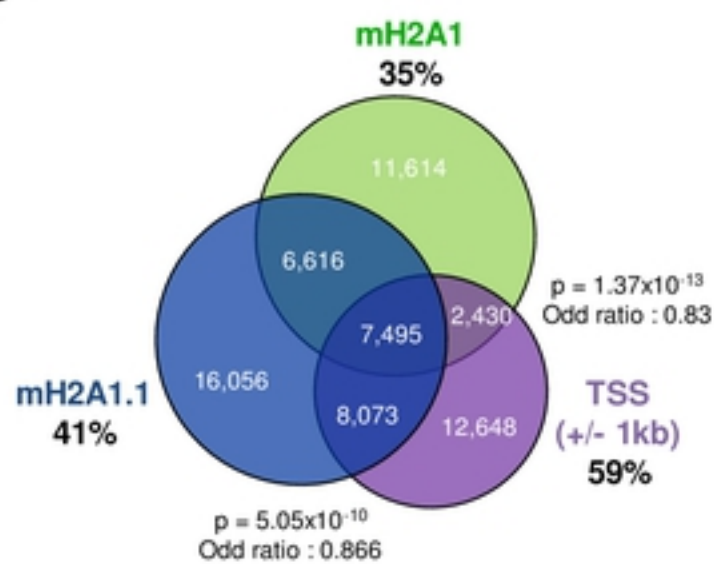
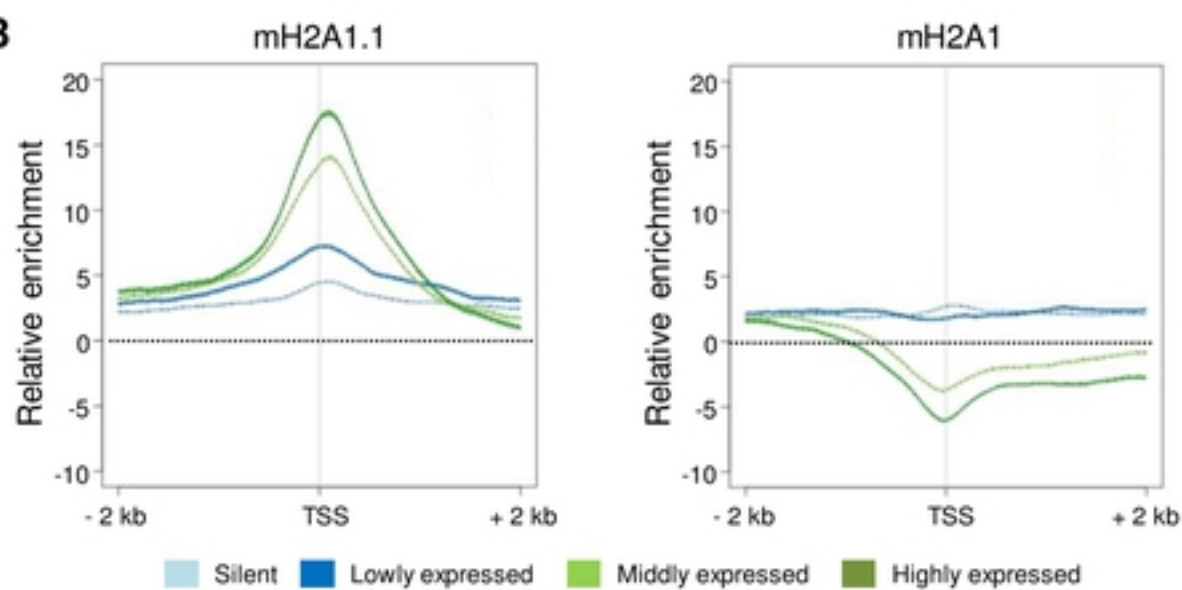


Fig 3

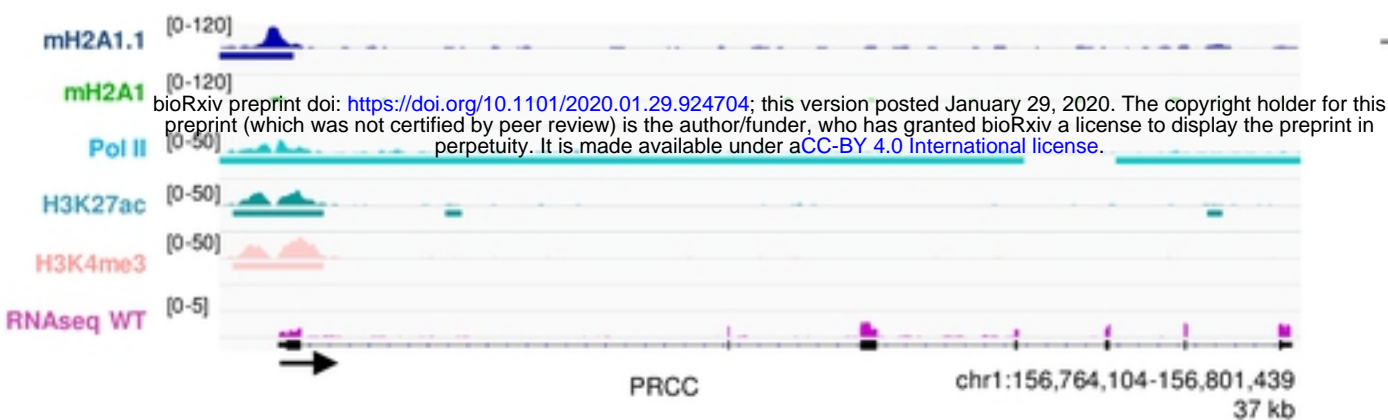
A



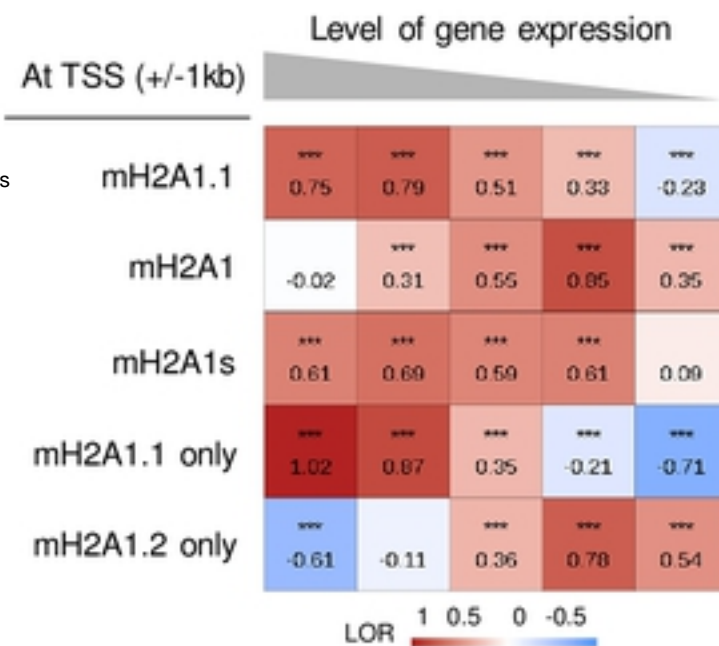
B



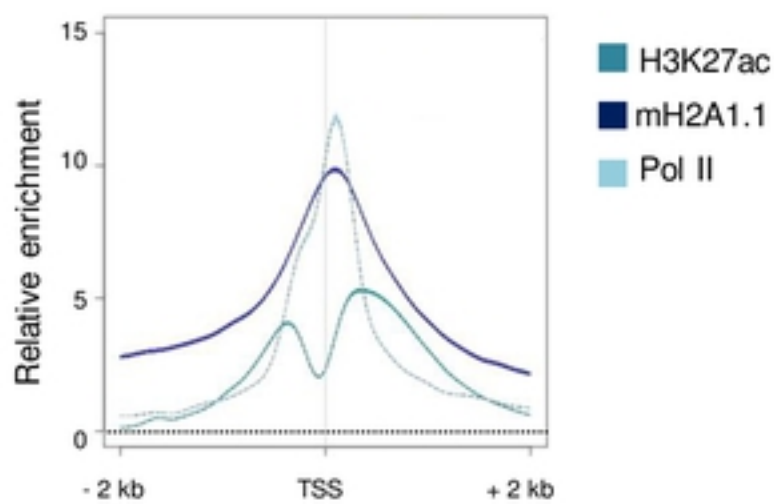
C



D



E



F

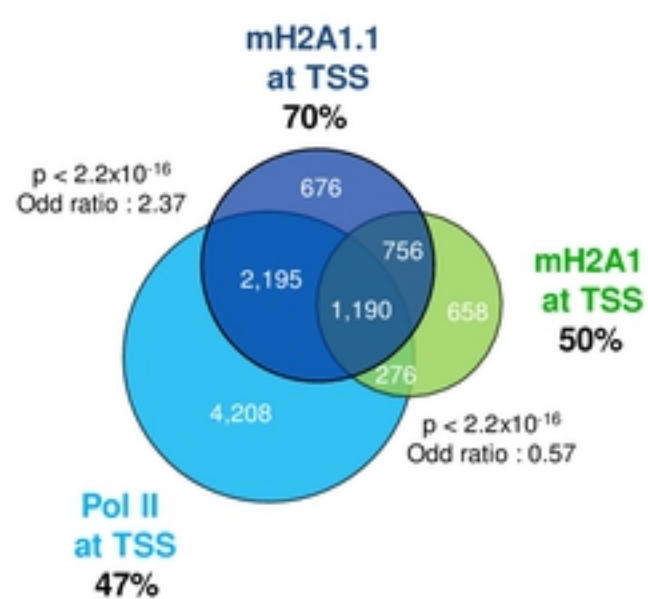


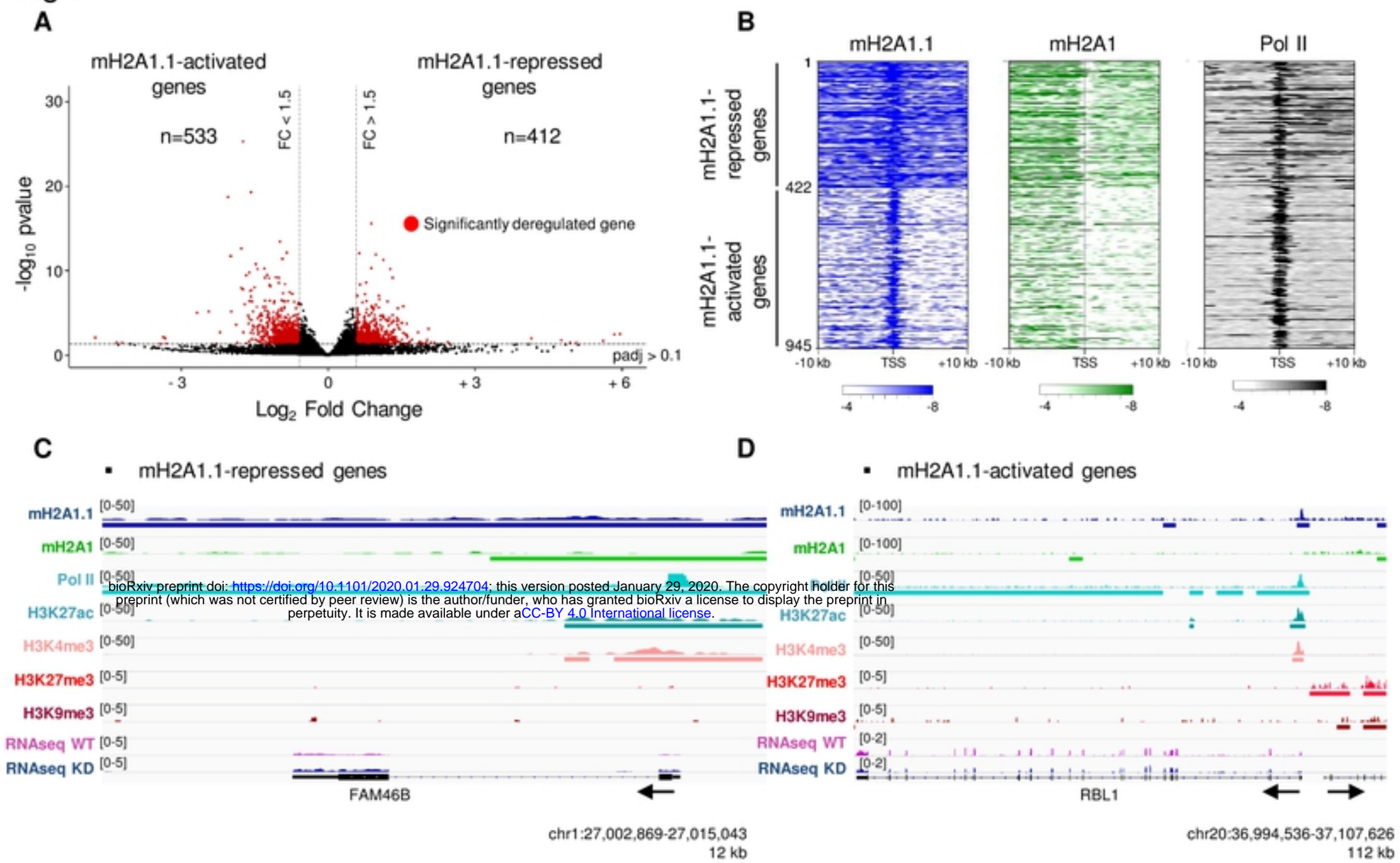
Fig 4

Fig 5

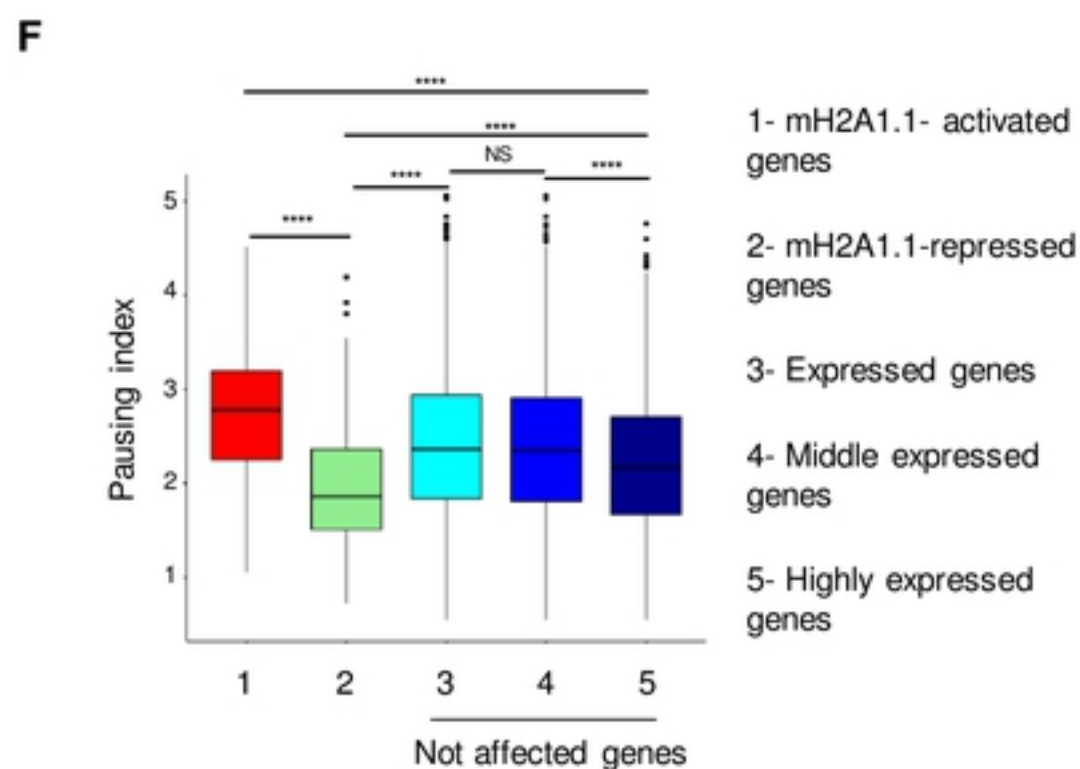
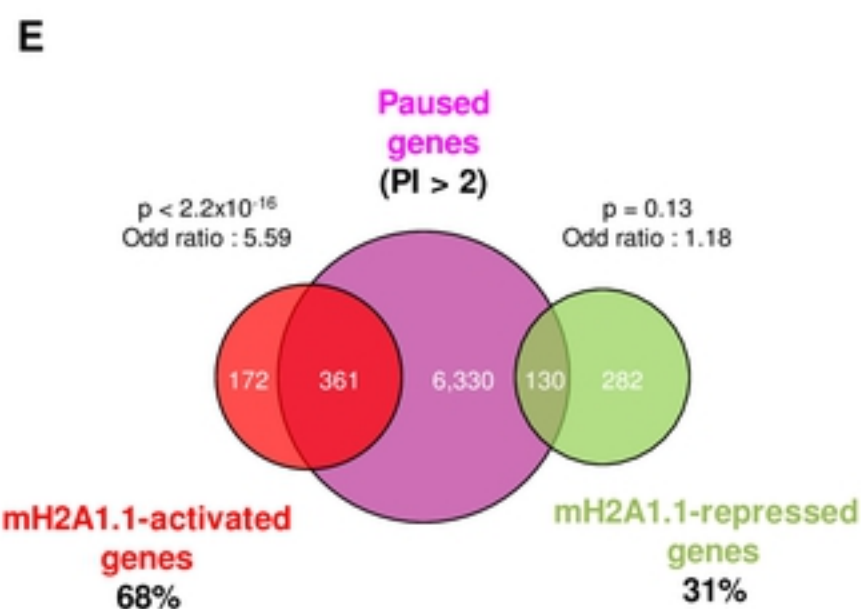
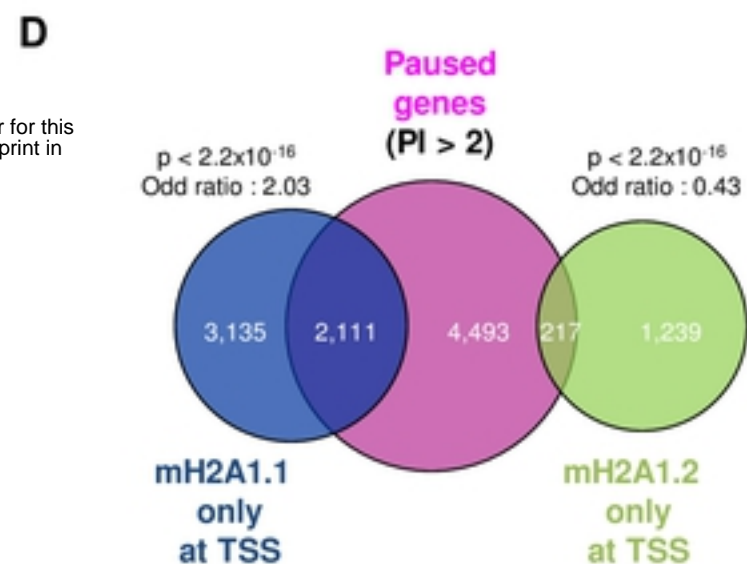
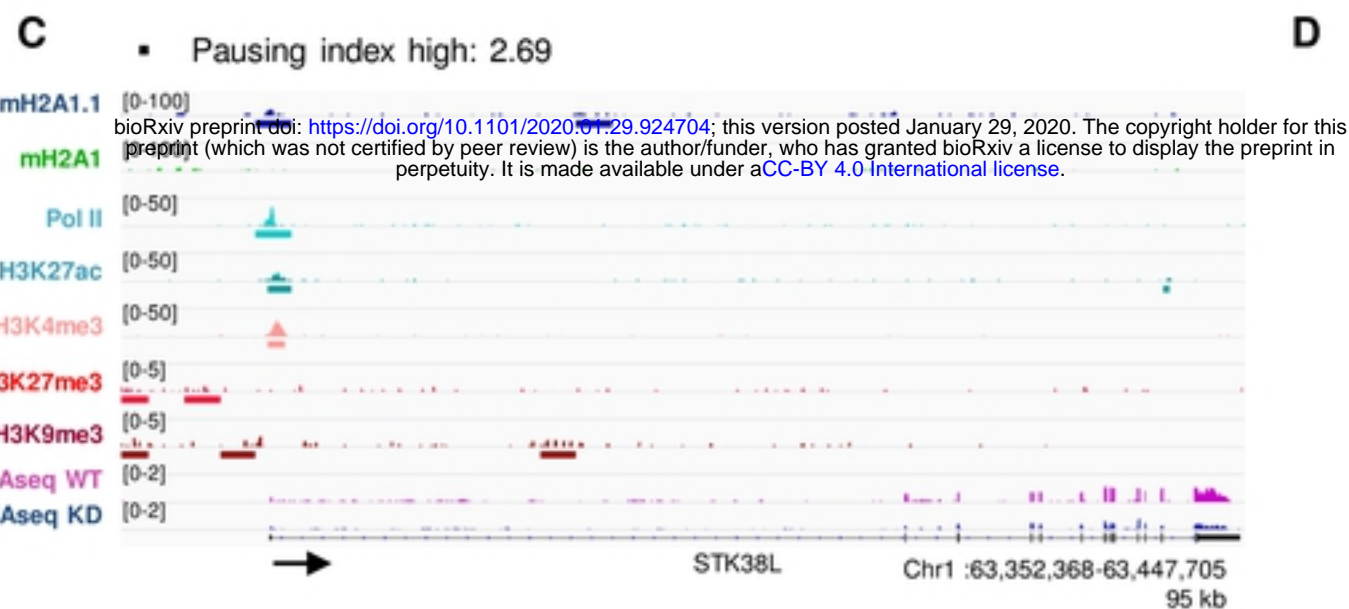
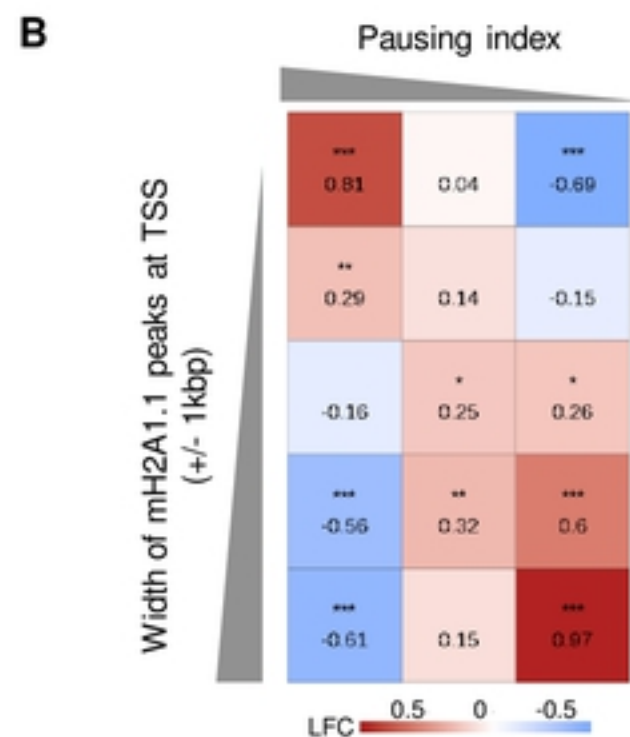
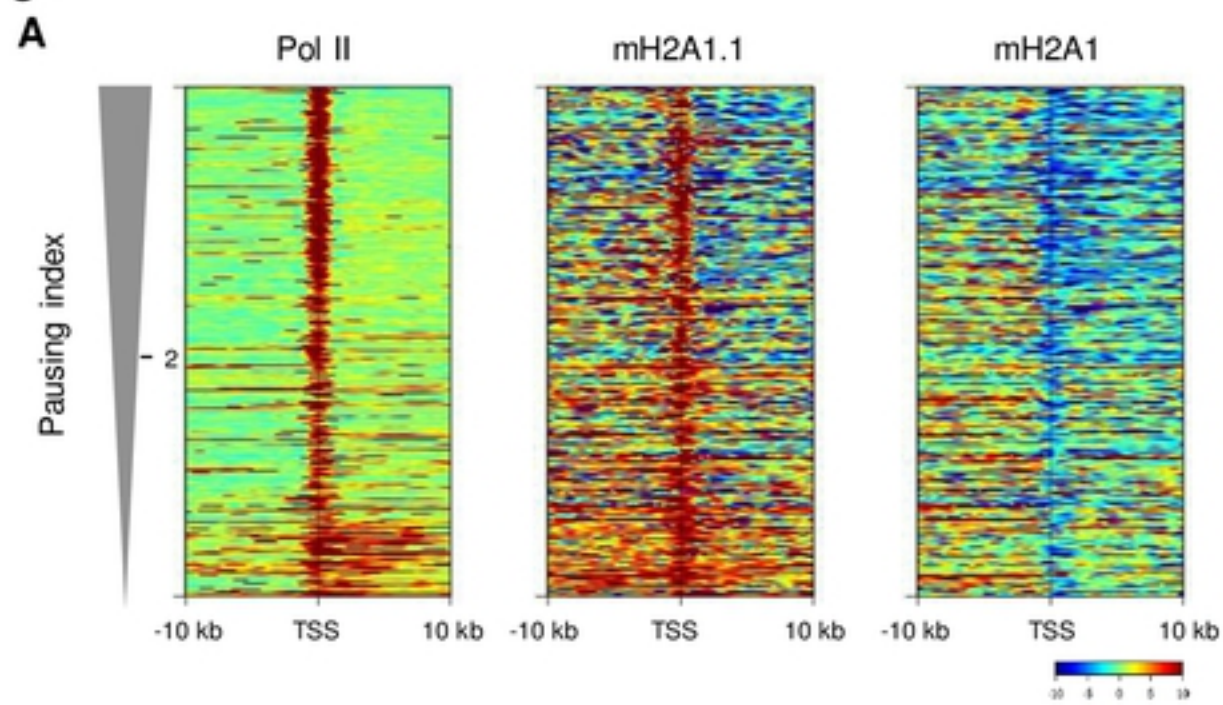
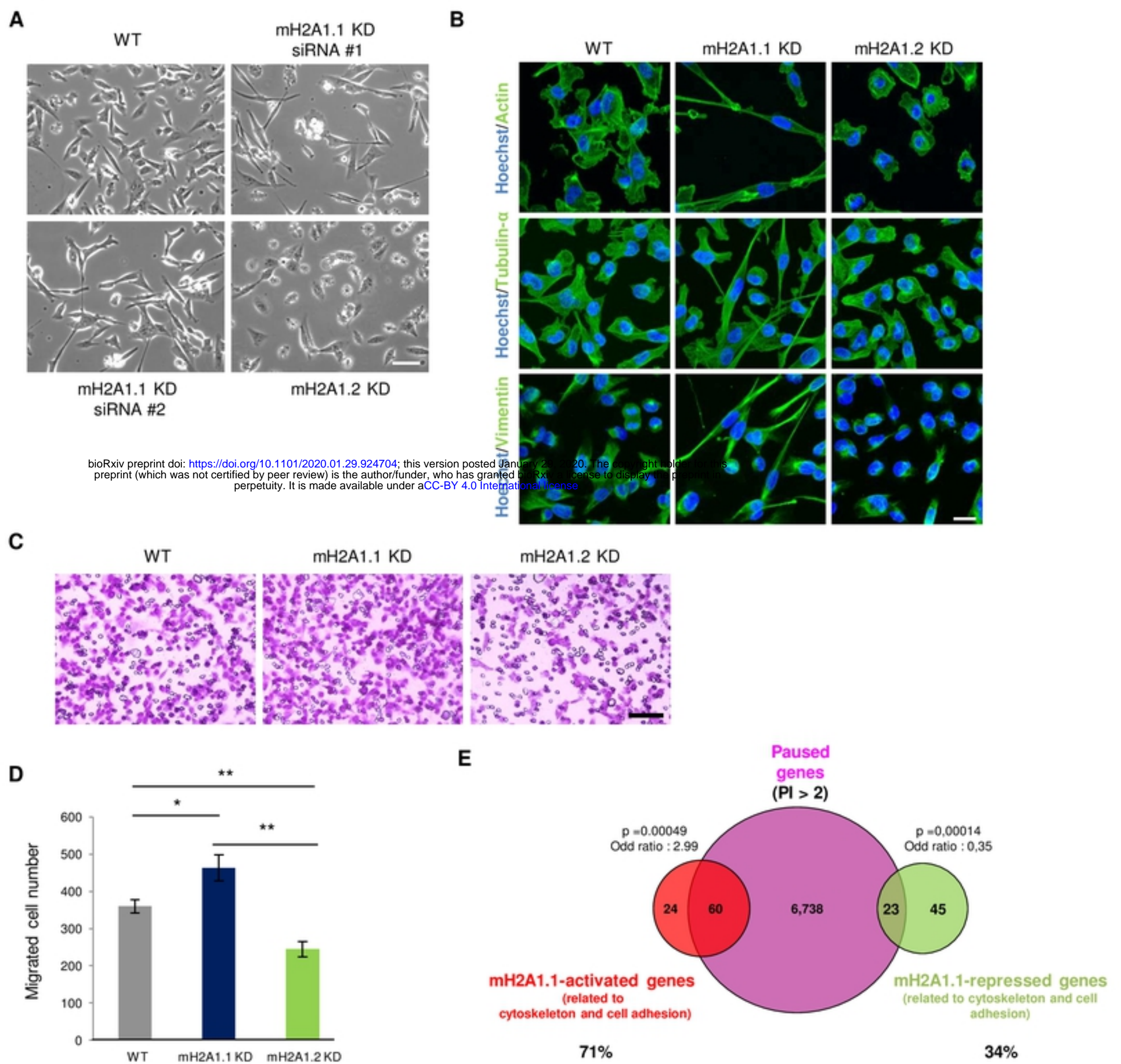
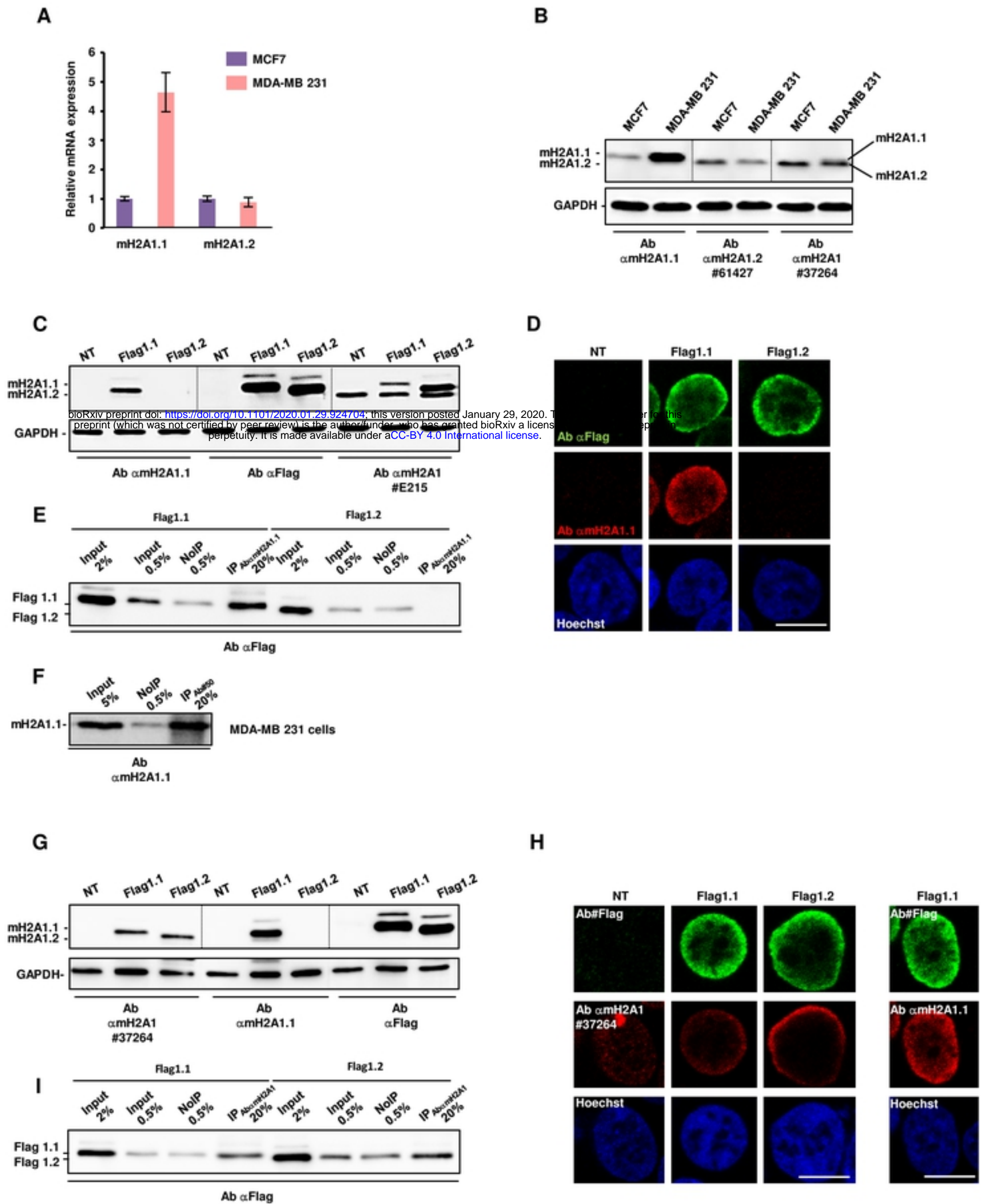
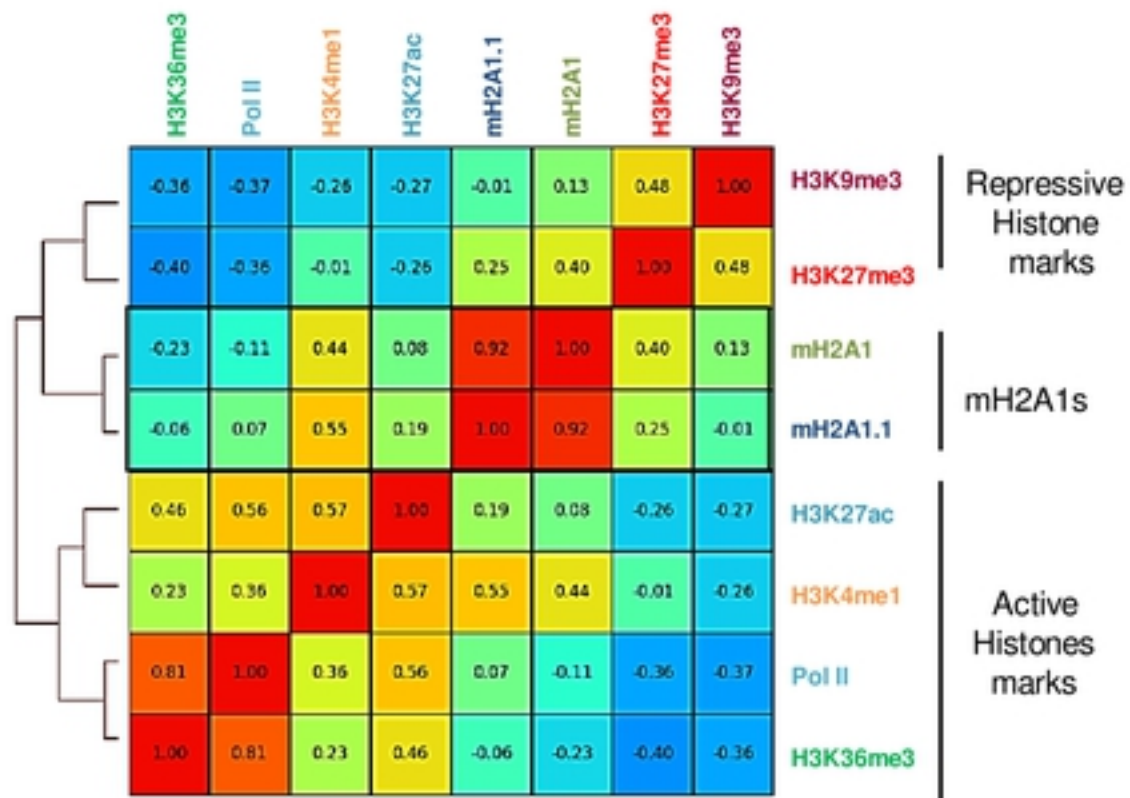


Fig 6

bioRxiv preprint doi: <https://doi.org/10.1101/2020.01.29.924704>; this version posted January 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



A

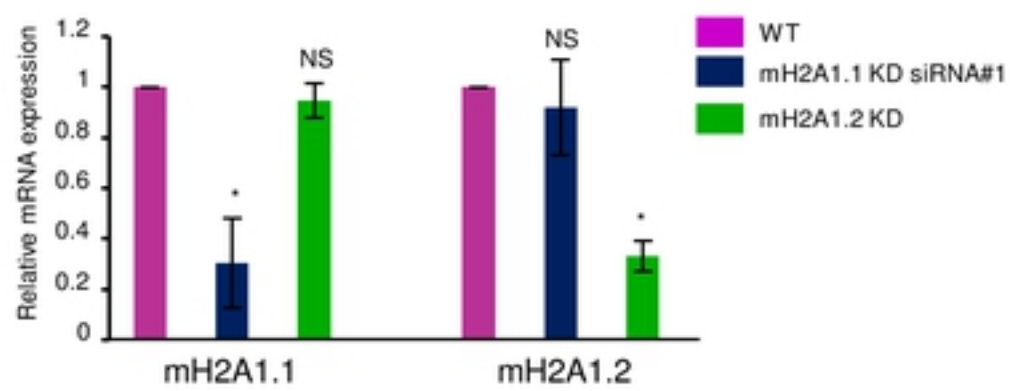


bioRxiv preprint doi: <https://doi.org/10.1101/2020.01.29.924704>; this version posted January 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

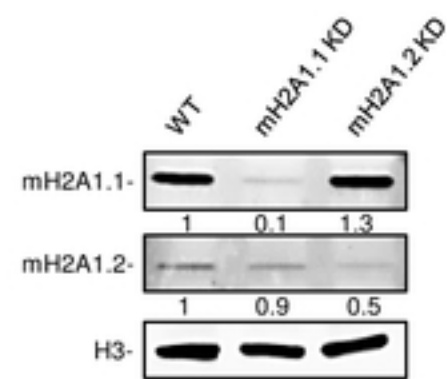
B

Genomic regions	PCC between mH2A1.1 and mH2A1
Heterochromatin	0.94
Enhancer	0.80
TSS +/- 1kb	0.41
TSS	-0.07

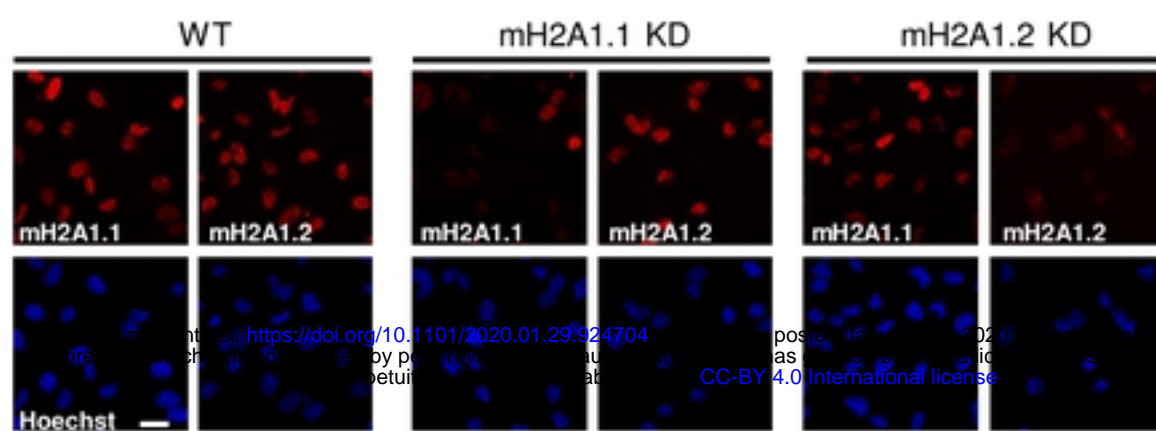
A



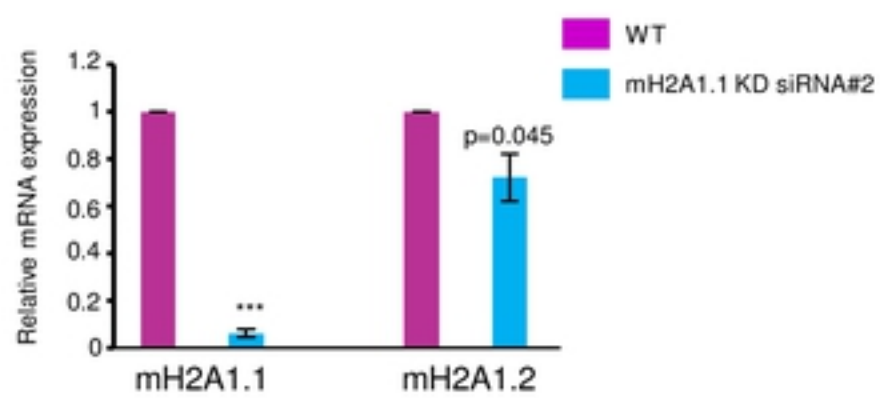
B



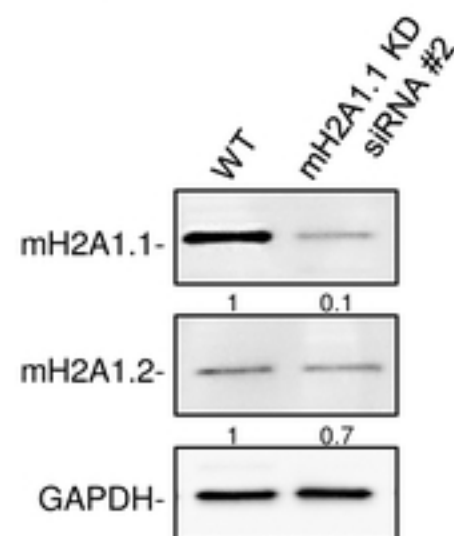
C



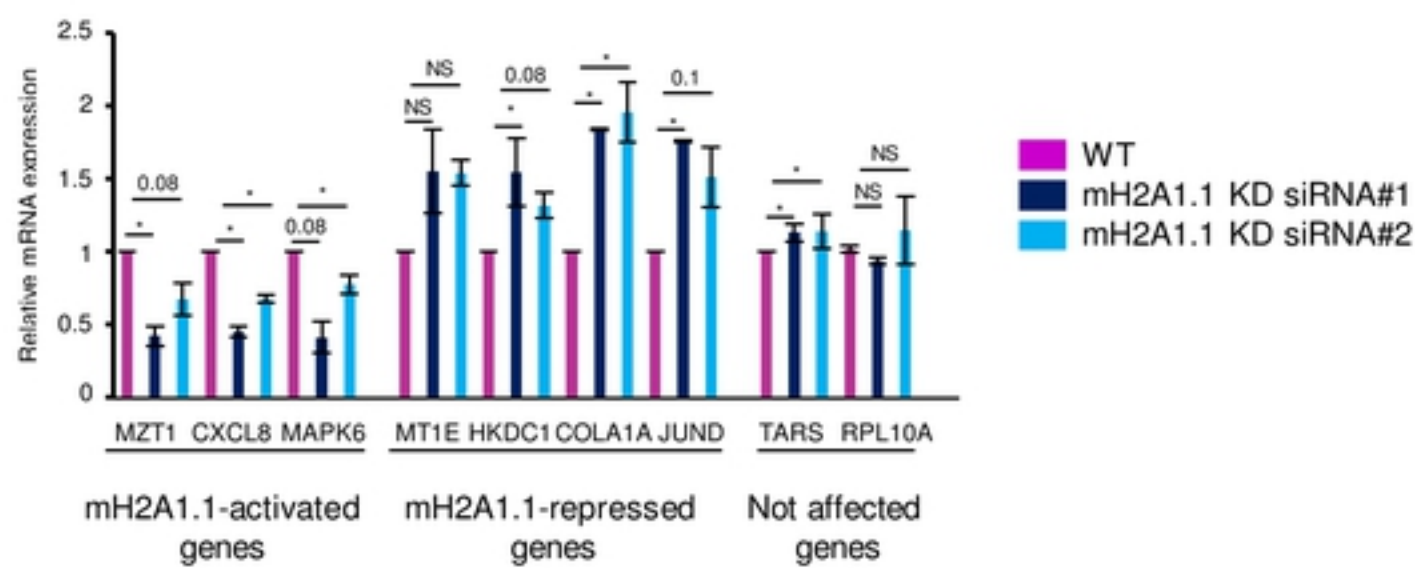
D



E

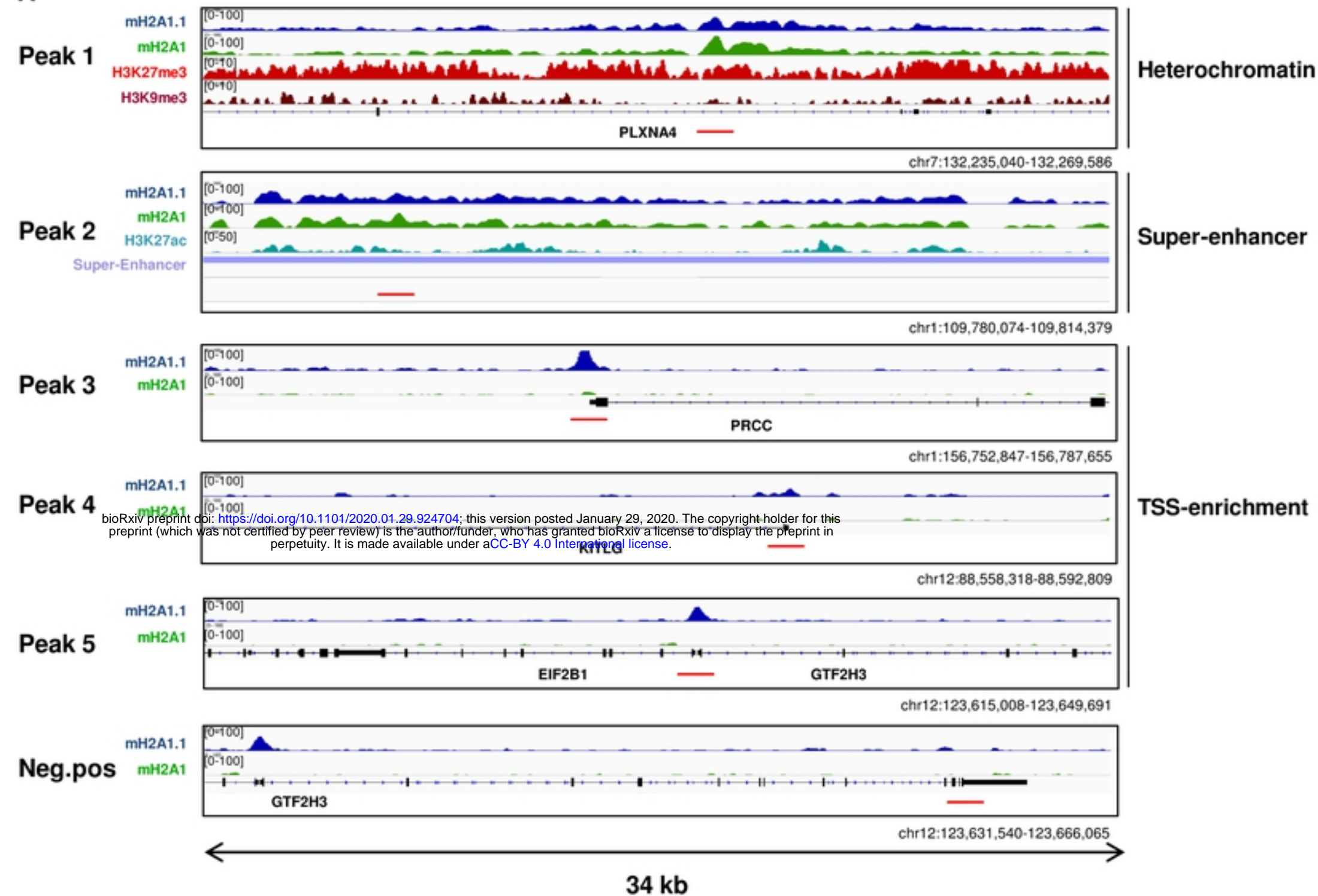


F



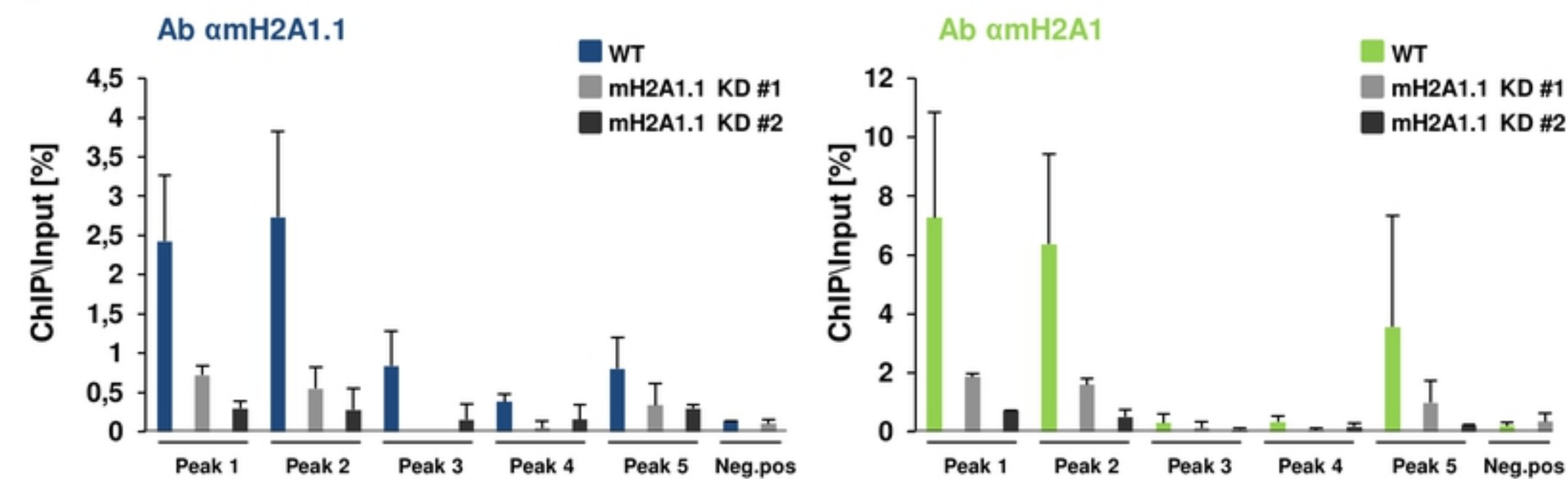
S4_Fig

A

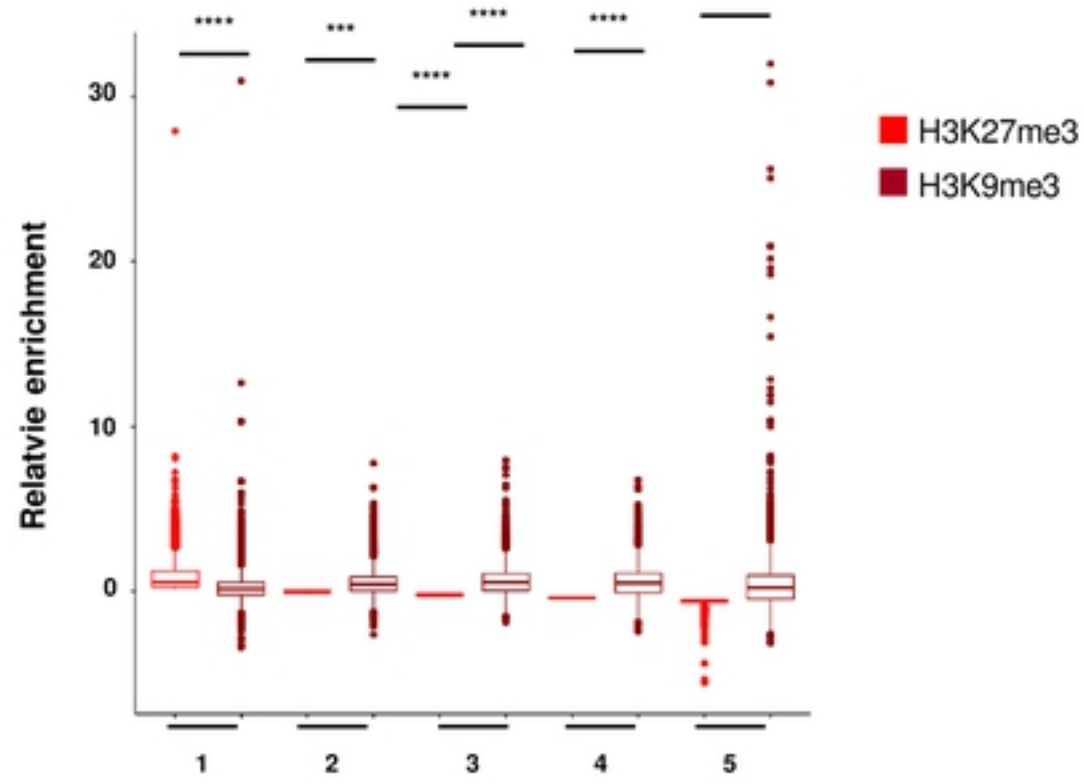


bioRxiv preprint doi: <https://doi.org/10.1101/2020.01.29.924704>; this version posted January 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

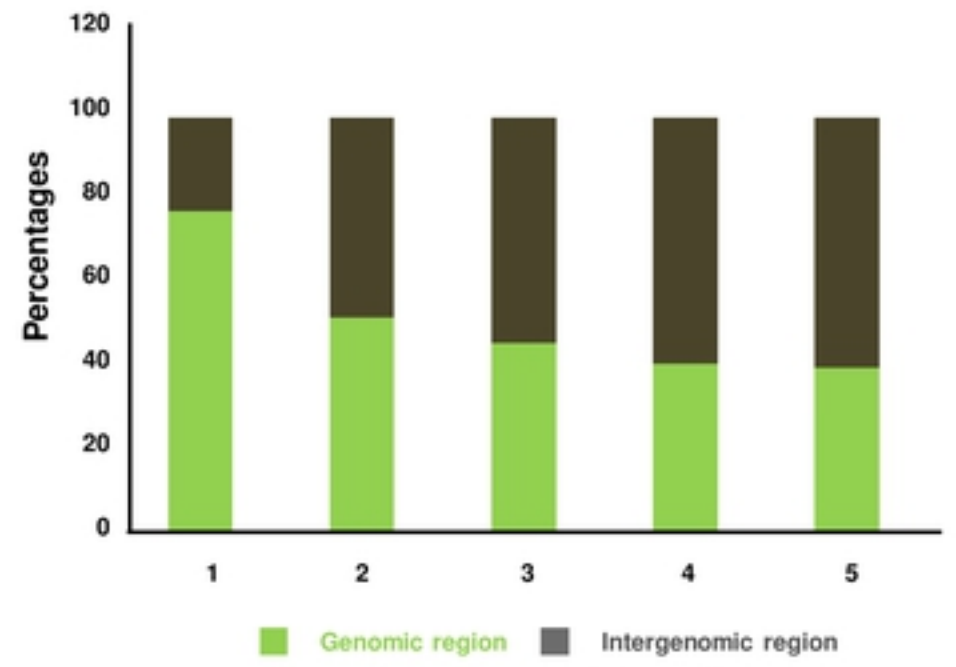
B



A



B



bioRxiv preprint doi: <https://doi.org/10.1101/2020.01.29.924704>; this version posted January 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 : Very high level of H3K27me3

2 : High level of H3K27me3

3 : Middle level of H3K27me3

4 : Low level of H3K27me3

5 : Very low level of H3K27me3

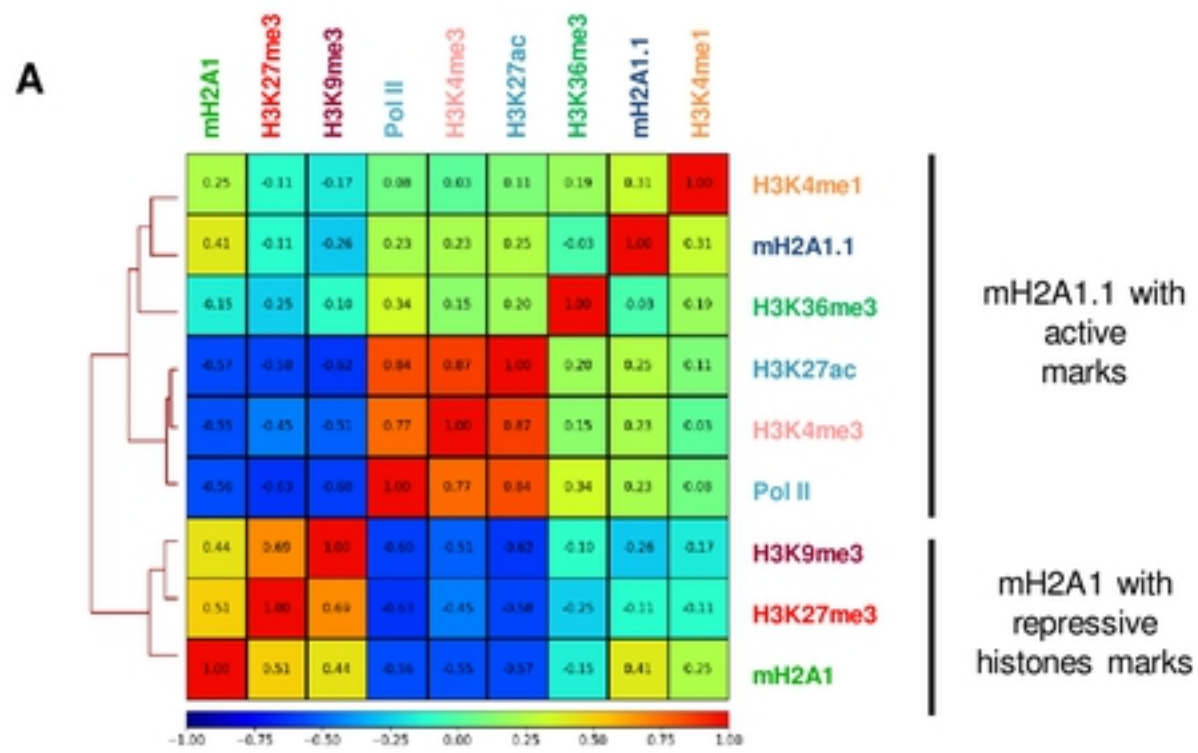
1 : Very high ratio of H3K27me3/H3K9me3

2 : High ratio of H3K27me3/H3K9me3

3 : Middle ratio of H3K27me3/H3K9me3

4 : Low ratio of H3K27me3/H3K9me3

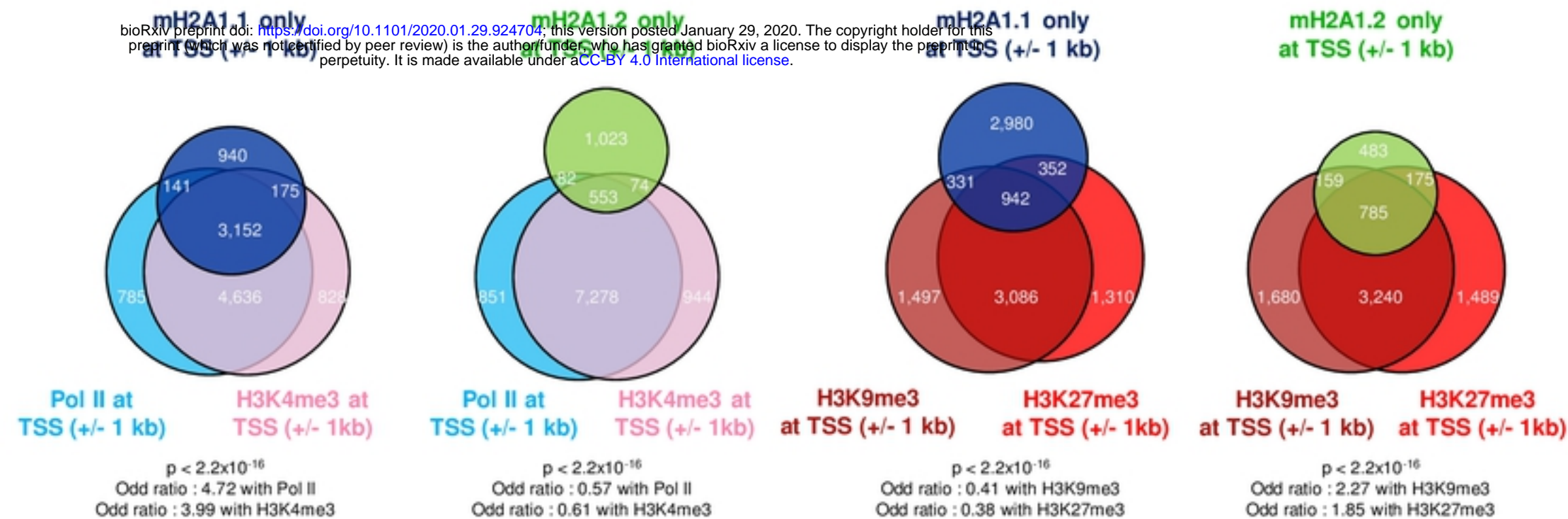
5 : Very low ratio of H3K27me3/H3K9me3

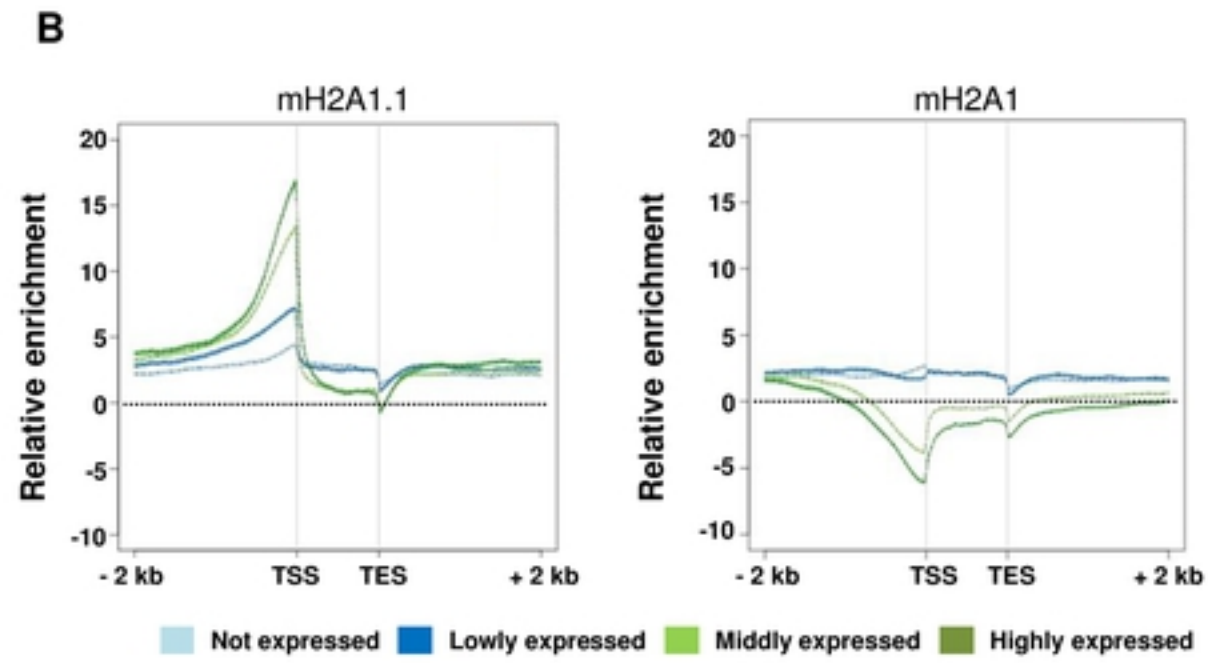
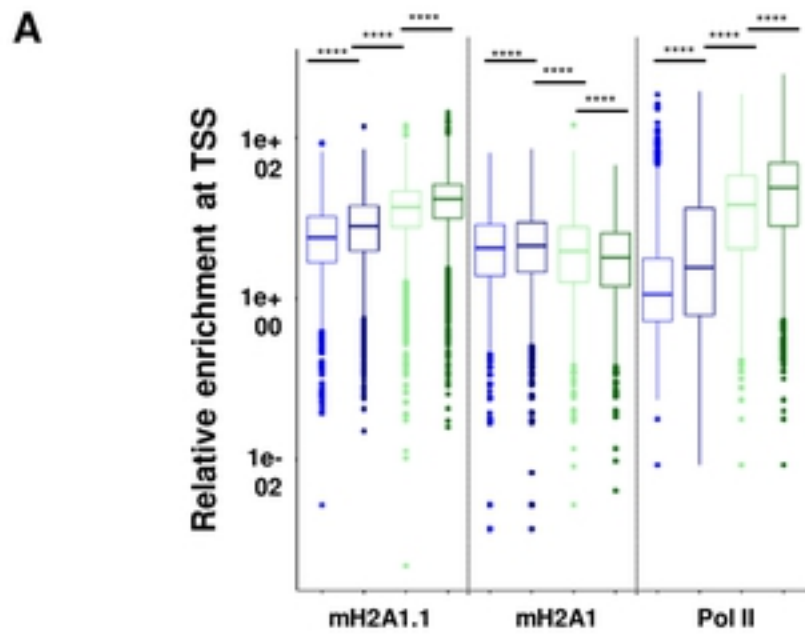


B

bioRxiv preprint doi: <https://doi.org/10.1101/2020.01.29.924704>; this version posted January 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

C

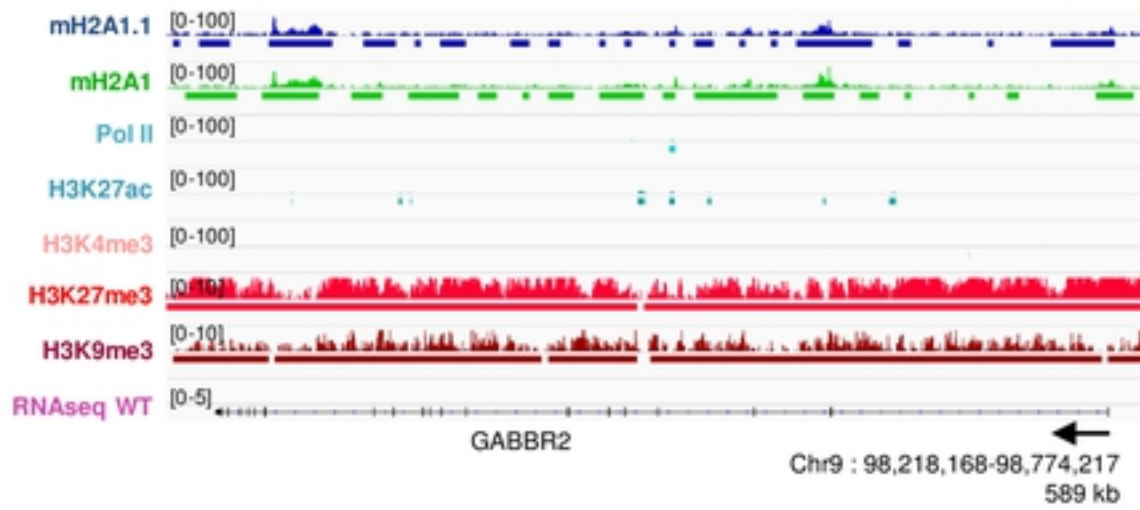




C

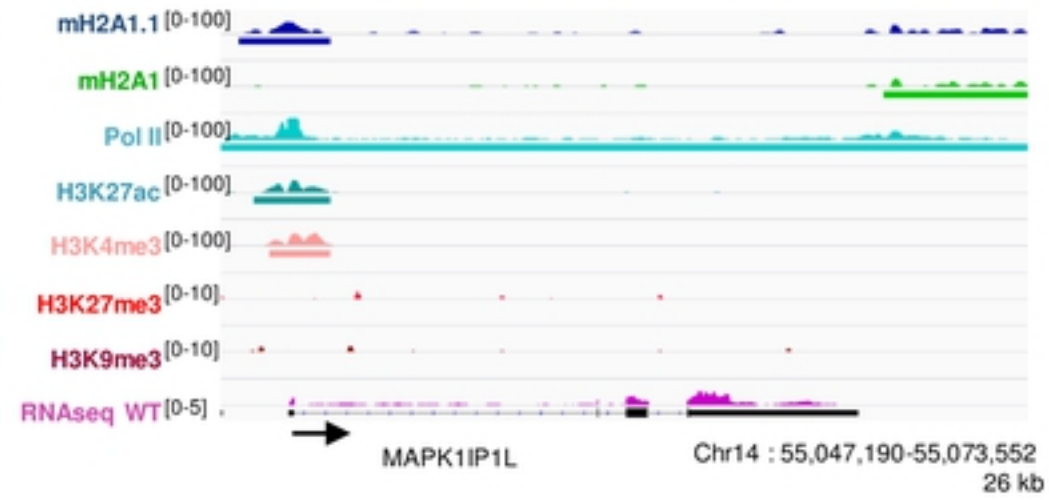
bioRxiv preprint doi: <https://doi.org/10.1101/2020.01.29.924704>; this version posted January 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

▪ Not expressed gene

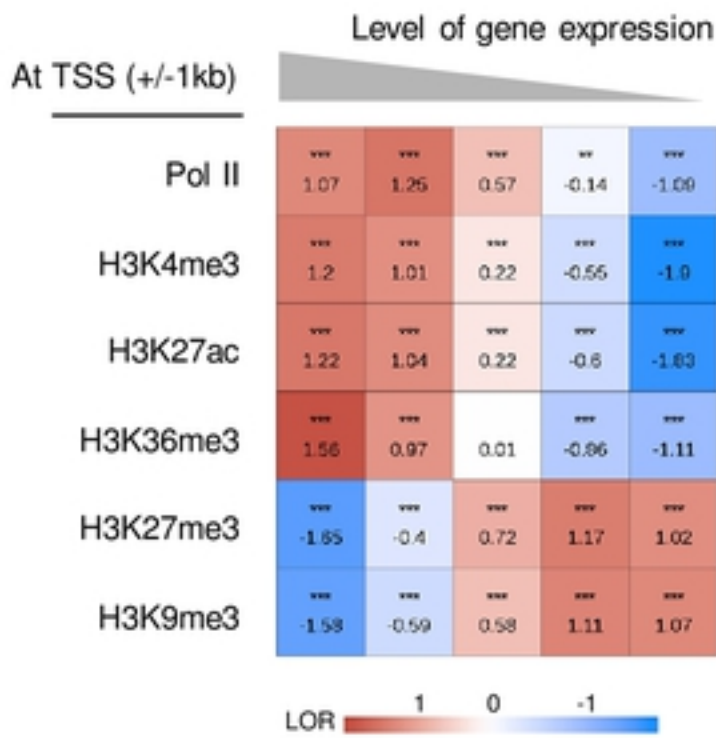


D

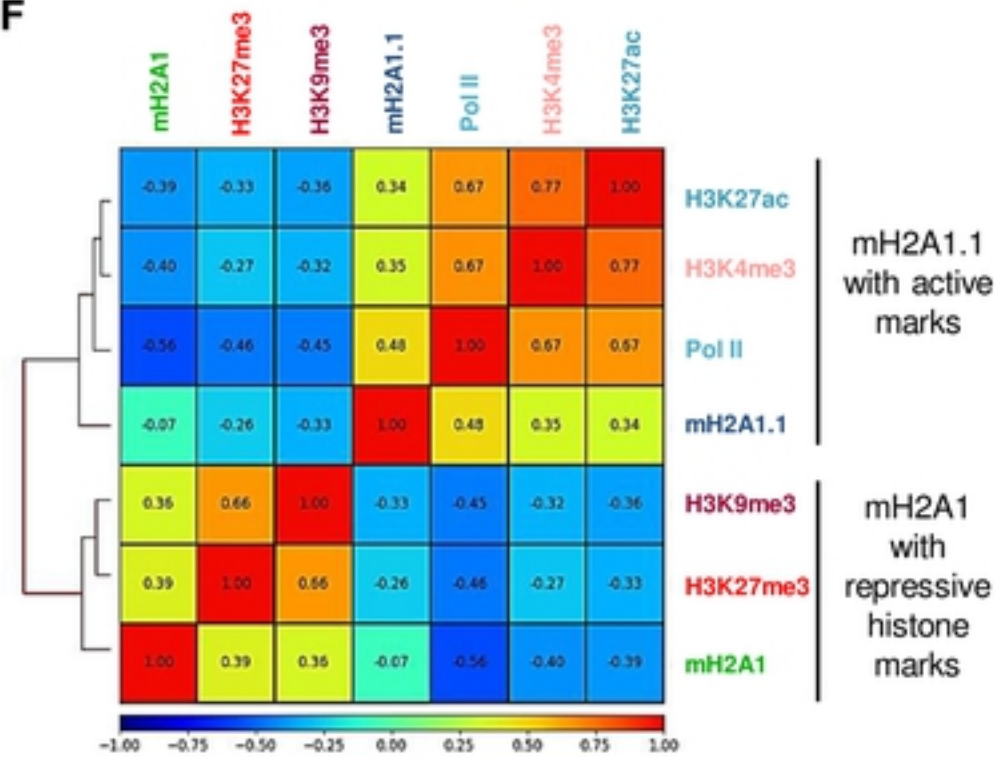
▪ Highly expressed gene



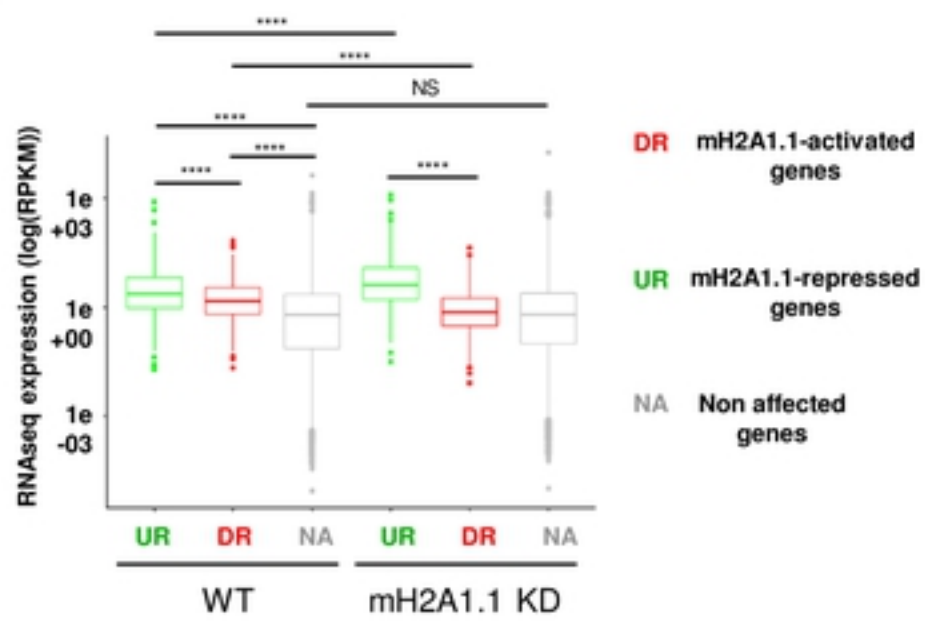
E



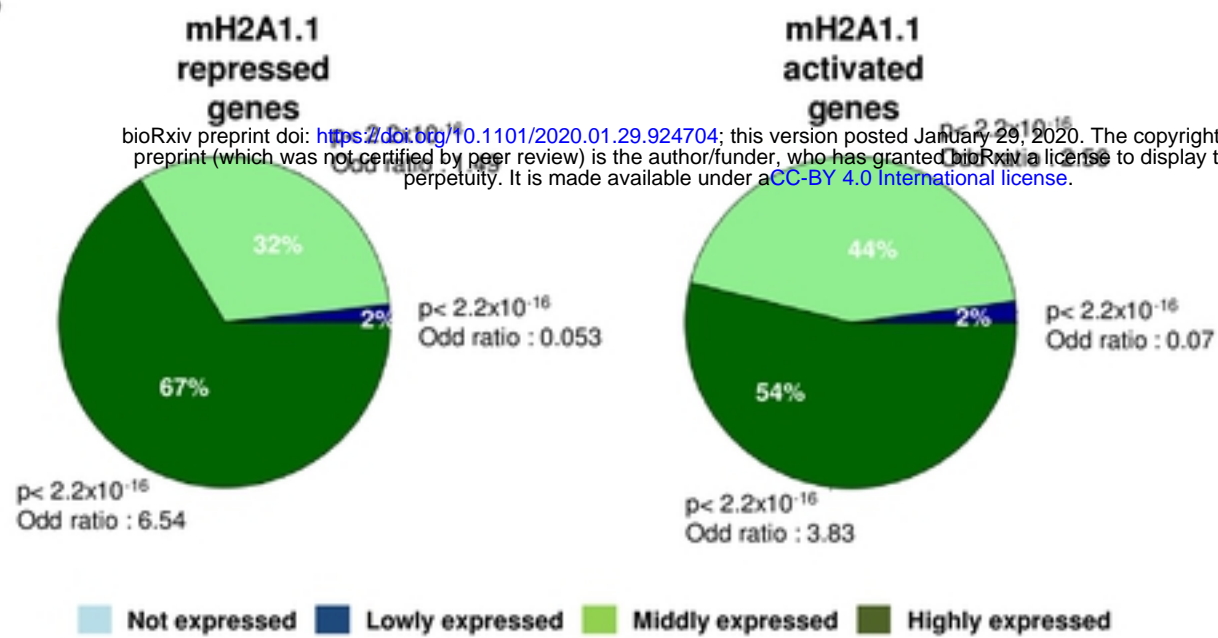
F



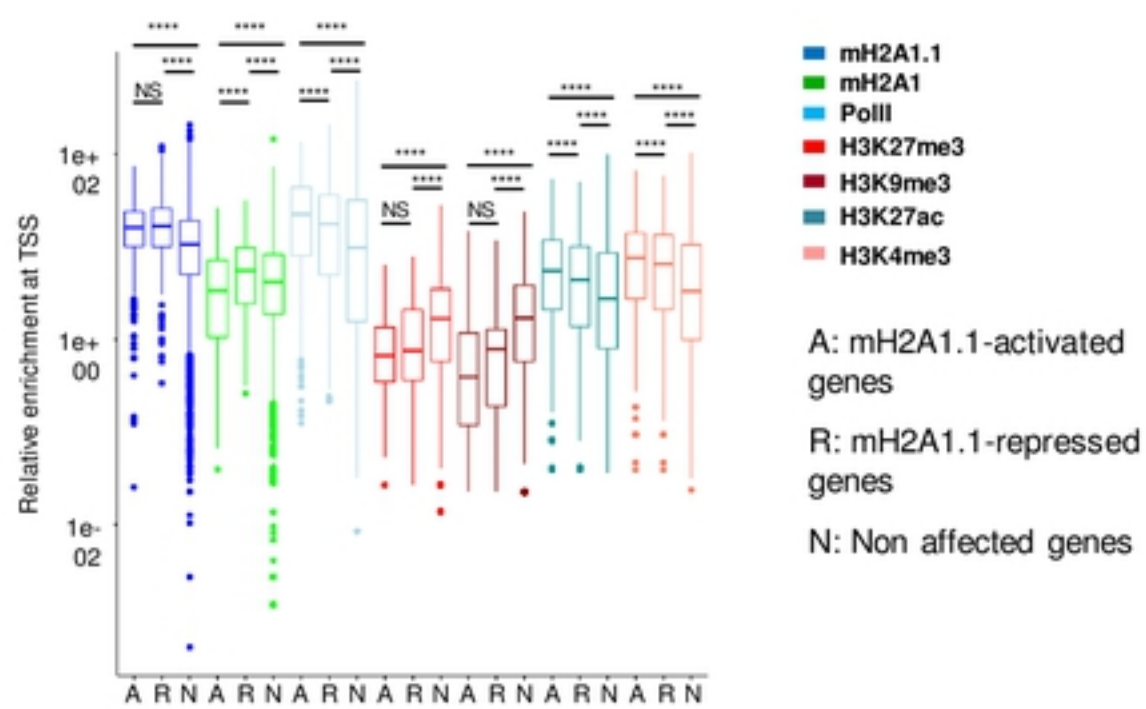
A



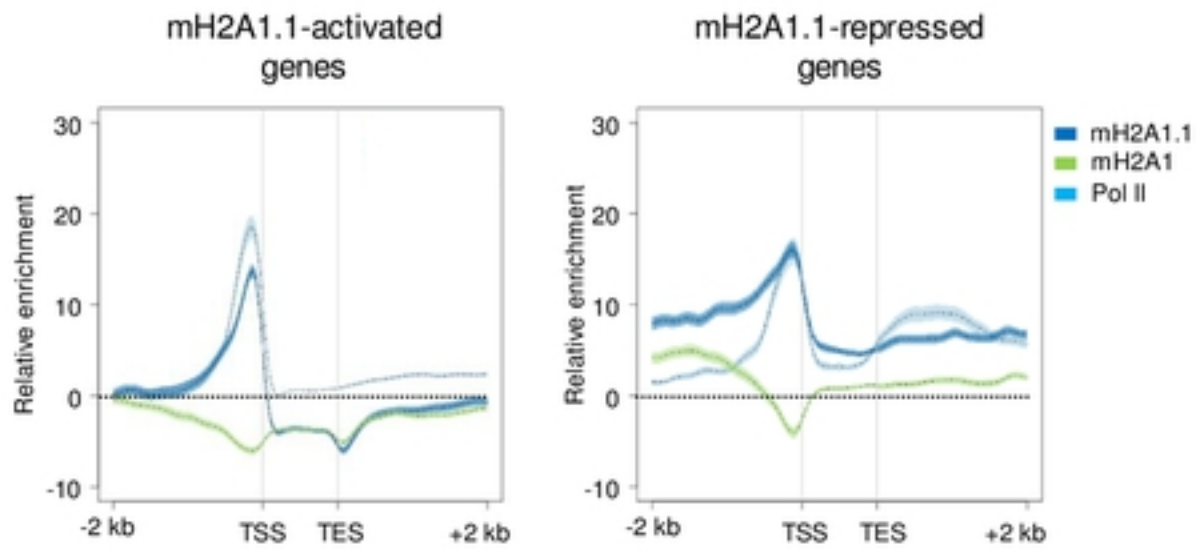
B



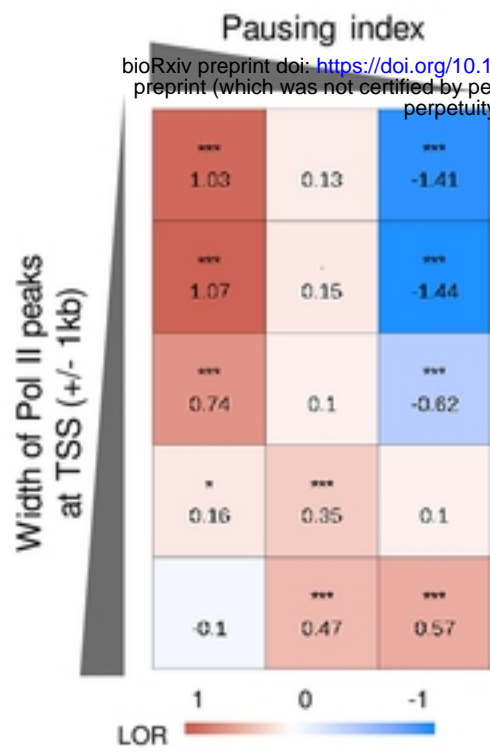
C



A



B



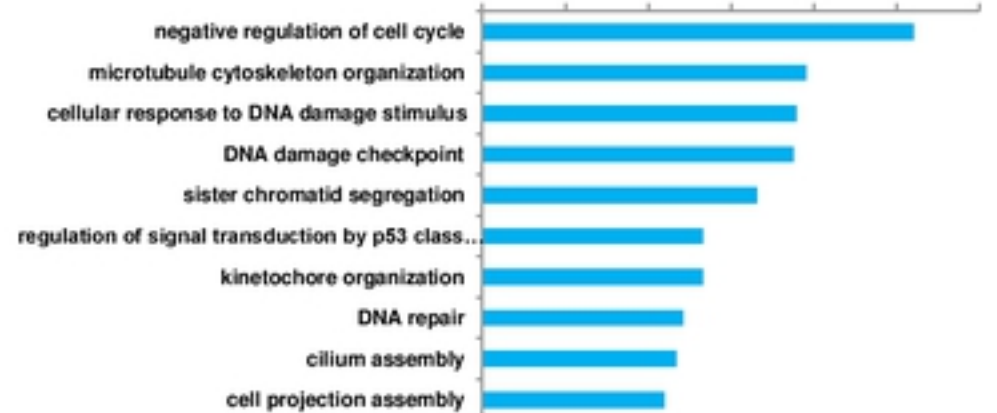
A

mH2A1.1-activated genes

Biological process

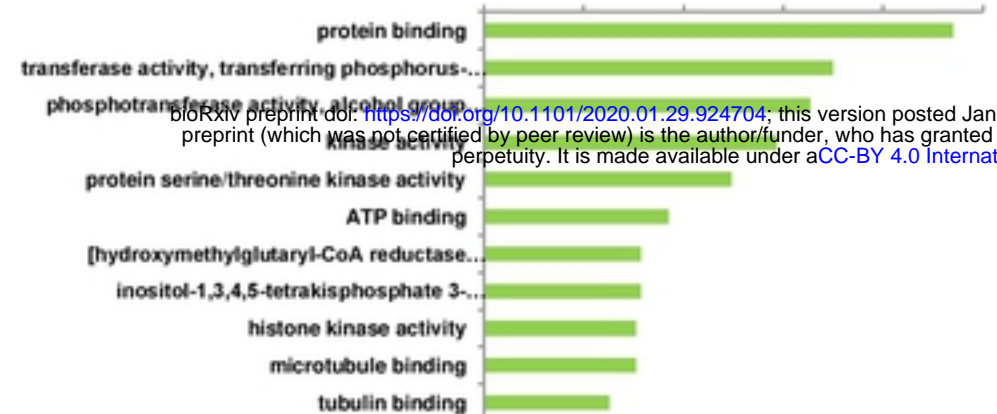
-log(pVal)

0 2 4 6 8 10 12



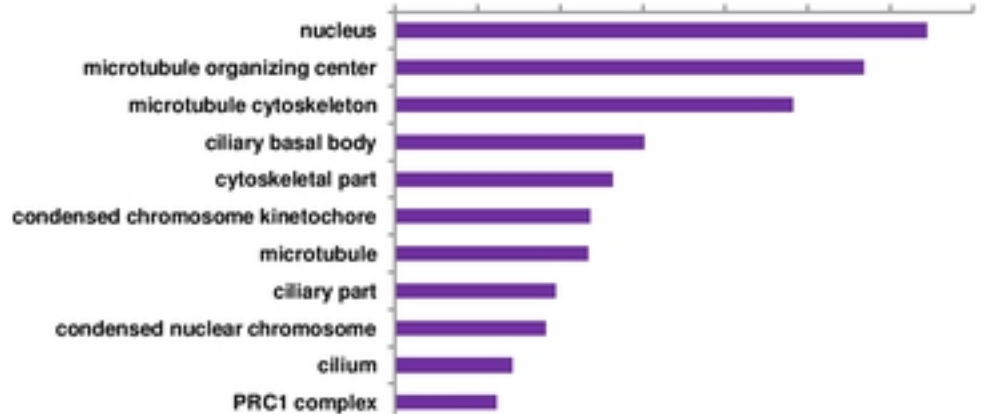
Molecular function

0 2 4 6 8 10



Cellular component

0 2 4 6 8 10 12 14



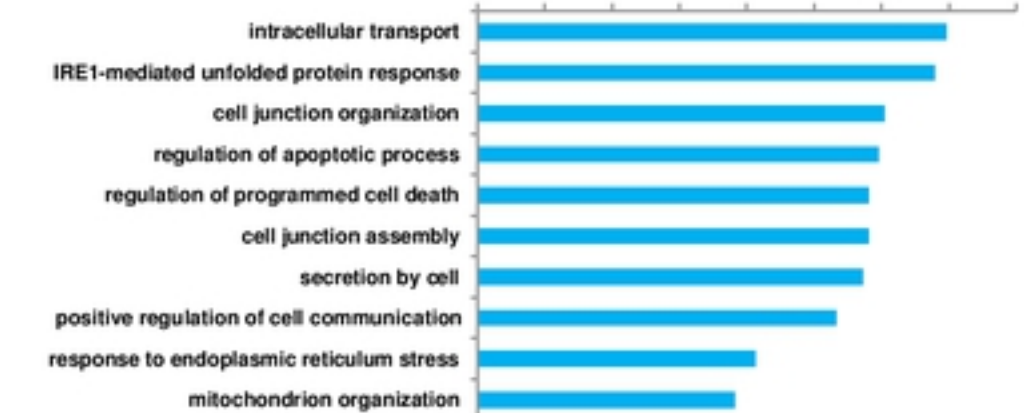
B

mH2A1.1-repressed genes

Biological process

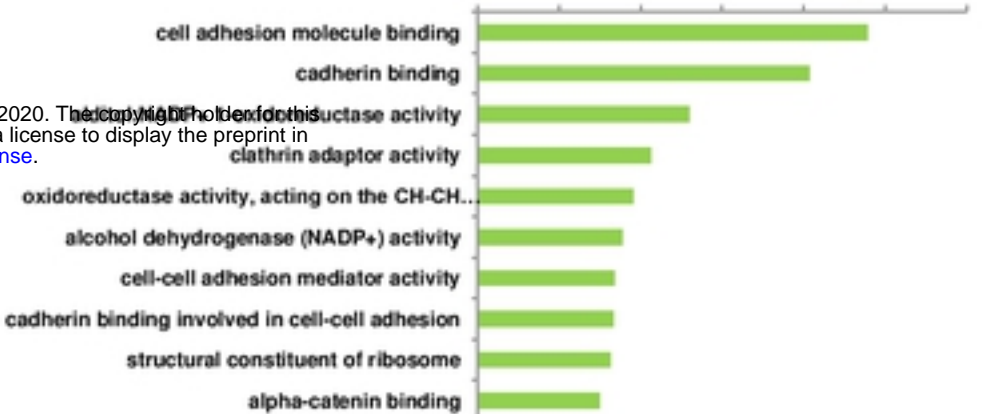
-log(pVal)

0 1 2 3 4 5 6 7 8



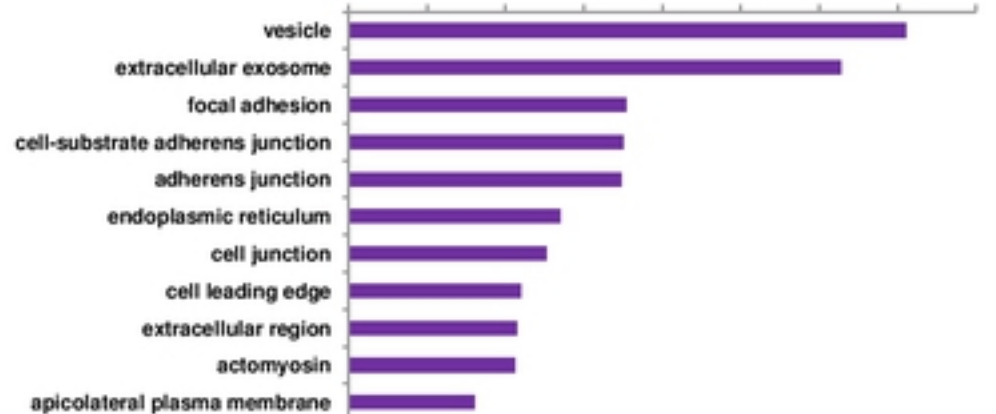
Molecular function

0 2 4 6 8 10 12



Cellular component

0 2 4 6 8 10 12 14 16



bioRxiv preprint doi: <https://doi.org/10.1101/2020.01.29.924704>; this version posted January 29, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.