

10 December 2024

Nucleosome dynamics render heterochromatin accessible in living human cells

Hemant K. Prajapati ^{1*}, Zhuwei Xu*, Peter R. Eriksson* and David J. Clark ¹

Division of Developmental Biology, Eunice Kennedy-Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda MD 20892, USA.

*These authors contributed equally to this work.

1. Corresponding authors: Building 6A Room 2A02, 6 Center Dr, Bethesda MD 20892.
DJC: Tel. (301) 496-6966. clarkda@mail.nih.gov; HKP: hemant.prajapati@nih.gov

ABSTRACT

The eukaryotic genome is packaged into chromatin, which is composed of a nucleosomal filament that coils up to form more compact structures. Chromatin exists in two main forms: euchromatin, which is relatively decondensed and enriched in transcriptionally active genes, and heterochromatin, which is condensed and transcriptionally repressed ¹⁻¹⁰. It is widely accepted that chromatin architecture modulates DNA accessibility, restricting the access of sequence-specific, gene-regulatory, transcription factors to the genome. Here, we measure genome accessibility at all GATC sites in living human MCF7 and MCF10A cells, using an adenovirus vector to express the sequence-specific *dam* DNA adenine methyltransferase. We find that the human genome is globally accessible in living cells, unlike in isolated nuclei. Active promoters are methylated somewhat faster than gene bodies and inactive promoters. Remarkably, both constitutive and facultative heterochromatic sites are methylated only marginally more slowly than euchromatic sites. In contrast, sites in centromeric chromatin are methylated slowly and are partly inaccessible. We conclude that *all* nucleosomes in euchromatin and heterochromatin are highly dynamic in living cells, whereas nucleosomes in centromeric α -satellite chromatin are static. A dynamic architecture implies that simple occlusion of transcription factor binding sites by chromatin is unlikely to be critical for gene regulation.

Chromatin consists of repeating units called nucleosomes, which contain ~147 bp of DNA coiled around a central histone octamer core. The octamer is composed of two molecules each of H2A, H2B, H3 and H4 ¹¹. Nucleosomes are regularly spaced on the DNA, resembling beads on a string. This nucleosomal filament undergoes additional compaction to form euchromatin or heterochromatin. In mammalian cells, heterochromatin may be constitutive, involving the same genomic regions in all cells (typically gene-poor regions composed of short repeated sequences), such as pericentromeric and telomeric regions, or facultative, involving cell type-specific fully repressed genes ¹⁻¹⁰. Heterochromatin is associated with specific post-translational histone modifications: constitutive heterochromatin is marked by H3K9me2/3, whereas facultative heterochromatin is marked by H3K27me3, although there is some overlap ¹². Conversely, other histone marks, such as H3K4me1, H3K4me3, H3K36me3 and H3K27ac, are generally associated with euchromatin.

47 The highly condensed nature of heterochromatin suggests that access to the DNA may be
48 limited or even prevented. However, large proteins and dextrans can penetrate heterochromatin
49 domains to some extent when injected into living cells, suggesting that heterochromatin may be
50 accessible¹³. Furthermore, heterochromatin protein 1 (HP1), which binds to H3K9me3 in
51 constitutive heterochromatin, is mobile in living mammalian cells^{14,15} and transcription of repeat
52 sequences in constitutive heterochromatin occurs at low levels^{6,16}. These data indicate that
53 constitutive heterochromatin is at least partially accessible some of the time. Liquid-liquid phase
54 separation may also be important in constitutive heterochromatin, resulting in exclusion of specific
55 proteins from the heterochromatin phase^{9,17-19}. These studies have led to a more nuanced view
56 concerning the accessibility of constitutive heterochromatin.

57
58 Facultative heterochromatin contains inactive genes that are subject to Polycomb-mediated
59 repression and are marked by H3K27me3 (reviewed by¹⁰). Genome-wide MNase-seq and ATAC-
60 seq studies on isolated nuclei from various organisms have shown that inactive genes lack
61 nucleosome-depleted regions (NDRs) at their promoters, unlike active genes. This observation
62 suggests that nucleosomes prevent transcription factor binding at regulatory elements, such as
63 promoters and enhancers, resulting in repression²⁰⁻²⁴. However, inactive promoters are partially
64 accessible in mouse liver cell nuclei²⁵. Although most transcription factors cannot access their
65 cognate binding sites when incorporated into a nucleosome²⁰, there is a class of transcription
66 factor, the "pioneer" factors, which bind to a nucleosomal site with high affinity²⁶. Pioneer factors
67 may be critical for initiating the process of nucleosome removal from regulatory elements by
68 facilitating the binding of other transcription factors and recruitment of ATP-dependent chromatin
69 remodelers to remove or displace blocking nucleosomes²⁷⁻²⁹.

70
71 These observations suggest that nucleosomes play a crucial role in gene regulation by
72 controlling access to regulatory elements. However, they are based primarily on experiments with
73 nuclei, which may not be representative of chromatin in living cells. Indeed, we have shown
74 recently that the budding yeast genome is globally accessible in living cells, except for the point
75 centromeres and the silenced loci³⁰. However, budding yeast chromatin is virtually all
76 euchromatin, and lacks heterochromatin resembling that found in higher eukaryotes. Here, we have
77 asked whether human euchromatin is generally accessible in living cells, like that of yeast, and
78 whether human heterochromatin is inaccessible, as might be expected. Surprisingly, we find that
79 both euchromatin and heterochromatin are generally accessible at the nucleosomal level in living
80 cells, and that only particular centromeric regions have limited accessibility in vivo.

81 82 **Global accessibility in live human cells**

83 We adapted our qDA-seq method to measure genome accessibility in human cells^{25,30}.
84 Specifically, we used *E. coli* Dam methyltransferase (Dam) as a probe for the accessibility of GATC
85 sites in chromatin. Dam methylates the 'A' in GATC to 'm⁶A'. It is challenging to express Dam
86 without leaky expression using inducible promoter systems³¹. Therefore, we employed an
87 adenovirus vector to transduce Dam fused to GFP and three HA tags into MCF7 cells (a human
88 breast cancer cell line). Following transduction, cells were collected at various time points to
89 monitor the kinetics of genome methylation (Fig. 1a) and of Dam production (Fig. 1b). Genomic
90 DNA was purified and digested with DpnI, a restriction enzyme that cuts at GATC only if the 'A'
91 is methylated on both strands. Agarose gel analysis of the extent of DpnI digestion revealed that
92 the genome became almost fully methylated over time (Fig. 1c; compare with purified control

93 unmethylated MCF7 DNA completely digested with MboI, which cuts at unmethylated GATC
94 sites). Thus, Dam can access a large fraction of the human genome.

95
96 The conclusion that most of the genome is accessible is further supported by genomic
97 analysis of GATC sites (Fig. 1d). The DpnI-digested DNA was sonicated to small fragments and
98 subjected to paired-end sequencing. For each GATC site, the fraction methylated is calculated as
99 the number of left or right DNA fragment ends divided by the coverage of that GATC site (Fig.
100 1a; see Methods). We calculated the methylation kinetics for each of ~5.9 million GATC sites in
101 the human genome. To visualise the data for all GATC sites, we plotted the fraction methylated for
102 the median GATC site, and for all sites within the 5%-95% methylated range, as a function of time
103 after transduction (Fig. 1d). The median GATC site was ~80% methylated after 72 h and still
104 trending upwards; 90% of all GATC sites show a similar trend.

105
106 We compared the methylation kinetics of transcriptionally active and inactive genes.
107 Analysis of published ATAC-seq data for MCF7 cells ³² (Extended Data Fig. 1a) delineated two
108 gene classes: one with high ATAC signal at the promoter, indicating the presence of an NDR, and
109 one with low or no ATAC signal, indicating the absence of an NDR (Extended Data Fig. 1a). To
110 confirm this interpretation, we performed MNase-seq on MCF7 nuclei and sorted the genes
111 according to ATAC signal (Extended Data Fig. 1a). We observed a clear correlation between ATAC
112 signal and the presence of a promoter NDR. Published gene expression (RNA-seq) data for MCF7
113 cells ³² also correlate with ATAC signal (Extended Data Fig. 1a).

114
115 We examined GATC site methylation at active and inactive promoters by plotting the mean
116 GATC site methylation as a function of distance from the transcription start site (TSS). After 12 h
117 of transduction, active genes show a weak nucleosome phasing signal that is exactly out of phase
118 with our nucleosome dyad data (MNase-seq) for MCF7 nuclei (grey profile) (Fig. 1e; Extended
119 Data Fig. 1b). This suggests that Dam methylation of the linkers and in promoter NDRs is slightly
120 faster than methylation within the first (+1) and second (+2) nucleosomes. Mean methylation
121 increases with time, reaching ~90% by 72 h. Promoter NDRs are methylated faster than gene
122 bodies (Fig. 1e). Inactive genes show the same trend, but are methylated slightly more slowly,
123 reaching ~83% after 72 h (Fig. 1f; Extended Data Fig. 1b). The methylation rate is almost uniform
124 across inactive promoter regions, with no phasing; promoters and gene bodies are methylated at
125 almost the same rate, consistent with the MNase-seq data (Fig. 1f). Nucleosome positioning
126 appears to be essentially random around inactive promoters. We conclude that both active and
127 inactive promoter regions are almost entirely accessible to Dam *in vivo*.

128
129 To determine whether higher transcription renders genes more accessible, we divided the
130 active genes into quintiles according to their mRNA levels in MCF7 cells ³². Quintile 5 has the
131 most active genes; inactive genes were placed in a separate group. We observed a small but
132 reproducible trend of increasing median methylation with increasing transcriptional activity; the
133 most active genes were methylated marginally faster than the least active genes (Fig. 1g; Extended
134 Data Fig. 1c). Thus, higher transcription correlates with a modest increase in methylation rate.
135 However, the overarching conclusion is that both active and inactive genes are accessible to Dam.

136
137
138

139 **Slow, limited methylation at centromeres**

140 To measure the accessibility of other genomic regions, we compared median methylation
141 rates for GATC sites in promoters, gene bodies, tRNA genes, enhancers, CpG islands, silencers,
142 replication origins and centromeres (Fig. 1h) using the hg38 genome annotations³³. All regions
143 are methylated at similar rates and to high levels, similar to gene bodies, except for the tRNA
144 genes, which are methylated even faster, and the centromeres, which are methylated much more
145 slowly and appear to be reaching a limit (Fig. 1h; Extended Data Fig. 1d,e). We quantified the
146 median methylation rates for the various regions relative to the median for all genomic GATC sites
147 by plotting the log of the unmethylated fraction as a function of time after transduction (Extended
148 Data Fig. 1f). Relative methylation rates for promoters, enhancers, silencers and CpG islands were
149 all slightly faster (1.2 to 1.7x; see replicates) than gene bodies and replication origins (1.0 to 1.1x).
150 tRNA genes were methylated ~2x faster, whereas centromeres were methylated ~3x more slowly
151 (0.3/0.4x the genomic median). Thus, the range in median methylation rate is ~8-fold, from
152 centromeres (slowest) to tRNA genes (fastest). We conclude that all genomic regions examined
153 are fully accessible, except for the centromeres.

154
155 To search for large regions of relatively inaccessible chromatin at the chromosome level,
156 we constructed a heat map showing the mean methylation rate for each 100 kb window along each
157 chromosome. The T2T (Telomere to Telomere) human genome was used for this analysis because
158 the centromeric regions have been thoroughly annotated³⁴. The rate was calculated as above, using
159 the mean methylation for all GATC sites in each window for each time point. We found that
160 windows of slow methylation tend to cluster predominantly at a few specific regions in each
161 chromosome (Fig. 2a). Most of these windows are situated within centromeres (Fig. 2a; red
162 rectangles indicate centromeres). The T2T genome has revealed the complexity of human
163 centromeres, which make up ~6% of the genome and include many different repeats, including α -
164 satellite repeats, human satellites (HSat1 to 5), β -satellite (β Sat), γ -satellite (γ Sat), as well as non-
165 satellite DNA³⁵ (Fig. 2b). The α -satellite repeats comprise variants of a ~171 bp repeat, which can
166 be divided into 20 supra-chromosomal families. Active α -satellite repeats are enriched in
167 centromeric histone H3 (CENP-A) and associate with the kinetochore. All of these centromeric
168 regions are methylated more slowly (0.4x to 0.7x) than the genomic average (1.0) (Fig. 2c, d); α -
169 satellite and HSats exhibit the slowest methylation rates (Fig. 2d; Extended Data Fig. 2). We
170 examined the methylation rates within the CENP-A-enriched active α -satellite SF1, SF2, SF3 and
171 SF01 supra-chromosomal families (Fig. 2e,f). The SF1 family has the slowest methylation rate
172 (0.2x) and reaches a limit at < 30% median methylation, indicating that a GATC site located in
173 SF1 α -satellite DNA is inaccessible in a large fraction of cells. In summary, centromeric GATC
174 sites are methylated more slowly than other genomic regions in living cells (Fig. 2f). This is
175 particularly true of the active α -satellite repeats (especially the SF1 family), which are associated
176 with CENP-A-containing nucleosomes, and are only partially accessible in vivo.

177 178 **TAD boundary chromatin is accessible**

179 The loop organisation of chromatin depends on the CTCF transcription factor/insulator-
180 binding protein and cohesin, which together define TAD boundaries^{36,37}. Consequently, CTCF
181 binding is expected to be quite stable, which is consistent with the exceptionally good nucleosome
182 phasing observed on both sides of CTCF binding sites (Fu et al., 2008; Wiechens et al., 2016). We
183 examined methylation patterns at and around CTCF sites (Fig. 1i). We detected very strong phasing
184 around CTCF motifs in our Dam data, which is out of phase with nucleosome dyads, as expected

185 (cf. the MNase-seq dyad plot, grey profile). Although phasing is strong, the nucleosomal DNA is
186 still accessible to Dam, since the mean methylation level increases with time, reaching ~85% after
187 72 h (Fig. 1i; Extended Data Fig. 1g). The plots also imply that the CTCF site is similarly accessible
188 to Dam, since methylation increases to high levels at the motif, even though the motif is in a small
189 trough in the methylation profiles (Fig. 1i). However, we note that the CTCF motif itself does not
190 contain a GATC site; it is therefore unclear whether it is protected from methylation.

191

192 **Heterochromatin is accessible in live cells**

193 We asked whether heterochromatin, as defined by specific histone marks, is accessible in
194 living cells using published ChIP-seq data for MCF7 cells³⁸. We grouped all GATC sites located
195 in H3K9me3 peaks (constitutive heterochromatin) or in H3K27me3 peaks (facultative
196 heterochromatin) and compared their methylation with GATC sites associated with euchromatin
197 marks (H3K4me1, H3K4me3, H3K27ac and H3K36me3). We observed that GATC sites in
198 euchromatin are methylated faster than those associated with heterochromatin (Fig. 3a; Extended
199 Data Fig. 3a). However, the rate difference is no more than ~2-fold: 1.2x - 1.6x the genomic
200 average for euchromatin; 0.8x and 0.9x for the two heterochromatic states (Fig. 3a; Extended Data
201 Fig. 3a). Most importantly, these heterochromatic regions are fully accessible to Dam.

202

203 In a more sophisticated approach, we compared methylation rates in euchromatin and
204 heterochromatin using a 15-state epigenetic ChromHMM model³⁹, which we derived using the
205 same MCF7 ChIP-seq data. ChromHMM models identify genomic regions associated with
206 various combinations of histone marks. Our ChromHMM model identifies 11 euchromatin states
207 based on the presence of H3K4me1, H3K4me3, H3K27ac and/or H3K36me3 (Fig. 3b). These are:
208 transcription start sites (TSS; state 1), TSS-flanking regions (states 2, 3 and 4), transcriptionally
209 active regions (state 5), weakly active regions (state 6), four types of enhancer (states 7, 8, 9 and
210 10), and regions with low levels of H3K27ac (state 11). Our model also defines two
211 heterochromatic states: constitutive (H3K9me3; state 13) and polycomb-repressed (facultative)
212 (H3K27me3; state 14). Some chromatin is in a bivalent state, characterised by the presence of both
213 the active H3K36me3 mark and the inactive H3K9me3 mark (state 12). State 15 has none of the
214 histone marks for which we have data and accounts for ~56% of the MCF7 genome. We plotted
215 the methylation kinetics for the active and repressed chromatin states separately for ease of
216 comparison (Fig. 3c,d; Extended Data Fig. 3b). The repressed states are methylated more slowly
217 than the active states, but importantly, even the repressed heterochromatin states trend toward
218 complete methylation (compare Figs. 3c and 3d).

219

220 With the exception of state 1 (TSS), the actual methylation rate differences are not large,
221 ranging from ~1.7x the median genomic rate for most of the euchromatin states to ~0.8x the
222 median genomic rate for the heterochromatin states (compare Figs. 3c and 3d). The relative
223 methylation rate for GATC sites in state 1 (H3K4me3 and H3K27ac) is relatively high (2.4x and
224 4.6x for biological replicates), consistent with the inclusion of NDRs in this state (Fig. 3c,d;
225 Extended Data Fig. 3b). We conclude that GATC sites located in heterochromatin are accessible
226 in living cells.

227

228 **Accessibility is not due to replication**

229 We considered the possibility that global genome accessibility in living MCF7 cells might
230 be due to DNA replication. It proved technically challenging to synchronise and maintain MCF7

231 cells in the G1 phase of the cell cycle prior to transduction. We were also unable to obtain fully
232 confluent MCF7 cells. Consequently, to confirm our findings more broadly and to test the possible
233 role of replication, we performed the same experiment using MCF10A cells, a normal human
234 breast epithelial cell line. Time course experiments after transduction of MCF10A cells with the
235 same Dam-expressing adenovirus produced similar results to those obtained for MCF7 cells
236 (Extended Data Fig. 4). Thus, the genome is also globally accessible in normal MCF10A cells,
237 suggesting that this accessibility is not due to the cancerous nature of MCF7 cells. We grew
238 MCF10A cells to confluence, when the cells cease to replicate their DNA, and then repeated the
239 transduction time course (Extended Data Fig. 5). We observed similarly high accessibility in these
240 arrested cells. We conclude that DNA replication is not a major contributor to genome accessibility.

241

242 **The X-chromosome is methylated slowly**

243 Dosage compensation results in inactivation of one of the two copies of the X-chromosome
244 in female cells and its condensation into heterochromatin (the 'Barr body'; reviewed by ⁴⁰). We
245 reasoned that the active X-chromosome would be methylated faster than the inactive X, predicting
246 an intermediate average methylation rate relative to the autosomes, because our method cannot
247 distinguish between the two copies. Since MCF7 cells have aberrant ploidy, whereas MCF10A
248 cells have normal ploidy, we focused our analysis on MCF10A cells. Using the average
249 methylation rate data for 100 kb chromosome windows (Extended Data Fig. 4i), we plotted the
250 fraction of windows with a given relative rate for each of the 23 chromosomes (Extended Data
251 Figs. 4l, 5k). This approach separates out the centromeric chromatin regions, which are methylated
252 more slowly in all chromosomes. We observed that the mean relative methylation rate of the most
253 common autosomal 100 kb window is 1.05, whereas that of the X-chromosome is 0.85. This result
254 is consistent with slower methylation of the inactive X-chromosome due to its heterochromatic
255 nature.

256

257 **Limited accessibility in isolated nuclei**

258 It has been shown previously that genome accessibility is limited in nuclei isolated from
259 both yeast and mouse liver cells ^{25,30,41}. To determine whether this is also true for MCF7, we treated
260 isolated MCF7 nuclei with increasing concentrations of purified Dam enzyme. After a 30 min
261 incubation at 37°C, genomic DNA was purified and digested with DpnI (Fig. 4a). In comparison
262 with the fully digested unmethylated control DNA ('MboI' lane), nuclei samples showed
263 incomplete DpnI digestion even at the highest Dam concentration. A clear nucleosome ladder
264 pattern is observed in all of the Dam-treated samples, suggesting that Dam methylates linker DNA,
265 but not nucleosomal DNA in nuclei. This result contrasts with the almost complete methylation
266 observed in living cells (cf. Fig. 1c).

267

268 Genomic analysis confirmed that methylation is limited in nuclei (Fig. 4b; Extended Data
269 Fig. 6). The median of all genomic GATC sites reached a plateau at ~38% methylation, indicating
270 that the median GATC site is accessible in ~38% of nuclei and inaccessible in the remaining ~62%
271 of nuclei. Examination of Dam methylation around the TSS revealed that active genes display
272 improved nucleosome phasing relative to living cells (compare Fig. 4c with Fig. 1e). Methylation
273 of active genes in nuclei reaches a limit at ~45% in the regions flanking the promoter NDR and a
274 limit of ~67% in the promoter NDR (Fig. 4c). In contrast, methylation of inactive genes in nuclei
275 is uniformly limited to ~45% over the entire region, including the inactive promoters, which are
276 not nucleosome-depleted (Fig. 4d; MNase-seq data: grey profile). Methylation is also limited

277 around CTCF motifs in nuclei (Extended Data Fig. 6d). The NDR associated with CTCF motifs is
278 ~70% accessible in nuclei and flanked by well-phased nucleosomes. Euchromatin and
279 heterochromatin regions have very similar accessibilities in nuclei (Extended Data Fig. 3c).
280 Application of our ChromHMM model to the nuclei data revealed that the active states reach a
281 limit methylation of 40% to 50%, except for the TSS state which reaches ~60% (Extended Data
282 Fig. 3d). The TSS state is higher because it includes promoter NDRs. All of the inactive states
283 show almost identical limit median methylation, at 35%-45%, which is slightly lower than the
284 median limit for the active states. We conclude that accessibility is severely limited in nuclei,
285 unlike in living cells.

286
287 Analysis of methylation at the various types of centromeric repeat in nuclei indicated that
288 the limit median methylation ranged from ~45% for non-satellite and other satellite repeats, similar
289 to that observed for gene bodies in nuclei, down to ~25%-35% for HSat1 and active and inactive
290 α -satellites (Extended Data Fig. 7a,b). The active α -satellite SF2, SF3 and SF01 families were
291 methylated to 25%-30% maximum (Extended Data Fig. 7c,d). Furthermore, the α -satellite SF1
292 family reached a limit methylation at only ~15%, indicating that GATC sites in SF1 repeats are
293 mostly inaccessible in nuclei. The methylation kinetics of these centromeric regions in nuclei are
294 similar to those observed in living cells, which also tend toward a limit (Fig. 2c, e), although the
295 methylation levels reached at α -satellite repeats in living cells are generally higher than in nuclei.
296 Notably, the SF1 α -satellite family is the slowest methylating region observed *in vivo* (Fig. 2f) and
297 the least accessible in isolated nuclei (Extended Data Fig. 7c,d).

298

299 **Human chromatin is dynamic in live cells**

300 We have measured the accessibility of GATC sites genome-wide *in vivo*. We expected to find that
301 human euchromatin would resemble yeast chromatin in being globally accessible and this is indeed
302 the case. We proposed that yeast nucleosomes are in continuous flux in living cells, but not in
303 nuclei, where they are static³⁰. Such a flux may occur either through nucleosome removal and
304 replacement, or by sliding along the DNA, and/or through reversible conformational changes (Fig.
305 4e). It is likely that all three mechanisms occur through the agencies of multiple ATP-dependent
306 chromatin remodelers. This flux renders the underlying DNA open to methylation by Dam and, by
307 inference, to sequence-specific transcription factors.

308

309 The general accessibility of both active and inactive genes to Dam suggests that the widely
310 accepted model that inactive genes are inactive because transcription factor binding sites in their
311 promoters are blocked by nucleosomes may no longer be tenable. It seems unlikely that
312 nucleosomes present a permanent block to transcription factor binding, thus maintaining genes in
313 the repressed state. The role of pioneer factors, which have similar affinity for nucleosomal and
314 non-nucleosomal sites, is unclear, given nucleosome flux. However, pioneer factors would be
315 predicted to bind faster, since they do not have to wait for nucleosome dynamics to expose their
316 cognate sites. Pioneer factors might also be important for initiating nucleosome dynamics.
317 Alternatively, gene regulation may occur primarily through regulation of transcription factor gene
318 expression, location (e.g. retention in the cytosol) or activity (e.g. post-translational modifications
319 and allosteric effects), and through non-coding RNA expression.

320

321 We also expected that human heterochromatin might either be resistant to Dam
322 methylation, because of its highly condensed state, or that it might be similar to chromatin in

323 nuclei, with immobile nucleosomes, such that only linkers are methylated. Instead, we observed a
324 trend towards complete methylation, albeit at a somewhat slower rate than for euchromatin. The
325 relatively slow methylation of heterochromatin may be due to a combination of various factors,
326 perhaps including slower nucleosome flux relative to euchromatin, the highly condensed nature of
327 heterochromatin, and the presence of heterochromatin proteins (e.g., HP1 or Polycomb
328 complexes). Nevertheless, both constitutive and facultative heterochromatin are generally
329 accessible in living human cells. This observation is inconsistent with models proposing that
330 heterochromatin condensation prevents access to the DNA it contains, resulting in gene repression
331 (discussed by ⁷). Our data show that heterochromatic DNA is generally accessible and highly
332 dynamic at the nucleosomal level in live cells, unlike in isolated nuclei.

333

334 The only genomic regions displaying limited accessibility in living human cells are the
335 centromeric active α -satellite repeats. These elements are methylated slowly relative to other
336 genomic regions and, unlike those regions, reach a limit at 50%-60% methylation. The SF1 α -
337 satellite repeats are methylated even more slowly, reaching a limit at only ~30% methylation. The
338 nucleosomes in active α -satellite repeats are enriched in centromeric H3 (CENP-A) and so
339 resemble yeast centromeric nucleosomes in their resistance to methylation in vivo ³⁰. The limited
340 accessibility of centromeric active α -satellite repeats in vivo is similar to that observed for all
341 genomic regions in isolated nuclei. This observation suggests that centromeric chromatin in live
342 cells is static, not dynamic, with little or no nucleosome flux, such that linkers are methylated and
343 nucleosomal DNA is protected (Fig. 4e).

344

345 In summary, we have measured the accessibility of the human genome in living cells. We
346 find that the genome is generally accessible at the nucleosomal level, including classical
347 heterochromatin regions marked by H3K9me3 or H3K27me3. The exception is the centromeric
348 active α -satellite repeats, which exhibit limited accessibility. We propose that nucleosome flux
349 creates a genome-wide open chromatin environment, in which the DNA is packaged but still
350 accessible, facilitating the search for cognate sites by sequence-specific transcription factors.

351

352 References

353

- 354 1 Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and
355 function. *Nat Rev Mol Cell Biol* **19**, 229-244 (2018). <https://doi.org/10.1038/nrm.2017.119>
- 356 2 Penagos-Puig, A. & Furlan-Magaril, M. Heterochromatin as an Important Driver of
357 Genome Organization. *Front Cell Dev Biol* **8**, 579137 (2020).
358 <https://doi.org/10.3389/fcell.2020.579137>
- 359 3 Richards, E. J. & Elgin, S. C. Epigenetic codes for heterochromatin formation and
360 silencing: rounding up the usual suspects. *Cell* **108**, 489-500 (2002).
361 [https://doi.org/10.1016/s0092-8674\(02\)00644-x](https://doi.org/10.1016/s0092-8674(02)00644-x)
- 362 4 Woodcock, C. L. & Ghosh, R. P. Chromatin higher-order structure and dynamics. *Cold*
363 *Spring Harb Perspect Biol* **2**, a000596 (2010).
364 <https://doi.org/10.1101/cshperspect.a000596>
- 365 5 Kireev, I. *et al.* In vivo immunogold labeling confirms large-scale chromatin folding
366 motifs. *Nat Methods* **5**, 311-313 (2008). <https://doi.org/10.1038/nmeth.1196>
- 367 6 Grewal, S. I. S. The molecular basis of heterochromatin assembly and epigenetic
368 inheritance. *Mol Cell* **83**, 1767-1785 (2023). <https://doi.org/10.1016/j.molcel.2023.04.020>

- 369 7 Bell, O., Burton, A., Dean, C., Gasser, S. M. & Torres-Padilla, M. E. Heterochromatin
370 definition and function. *Nat Rev Mol Cell Biol* **24**, 691-694 (2023).
371 <https://doi.org/10.1038/s41580-023-00599-7>
- 372 8 Padeken, J., Methot, S. P. & Gasser, S. M. Establishment of H3K9-methylated
373 heterochromatin and its functions in tissue differentiation and maintenance. *Nat Rev Mol*
374 *Cell Biol* **23**, 623-640 (2022). <https://doi.org/10.1038/s41580-022-00483-w>
- 375 9 Janssen, A., Colmenares, S. U. & Karpen, G. H. Heterochromatin: Guardian of the
376 Genome. *Annu Rev Cell Dev Biol* **34**, 265-288 (2018). <https://doi.org/10.1146/annurev-cellbio-100617-062653>
- 377
378 10 Kim, J. J. & Kingston, R. E. Context-specific Polycomb mechanisms in development. *Nat*
379 *Rev Genet* **23**, 680-695 (2022). <https://doi.org/10.1038/s41576-022-00499-0>
- 380 11 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal
381 structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-260 (1997).
382 <https://doi.org/10.1038/38444>
- 383 12 Nicetto, D. & Zaret, K. S. Role of H3K9me3 heterochromatin in cell identity establishment
384 and maintenance. *Curr Opin Genet Dev* **55**, 1-10 (2019).
385 <https://doi.org/10.1016/j.gde.2019.04.013>
- 386 13 Verschure, P. J. *et al.* Condensed chromatin domains in the mammalian nucleus are
387 accessible to large macromolecules. *EMBO Rep* **4**, 861-866 (2003).
388 <https://doi.org/10.1038/sj.embor.embor922>
- 389 14 Cheutin, T. *et al.* Maintenance of stable heterochromatin domains by dynamic HP1 binding.
390 *Science* **299**, 721-725 (2003). <https://doi.org/10.1126/science.1078572>
- 391 15 Festenstein, R. *et al.* Modulation of heterochromatin protein 1 dynamics in primary
392 Mammalian cells. *Science* **299**, 719-721 (2003). <https://doi.org/10.1126/science.1078694>
- 393 16 Saksouk, N., Simboeck, E. & Dejardin, J. Constitutive heterochromatin formation and
394 transcription in mammals. *Epigenetics Chromatin* **8**, 3 (2015).
395 <https://doi.org/10.1186/1756-8935-8-3>
- 396 17 Zhang, H., Qin, W., Romero, H., Leonhardt, H. & Cardoso, M. C. Heterochromatin
397 organization and phase separation. *Nucleus* **14**, 2159142 (2023).
398 <https://doi.org/10.1080/19491034.2022.2159142>
- 399 18 Larson, A. G. & Narlikar, G. J. The Role of Phase Separation in Heterochromatin
400 Formation, Function, and Regulation. *Biochemistry* **57**, 2540-2548 (2018).
401 <https://doi.org/10.1021/acs.biochem.8b00401>
- 402 19 Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**,
403 241-245 (2017). <https://doi.org/10.1038/nature22989>
- 404 20 Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome.
405 *Nature* **562**, 76-81 (2018). <https://doi.org/10.1038/s41586-018-0549-5>
- 406 21 Kornberg, R. D. & Lorch, Y. Primary Role of the Nucleosome. *Mol Cell* **79**, 371-375
407 (2020). <https://doi.org/10.1016/j.molcel.2020.07.020>
- 408 22 Isbel, L., Grand, R. S. & Schubeler, D. Generating specificity in genome regulation through
409 transcription factor sensitivity to chromatin. *Nat Rev Genet* **23**, 728-740 (2022).
410 <https://doi.org/10.1038/s41576-022-00512-6>
- 411 23 Michael, A. K. & Thoma, N. H. Reading the chromatinized genome. *Cell* **184**, 3599-3611
412 (2021). <https://doi.org/10.1016/j.cell.2021.05.029>

- 413 24 Lai, W. K. M. & Pugh, B. F. Understanding nucleosome dynamics and their links to gene
414 expression and DNA replication. *Nat Rev Mol Cell Biol* **18**, 548-562 (2017).
415 <https://doi.org/10.1038/nrm.2017.47>
- 416 25 Chereji, R. V., Eriksson, P. R., Ocampo, J., Prajapati, H. K. & Clark, D. J. Accessibility of
417 promoter DNA is not the primary determinant of chromatin-mediated gene regulation.
418 *Genome Res* **29**, 1985-1995 (2019). <https://doi.org/10.1101/gr.249326.119>
- 419 26 Bulyk, M. L., Drouin, J., Harrison, M. M., Taipale, J. & Zaret, K. S. Pioneer factors - key
420 regulators of chromatin and gene expression. *Nat Rev Genet* **24**, 809-815 (2023).
421 <https://doi.org/10.1038/s41576-023-00648-z>
- 422 27 Iurlaro, M. *et al.* Mammalian SWI/SNF continuously restores local accessibility to
423 chromatin. *Nat Genet* **53**, 279-287 (2021). <https://doi.org/10.1038/s41588-020-00768-w>
- 424 28 Prajapati, H. K., Ocampo, J. & Clark, D. J. Interplay among ATP-Dependent Chromatin
425 Remodelers Determines Chromatin Organisation in Yeast. *Biology (Basel)* **9** (2020).
426 <https://doi.org/10.3390/biology9080190>
- 427 29 Clapier, C. R., Iwasa, J., Cairns, B. R. & Peterson, C. L. Mechanisms of action and
428 regulation of ATP-dependent chromatin-remodelling complexes. *Nat Rev Mol Cell Biol* **18**,
429 407-422 (2017). <https://doi.org/10.1038/nrm.2017.26>
- 430 30 Prajapati, H. K., Eriksson, P.R., Elizalde, P.A., Coey, C.T., Xu, Z., Clark, D.J. The yeast
431 genome is globally accessible in living cells. *Nat Struct Mol Biol* (2024).
432 [https://doi.org:https://doi.org/10.1038/s41594-024-01318-2](https://doi.org/https://doi.org/10.1038/s41594-024-01318-2)
- 433 31 Aughey, G. N. & Southall, T. D. Dam it's good! DamID profiling of protein-DNA
434 interactions. *Wiley Interdiscip Rev Dev Biol* **5**, 25-37 (2016).
435 <https://doi.org/10.1002/wdev.205>
- 436 32 Tian, C. *et al.* Impaired histone inheritance promotes tumor progression. *Nat Commun* **14**,
437 3429 (2023). <https://doi.org/10.1038/s41467-023-39185-y>
- 438 33 Schneider, V. A. *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies
439 demonstrates the enduring quality of the reference assembly. *Genome Res* **27**, 849-864
440 (2017). <https://doi.org/10.1101/gr.213611.116>
- 441 34 Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44-53 (2022).
442 <https://doi.org/10.1126/science.abj6987>
- 443 35 Altemose, N. *et al.* Complete genomic and epigenetic maps of human centromeres. *Science*
444 **376**, eabl4178 (2022). <https://doi.org/10.1126/science.abl4178>
- 445 36 Ghirlando, R. & Felsenfeld, G. CTCF: making the right connections. *Genes Dev* **30**, 881-
446 891 (2016). <https://doi.org/10.1101/gad.277863.116>
- 447 37 Kim, E., Barth, R. & Dekker, C. Looping the Genome with SMC Complexes. *Annu Rev*
448 *Biochem* **92**, 15-41 (2023). <https://doi.org/10.1146/annurev-biochem-032620-110506>
- 449 38 Franco, H. L. *et al.* Enhancer transcription reveals subtype-specific gene expression
450 programs controlling breast cancer pathogenesis. *Genome Res* **28**, 159-170 (2018).
451 <https://doi.org/10.1101/gr.226019.117>
- 452 39 Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with
453 ChromHMM. *Nat Protoc* **12**, 2478-2492 (2017). <https://doi.org/10.1038/nprot.2017.124>
- 454 40 Galupa, R. & Heard, E. X-Chromosome Inactivation: A Crossroads Between Chromosome
455 Architecture and Gene Regulation. *Annu Rev Genet* **52**, 535-566 (2018).
456 <https://doi.org/10.1146/annurev-genet-120116-024611>

457 41 Oberbeckmann, E. *et al.* Absolute nucleosome occupancy map for the *Saccharomyces*
458 *cerevisiae* genome. *Genome Res* **29**, 1996-2009 (2019).
459 <https://doi.org/10.1101/gr.253419.119>
460
461
462

463 **Methods**

464 **Cell culture**

466 MCF7 cells (ATCC HTB-22) were cultured in RPMI 1640 with L-glutamine (Corning 10-040-
467 CV) supplemented with 10% Fetal Bovine Serum (FBS) (Corning 35-010-CV) and 1% penicillin/
468 streptomycin (Gibco 15140-148) at 37°C with 5% CO₂/95% air in a humidified incubator.
469 MCF10A cells (ATCC CRL-10317) were cultured in DMEM/F12 (Gibco 11320-033)
470 supplemented with 5% horse serum (Gibco 26050070), 10 µg/ml insulin (Gibco 12585-014), 20
471 ng/ml epidermal growth factor (Gibco PHG0311), 500 ng/ml hydrocortisone (Sigma H0888), 100
472 ng/ml cholera toxin (Sigma C80520), 1% penicillin/streptomycin and 2 mM L-glutamine (Gibco
473 25030149). Confluent MCF10A cells were obtained by culturing in a 12-well plate for 46 h in the
474 same medium and then transduced in medium lacking all growth factors except horse serum.
475

476 **Adenovirus transduction and Dam methylation in living cells**

477 The Dam-3HA-eGFP cassette, codon-optimised for mouse, was constructed by gene synthesis
478 (Thermo Fisher GeneArt) (sequence available on request). The Dam expression cassette was
479 expressed from a CMV promoter in an adenovirus vector (human adenovirus type 5 (dE1/E3);
480 Vector Biolabs), packaged, amplified and purified by Vector Biolabs. The final viral yield was 4.2
481 x 10¹⁰ plaque-forming units (pfu) per ml, equivalent to approximately 10¹² viral particles per ml.
482 About a million cells were seeded in each well of a 6-well plate one day prior to adenovirus
483 transduction. The next day, the cells were counted and the amount of adenovirus required to
484 achieve a multiplicity of infection (MOI) of 1000 was pre-incubated at 37°C for 30 min to improve
485 transfer efficiency⁴². The virus was mixed with 300 µl per well of serum-free, antibiotic-free
486 medium (RPMI 1640 for MCF7; DMEM for MCF10A) and incubated for 5-10 min at room
487 temperature. Meanwhile, the culture medium was aspirated from the 6-well plate and 450 µl of
488 complete medium containing FBS was added. The transduction mixture was added to each well,
489 mixed by gently swirling the plate a few times, and incubated at 37°C for 4 h. Then 1.8 ml of
490 complete culture medium containing serum was added to each well.

491 Confluent MCF10A cells were treated slightly differently: a 12-well plate and an MOI of
492 2000 was used. The virus was mixed with 136 µl DMEM per well and incubated for 5-10 min at
493 room temperature. Meanwhile, the culture medium was aspirated and 204 µl of complete medium
494 containing FBS was added. The transduction mixture was added to each well, mixed and incubated
495 at 37°C for 8 h. Then 1 ml of complete culture medium containing serum was added to each well.

496 The time course started at this point, with cells harvested after 12 h, 24 h, 48 h and 72 h
497 (for MCF10A the time points were 12 h, 24 h, 36 h and 48 h). Cells were detached from the well
498 by washing with 1 ml PBS and then incubating in 0.4 ml 0.25% (w/v) trypsin-0.53 mM EDTA
499 solution (ATCC 30-2101) at 37°C for 5 min (15 min for MCF10A). Next, 1 ml medium was added
500 to the well, cells were collected by centrifugation at 100 g for 5 min, after which the medium was
501 aspirated, and the cells were resuspended in 1 ml medium. The cells were counted, divided up for
502 DNA and protein extraction, quickly frozen on dry ice, and stored at -80°C. DNA extraction was
503 performed using the PureLink Genomic DNA Mini Kit (Invitrogen) according to the

504 manufacturer's guidelines. Purified genomic DNA (1.2 - 1.5 μ g) was digested with 10 units of
505 DpnI (New England Biolabs (NEB) R0176L) in NEB CutSmart buffer for 2 h at 37°C.

506

507 **Immunoblotting**

508 A pellet containing 0.3-0.5 million cells was resuspended in 0.25 to 0.4 ml of 1x lithium dodecyl
509 sulfate buffer (Invitrogen NP0007) supplemented with 0.2 M 2-mercaptoethanol and heated for 5
510 min at 99°C; 10 μ l was loaded on to each of two 4-12% bis-Tris polyacrylamide gels (Invitrogen
511 NP0336) and run using MOPS/SDS running buffer (Invitrogen). Transfer of proteins to a
512 membrane and signal development with horseradish peroxidase-conjugated anti-HA (3F10; Roche
513 12013819001) or anti-tubulin (Abcam ab-185067) antibodies were performed as described³⁰.

514

515 **FACS analysis**

516 Propidium iodide staining and flow cytometric DNA analysis of MCF7 and MCF10A cells were
517 performed as described (Mullen (2004)). Cells (0.1-0.2 million) were resuspended in 50 μ l cold
518 buffer (250 mM sucrose, 40 mM trisodium citrate, 5% v/v DMSO) and frozen at -80°C. For FACS,
519 cells were thawed and 200 μ l of ice-cold Solution A (0.03 mg/ml trypsin, 3.4 mM trisodium citrate,
520 0.1% v/v NP-40, 1.5 mM spermine tetrahydrochloride, 0.5 mM Tris-HCl pH 7.6) was added. The
521 mixture was incubated at room temperature for 5 min. Subsequently, 100 μ l of ice-cold Solution
522 B (0.5 mg/ml trypsin inhibitor, 0.1 mg/ml RNase A, 3.4 mM trisodium citrate, 0.1% v/v NP-40,
523 1.5 mM spermine tetrahydrochloride, 0.5 mM Tris-HCl pH 7.6) was added and incubated for
524 another 5 min at room temperature. Finally, 20 μ l propidium iodide at 1 mg/ml (Invitrogen P3566)
525 was added and incubated at room temperature in the dark to prevent photobleaching. The cells
526 were analysed using a FACSCalibur flow cytometer (Becton Dickinson) and Cell Quest Pro
527 software, following the manufacturer's instructions.

528

529 **Dam methylation of isolated nuclei**

530 MCF7 cells were cultured in complete medium in a 75 cm² flask and re-passaged into a new flask
531 after 2-3 days of growth. When the cells reached approximately 80% confluency, they were
532 trypsinized and harvested. To extract nuclei, a pellet of 3 - 4 million cells was resuspended in 2 ml
533 Buffer A (15 mM Tris-HCl pH 8.0, 15 mM NaCl, 60 mM KCl, 1.5 mM EDTA, 0.5 mM spermidine,
534 15 mM β -mercaptoethanol, and protease inhibitors) with 0.03% NP-40. The mixture was gently
535 but thoroughly mixed by pipetting and incubated on ice for 10 min, inverting the tube 2 or 3 times
536 during the incubation. The lysate was centrifuged at 500 g for 2 min at 4°C, and the supernatant
537 was removed. The nuclei were washed with 1 ml Buffer A. The nuclei were resuspended in 1 ml
538 Buffer A supplemented with fresh S-adenosylmethionine to 0.5 mM, and divided into five 200 μ l
539 aliquots. Dam methyltransferase (NEB M0222B-HC2 at 40 U/ μ l; 8 μ g Dam/ml) was added to the
540 aliquots of nuclei: 0, 25, 50, 100, and 200 units (0, 0.8, 1.6, 3.1, 6.3 nM, respectively), gently
541 mixed, and incubated for 30 min at 37°C. Genomic DNA was extracted using the PureLink
542 Genomic DNA Kit (Invitrogen 2666617). Finally, 1.2 to 1.5 μ g purified genomic DNA was
543 digested with 10 units of DpnI as above.

544

545 **MNase-seq**

546 MNase (Worthington LS004798) was dissolved to 10 units/ μ l in 5 mM Na-phosphate buffer pH
547 7.0, 0.025 mM CaCl₂, aliquotted out, and stored at -80°C. MCF7 cells (3 to 4 million) were
548 resuspended in 2 ml Buffer B (15 mM Tris-HCl, pH 8.0, 15 mM NaCl, 60 mM KCl, 1 mM EDTA,
549 2 mM CaCl₂, 0.5 mM spermidine, 0.03% NP-40, 15 mM 2-mercaptoethanol and protease

550 inhibitors). The cells were gently lysed by pipetting and incubated on ice for 10 min, inverting the
551 tube 2-3 times during incubation. The lysate was centrifuged at 500 g for 2 min at 4°C and the
552 supernatant was removed. The nuclei were washed with 1 ml Buffer B without NP40 and
553 resuspended in 1.3 ml Buffer B without NP40. MNase was added to six tubes of 200 µl nuclei, as
554 follows: 12.5 U, 25 U, 50 U, 100 U, 200 U and 400 U, gently mixed, and incubated for 3 min at
555 25°C. MNase-digested DNA was purified using the PureLink Genomic DNA kit (Invitrogen
556 2666617) and analysed in an agarose gel. For accurate and even nucleosome mapping, we chose
557 digests with a dominant band at ~150 bp corresponding to >80% of the DNA (typically 25 U, 50
558 U and 100 U), prepared paired-end libraries, and performed low-coverage sequencing to identify
559 the digest with the most optimal DNA fragment length distribution⁴³. The 25 U sample (Replicate
560 1) and the 50 U sample (Replicate 2) were chosen for high-coverage sequencing.

561

562 **Library preparation for paired-end Illumina sequencing**

563 For both nuclei and live cell experiments, DpnI-digested genomic DNA was purified using 1.8 vol.
564 AMPure XP beads (Beckman). Paired-end libraries were prepared as described³⁰ except for the
565 sonication step, in which the DNA was fragmented using a Covaris ME220 ultrasonicator (350 bp
566 program, peak power, 50 W; duty factor 10%; 1,000 cycles per burst; average power 5; total time
567 170 s per tube). All sequencing was performed using an Illumina NextSeq 2000 machine.

568

569 **Computational analysis of methylated fractions**

570 We developed two packages for methylated fraction analysis: `snakemakeMethylFrac` and
571 `methylFracAnalyzer`. `SnakemakeMethylFrac`, a `snakemake` workflow⁴⁴, processes raw Illumina
572 paired-end reads to determine methylated fractions at all GATC sites. `Bowtie2` v2.5.1⁴⁵ is used
573 for alignment and `bedtools` v2.31.1⁴⁶ is used to calculate the occupancy (fragment coverage) and
574 5'-end counting. GATC sites that overlap with CpG sites were filtered out, because DpnI cannot
575 cut GATm⁵C. GATC half-sites that are within 150 bp of each other were also filtered out because
576 small DNA fragments < 150 bp tend to be lost during sample purification. The output includes
577 SQLite database and bigwig files. We use `pandas`⁴⁷, `pyBigWig`
578 (<https://github.com/deeptools/pyBigWig>), `biopython`⁴⁸, `matplotlib`⁴⁹ and `seaborn`⁵⁰ in this
579 workflow. `MethylFracAnalyzer` processes bigwig files from `SnakemakeMethylFrac` for
580 downstream analysis. It calculates percentiles for each feature, methylation rates from median
581 methylated fractions and relative methylation rates. It computes the average methylated fraction
582 in 100-kb windows (T2T v1.1 assembly) and methylation rates using average methylated fractions.
583 It calculates the average methylated fraction relative to the TSS of active and inactive genes, and
584 relative to CTCF sites, smoothed in 21-bp windows. Finally, it generates the associated figures.
585 This software uses `pandas`⁴⁷, `pyBigWig`, `matplotlib`⁴⁹, `seaborn`⁵⁰ and `statsmodels`⁵¹.

586

587 **Analysis of RNA-seq, ATAC-seq and MNase-seq data**

588 We used RNA-seq datasets from the GEO database (GSE201262 for MCF7³²; GSE237066 for
589 MCF10A⁵²) using `salmon` v1.10.0⁵³ to normalise read counts. We averaged counts across
590 replicates for each transcript using the GENCODE v43 annotation of the Hg38 assembly⁵⁴. For
591 multi-transcript genes, we selected the most highly expressed transcripts and used their TSSs. We
592 used published bigwig files for ATAC-seq datasets for MCF7 (GSE201262³²) and MCF10A
593 (GSE152410⁵⁵). ATAC-seq data for 1-kb regions flanking TSSs were extracted; active genes were
594 assigned based on average signal (>0.4 for MCF7 and >2 for MCF10A). Our MNase-seq data were
595 aligned using `Bowtie2` v2.5.1⁴⁵. We selected read fragments for single nucleosomes (fragment

596 length: 120-180 bp) and counted the nucleosome dyads in 2010-bp regions flanking active gene
597 TSSs. We normalised dyad counts per gene using average dyad counts per flanking region, then
598 averaged dyad counts across all active genes, smoothing in 21-bp windows.
599

600 **CTCF sites and ChromHMM**

601 CTCF narrowPeaks were obtained from ENCODE ^{56,57} (ENCSR000AHD for MCF7 and
602 ENCSR193SZD for MCF10A). CTCF motifs were predicted using HOMER v4.11.1 ⁵⁸. For CTCF
603 phasing analysis, the peaks containing a single copy of the highest frequency motif were selected.
604 For ChromHMM analysis, we used ChIP-seq data from GSE85158 for both MCF7 and MCF10A
605 cells ³⁸, processed with the ENCODE ChIP-seq pipeline v2 ([https://github.com/ENCODE-](https://github.com/ENCODE-DCC/chip-seq-pipeline2)
606 [DCC/chip-seq-pipeline2](https://github.com/ENCODE-DCC/chip-seq-pipeline2)). ChromHMM v1.23 predicted 15 chromatin states using the T2T v1.1
607 assembly ³⁹. JHU RefSeqv110 + Liftoff v5.1 annotation (<https://github.com/marbl/CHM13>) were
608 used for feature enrichment. Final chromatin state annotations are available in our GitHub
609 repository (<https://github.com/zhuweix/methylFracAnalyzer>).
610

611 **Statistics and Reproducibility**

612 Two biological replicate experiments were performed. The panels shown in each main figure
613 belong to the same experiment. Extended Data figures generally show the results from both
614 replicate experiments, except for immunoblots, micrographs and DNA gels. The immunoblots and
615 DNA gel analyses were similar in both experiments. Correlations between biological replicates at
616 the chromosomal level are presented in Extended Data Fig. 8.
617

618 **References for Methods**

- 619 42 Kossila, M. *et al.* Improvement in adenoviral gene transfer efficiency after preincubation
620 at +37 degrees C in vitro and in vivo. *Mol Ther* **5**, 87-93 (2002).
621 <https://doi.org/10.1006/mthe.2001.0516>
- 622 43 Cole, H. A., Howard, B. H. & Clark, D. J. Genome-wide mapping of nucleosomes in yeast
623 using paired-end sequencing. *Methods Enzymol* **513**, 145-168 (2012).
624 <https://doi.org/10.1016/B978-0-12-391938-0.00006-9>
- 625 44 Molder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Res* **10**, 33 (2021).
626 <https://doi.org/10.12688/f1000research.29032.2>
- 627 45 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
628 **9**, 357-359 (2012). <https://doi.org/10.1038/nmeth.1923>
- 629 46 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
630 features. *Bioinformatics* **26**, 841-842 (2010).
631 <https://doi.org/10.1093/bioinformatics/btq033>
- 632 47 McKinney, W. *Data Structures for Statistical Computing in Python*. (2010).
- 633 48 Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular
634 biology and bioinformatics. *Bioinformatics* **25**, 1422-1423 (2009).
635 <https://doi.org/10.1093/bioinformatics/btp163>
- 636 49 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science &*
637 *Engineering* **9**, 90-95 (2007). <https://doi.org/10.1109/mcse.2007.55>
- 638 50 Waskom, M. seaborn: statistical data visualization. *Journal of Open Source Software* **6**,
639 3021 (2021). <https://doi.org/10.21105/joss.03021>
- 640 51 Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python.
641 *Proceedings of the 9th Python in Science Conference 2010* (2010).

- 642 52 Dorgham, M. G., Elliott, B. A., Holley, C. L. & Mansfield, K. D. m6A regulates breast
643 cancer proliferation and migration through stage-dependent changes in Epithelial to
644 Mesenchymal Transition gene expression. *Front Oncol* **13**, 1268977 (2023).
645 <https://doi.org/10.3389/fonc.2023.1268977>
- 646 53 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast
647 and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419 (2017).
648 <https://doi.org/10.1038/nmeth.4197>
- 649 54 Frankish, A. *et al.* GENCODE: reference annotation for the human and mouse genomes in
650 2023. *Nucleic Acids Res* **51**, D942-D949 (2023). <https://doi.org/10.1093/nar/gkac1071>
- 651 55 Gross, S. M. *et al.* A multi-omic analysis of MCF10A cells provides a resource for
652 integrative assessment of ligand-mediated molecular and phenotypic responses. *Commun*
653 *Biol* **5**, 1066 (2022). <https://doi.org/10.1038/s42003-022-03975-9>
- 654 56 Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data
655 portal. *Nucleic Acids Res* **48**, D882-D889 (2020). <https://doi.org/10.1093/nar/gkz1062>
- 656 57 Hitz, B. C. *et al.* The ENCODE Uniform Analysis Pipelines. *Res Sq* (2023).
657 <https://doi.org/10.21203/rs.3.rs-3111932/v1>
- 658 58 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime
659 cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-
660 589 (2010). <https://doi.org/10.1016/j.molcel.2010.05.004>

661
662

663 **Acknowledgements**

664 We thank Alex Vassilev for help with the FACS analysis. This study utilised the high-performance
665 computational capabilities of the Biowulf Linux cluster at the National Institutes of Health (NIH).
666 This research was supported by the Intramural Research Program of the NIH (NICHD).

667

668 **Author contributions**

669 H.P. and P.E. performed the experiments; Z.X. performed the bioinformatic analysis; H.P. and D.C.
670 wrote the manuscript.

671

672 **Competing interests**

673 The authors have no competing interests.

674

675 Correspondence and requests for materials should be addressed to DJC.

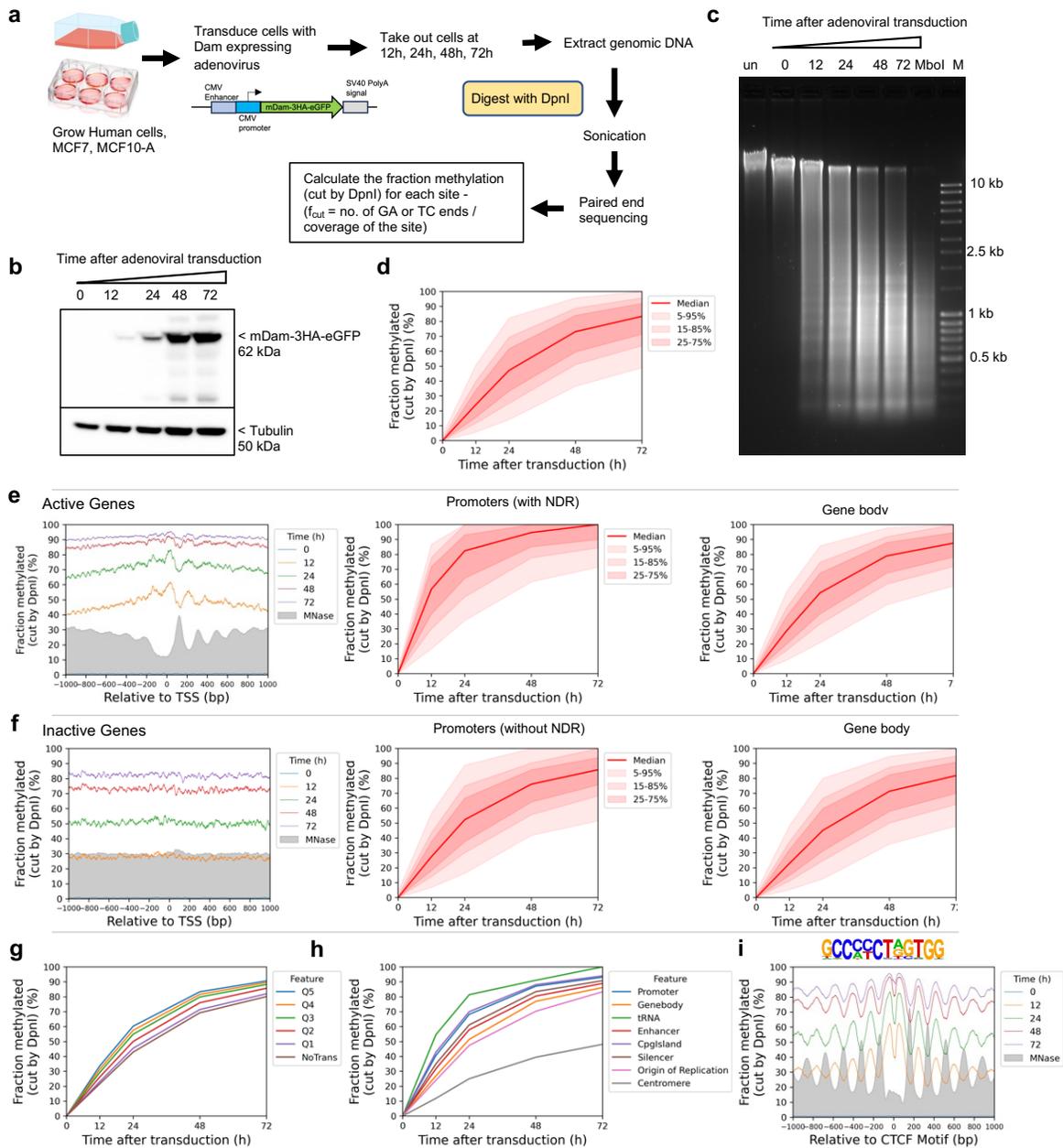


Fig.1 | The human genome is globally accessible in living MCF7 cells. **a**, Schematic of adenovirus transduction and time course experiment to express Dam methylase in live cells. **b**, Anti-HA immunoblot to detect Dam-3HA-eGFP expression in MCF7 cells. **c**, Agarose gel electrophoresis of DpnI-digested genomic DNA purified from MCF7 cells as a function of time of adenovirus treatment. 'un', undigested genomic DNA; 'MboI', DNA from non-transduced cells digested with MboI; M, DNA size marker. **d**, Almost complete methylation of GATC sites in MCF7 cells after transduction. Red line and shading: median GATC site methylation with data range indicated. **e**, **f**, Nucleosome phasing with respect to the TSS for active and inactive genes, as defined by ATAC-seq³². Grey profile: nucleosome dyad distribution in nuclei (MNase-seq data for MCF7 cells arbitrarily normalised to 30%). **g**, The effect of transcriptional activity on methylation rate. Active genes were divided into quintiles Q1 to Q5 based on increasing transcriptional activity (Q5 is the highest) using RNA-seq data from³²; methylation of the median GATC site in each quintile is shown. Inactive genes are treated as a single separate group ('NoTrans'). **h**, Median GATC methylation for various genomic regions. **i**, Nucleosome phasing around CTCF motifs using the motif shown.

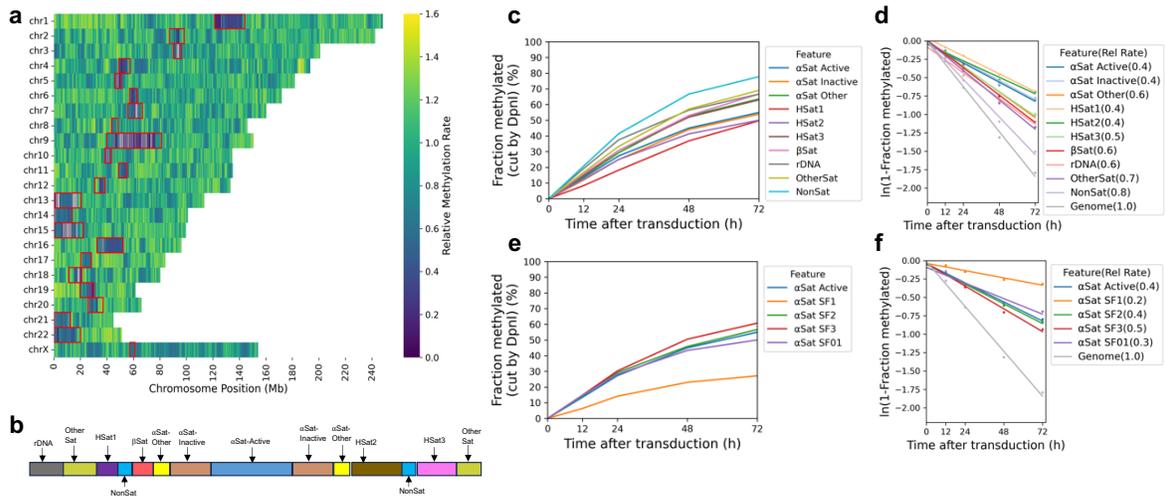


Fig.2 | Centromeres are methylated slower and reach a limit, unlike other genomic regions.

a, Heat map showing the variation in methylation rate at the chromosomal level. The average methylation rate was calculated for all GATC sites in each 100 kb window in the T2T genome by plotting ' $\ln(1 - \text{fraction methylated})$ ' against time after adenovirus transduction, and then normalised to the genomic average rate to obtain relative rates. Red rectangles: centromeric regions. **b**, Schematic of the organisation of the various centromeric elements defined in the T2T genome assembly (adapted from ³⁵). Divergent α -Sat and monomeric α -Sat were combined as " α -Sat other". **c**, Median GATC methylation time courses and **d**, methylation rates for the various centromeric elements. **e**, Median GATC methylation time courses and **f**, methylation rates for the various active supra-chromosomal α -satellite families.

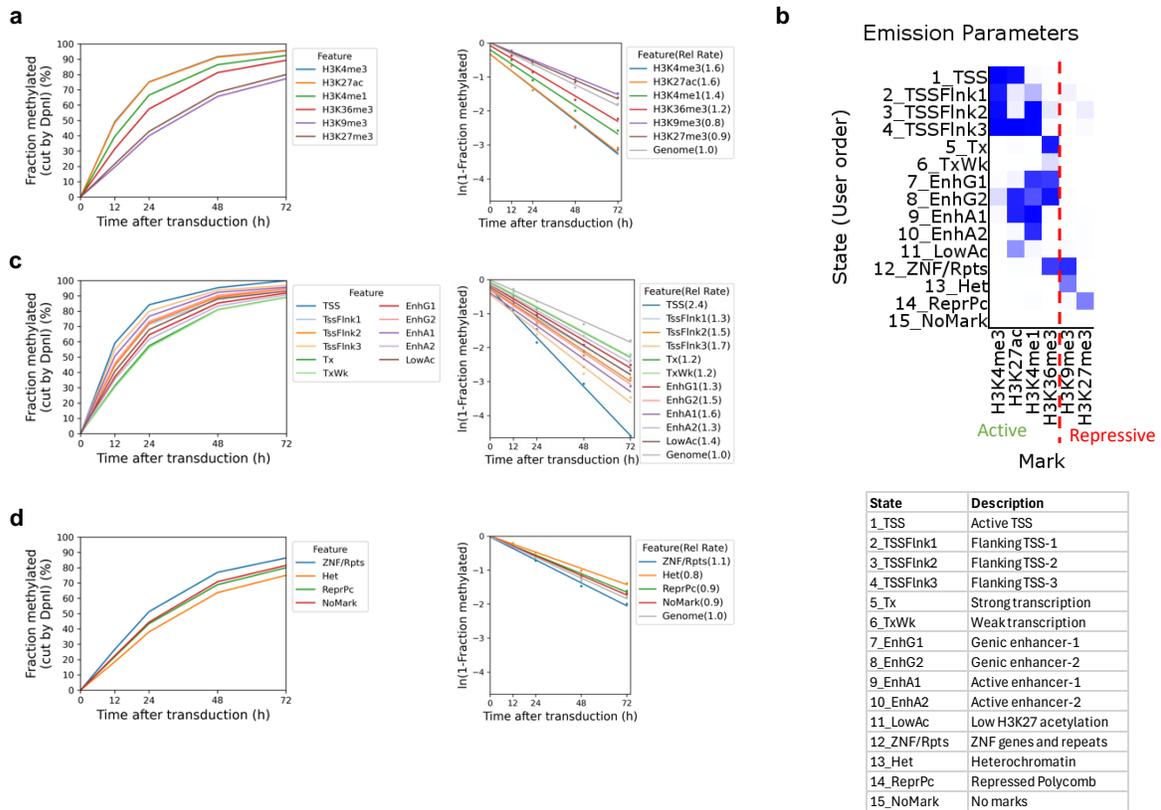


Fig.3 | Heterochromatin is accessible but methylated at a slower rate than euchromatin.

a, Methylation time courses and methylation rates for the median GATC site in regions with active histone marks (H3K4me1, H3K4me3, H3K27ac, H3K36me3) and inactive histone marks (H3K9me3 and H3K27me3). **b**, ChromHMM model defining 15 epigenetic states in MCF7 cells, defined as active or inactive chromatin based on the histone marks. **c**, Methylation time courses and methylation rates for the median GATC site in the active chromatin (euchromatin) states defined by the HMM model. **d**, Methylation time courses and methylation rates for the median GATC site in the inactive chromatin (heterochromatin) states defined by the HMM model.

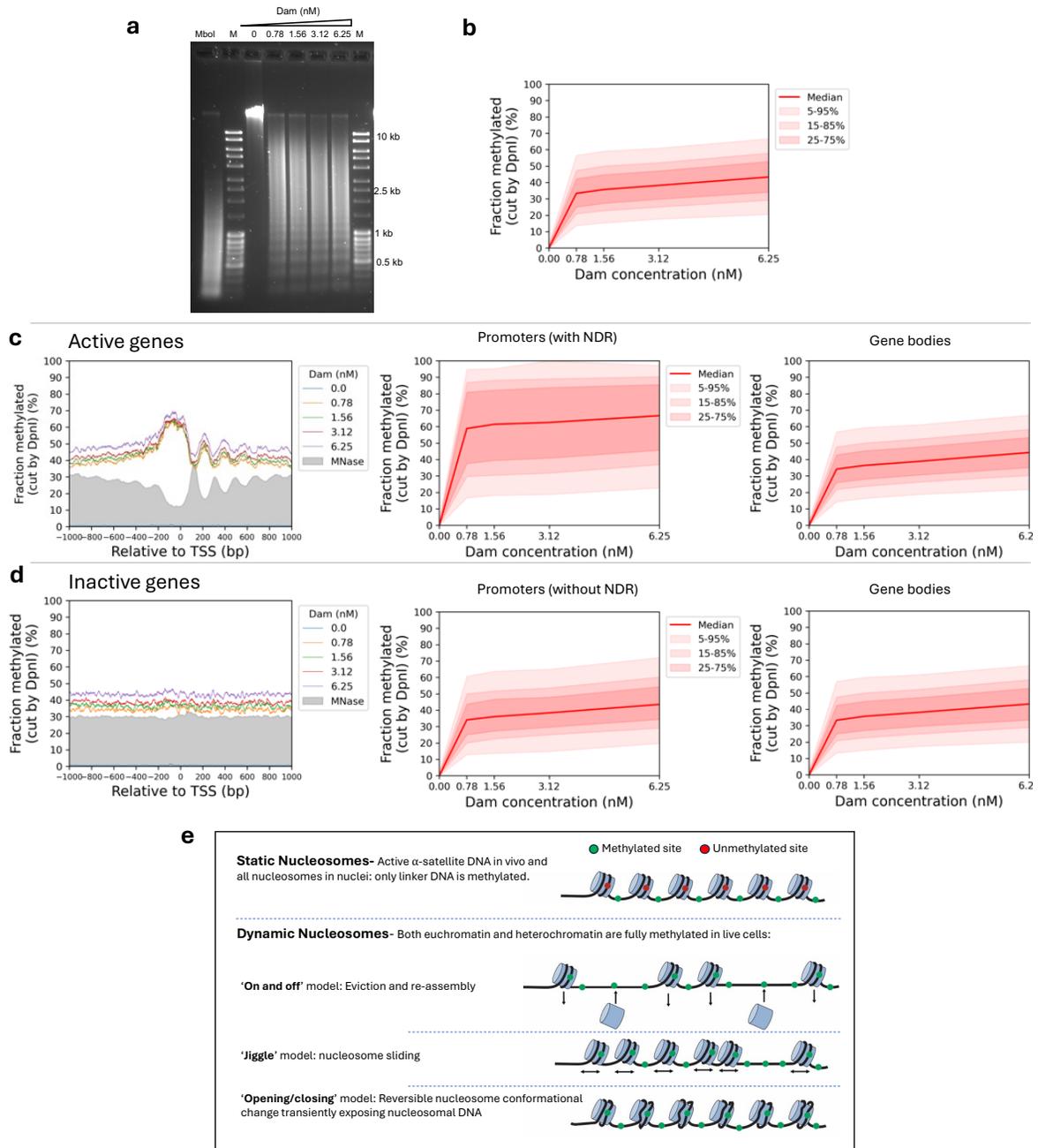
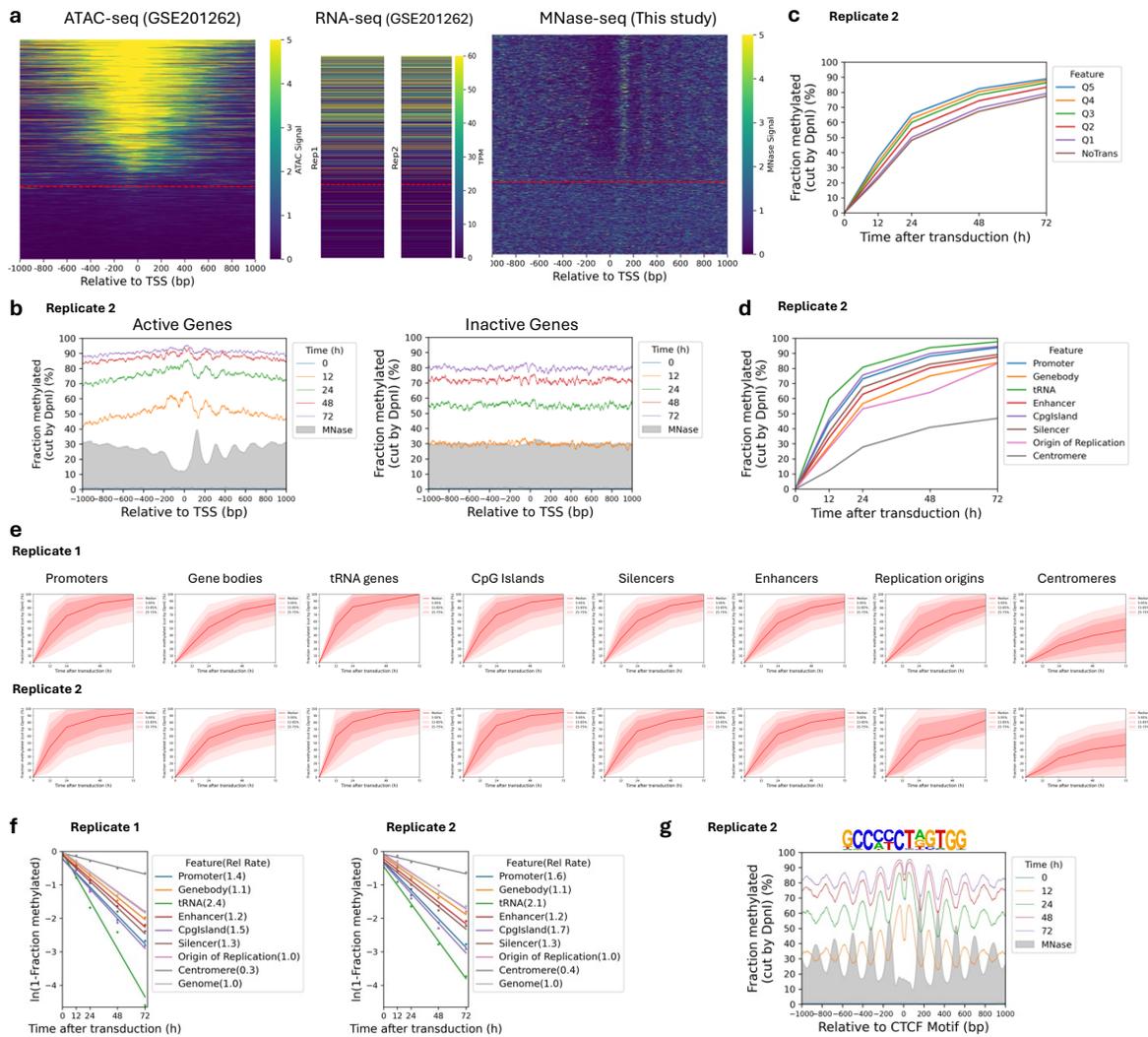
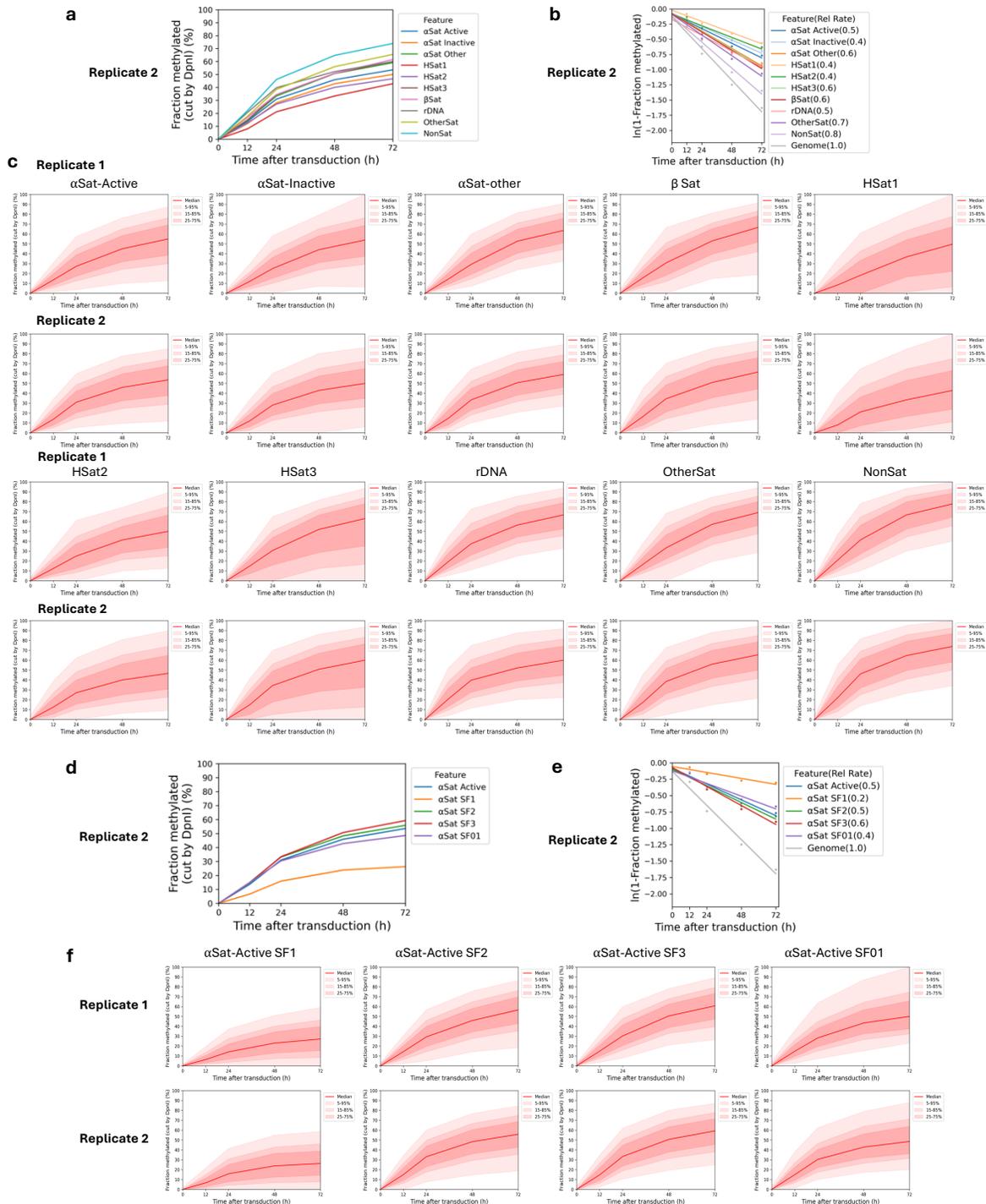


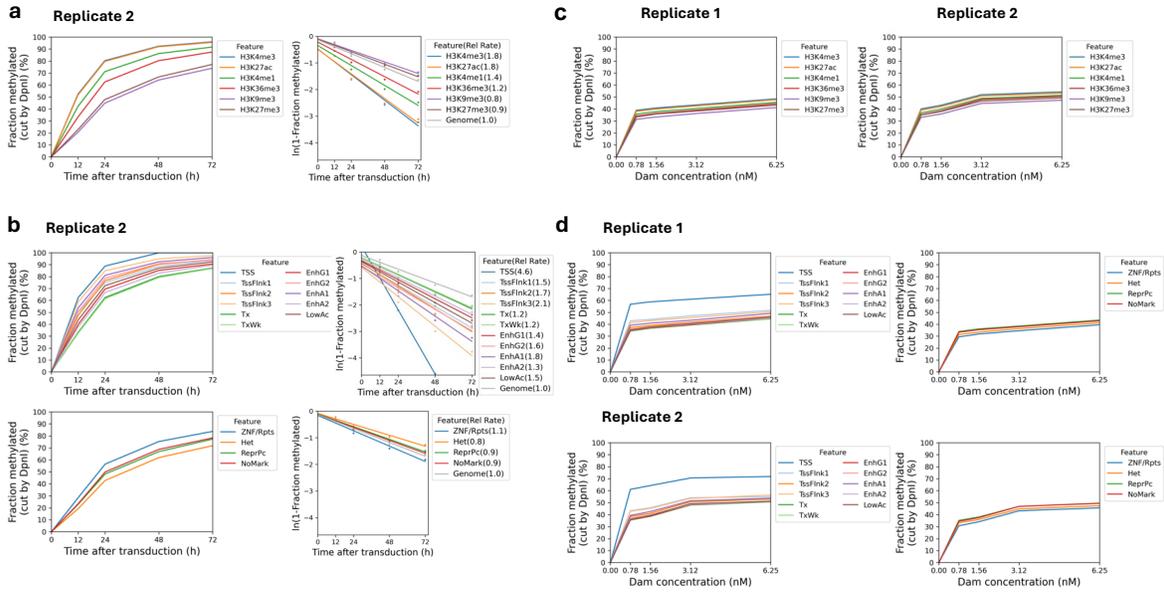
Fig.4 | Genome accessibility is limited in isolated MCF7 nuclei. a, Agarose gel electrophoresis of DpnI-digested genomic DNA purified from nuclei treated with increasing amounts of Dam. M, DNA marker. MboI, unmethylated control DNA fully digested at GATC sites by MboI. **b**, Methylation of all the GATC sites in the human genome as a function of Dam concentration. Red line: methylation of the median GATC site; shading indicates the data range. **c**, Active genes: Nucleosome phasing and methylation of the median GATC site in promoters or gene bodies as a function of Dam concentration. Red line and shading: median GATC site methylation with data range indicated. Grey profile: nucleosome dyad distribution in nuclei (MNase-seq data for MCF7 cells arbitrarily normalised to 30%). **d**, The same analysis for inactive genes. **e**, Possible mechanisms for generating accessibility in living cells (based on the known activities of various ATP-dependent chromatin remodelers).



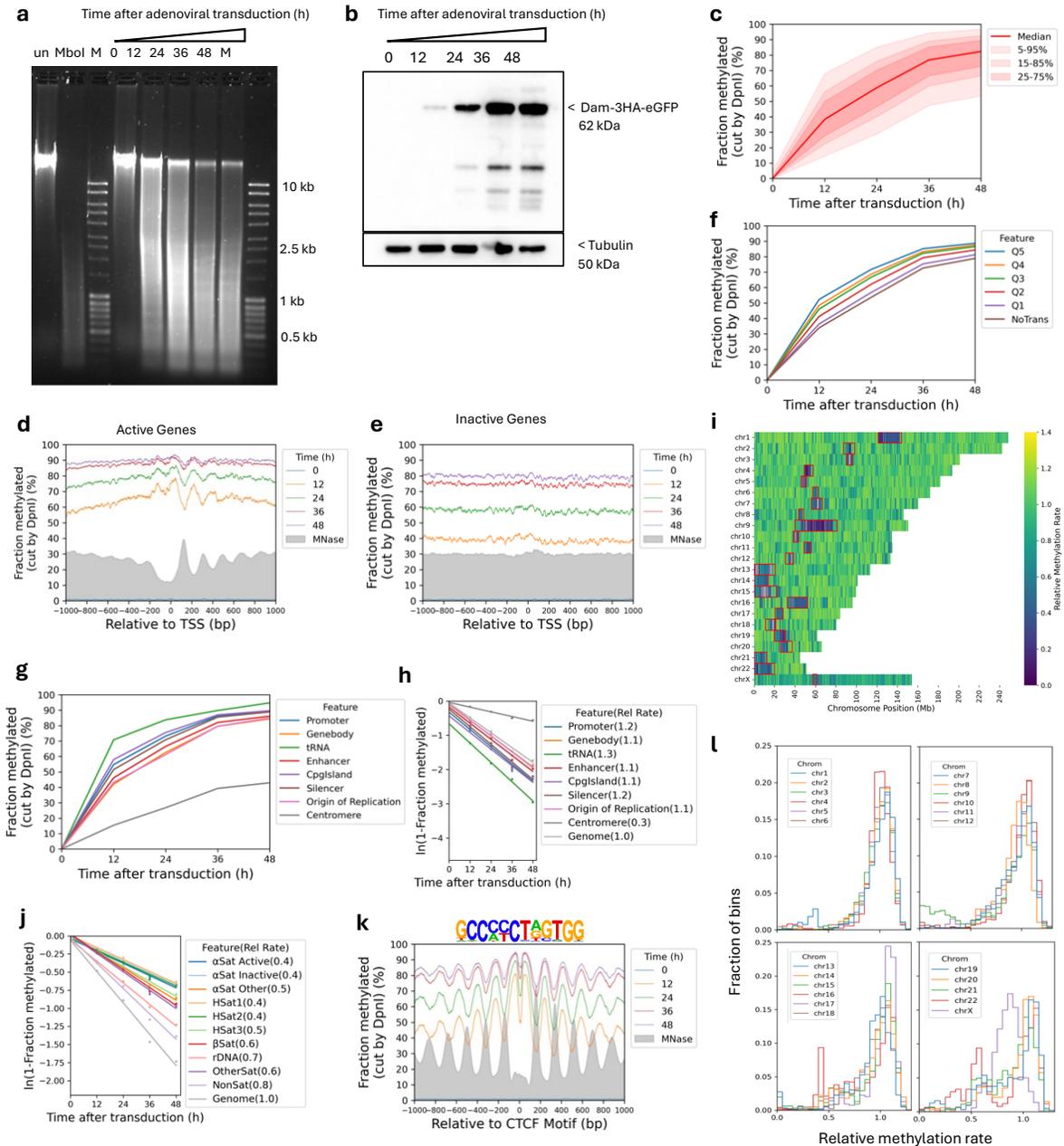
Extended Data Fig.1 | Rates of Dam methylation in various genomic regions in MCF7 cells. **a**, Left panel: All human genes sorted by ATAC-seq signal at their promoters in MCF7 cells³² relative to the major TSS. The red line separates genes with NDRs (active) from those that have no NDR (inactive). Middle panel: RNA-seq data for MCF7 cells (from³²) sorted as in the left panel. Right panel: MNase-seq data sorted as in the left panel. **b**, Nucleosome phasing in vivo detected by Dam methylation for Replicate 2 (see Fig.1e,f for Replicate 1). Methylation data for GATC sites across active and inactive genes at each time point are plotted relative to the TSS (smoothed with a 21-bp window). Grey profile: nucleosome dyad distribution in nuclei (MNase-seq data normalised to 30%). **c**, Effect of transcription on median GATC site methylation. The active genes were divided into quintiles, Q1 to Q5, with increasing transcriptional activity; inactive genes were treated as a single group ("NoTrans"). Data for Replicate 2 (see Fig.1g for Replicate 1). **d**, Methylation time courses for the median GATC site in various genomic regions. Data for Replicate 2 (see Fig.1h for Replicate 1). **e**, Methylation time courses for various genomic regions defined by hg38 annotations. Red line: median GATC site; shading: data range as indicated. **f**, Relative methylation rates for various genomic regions in vivo. Rates are relative to the genomic average for all GATC sites. **g**, Nucleosome phasing around CTCF motifs in vivo using the motif shown. Data for Replicate 2 (see Fig.1i for Replicate 1).



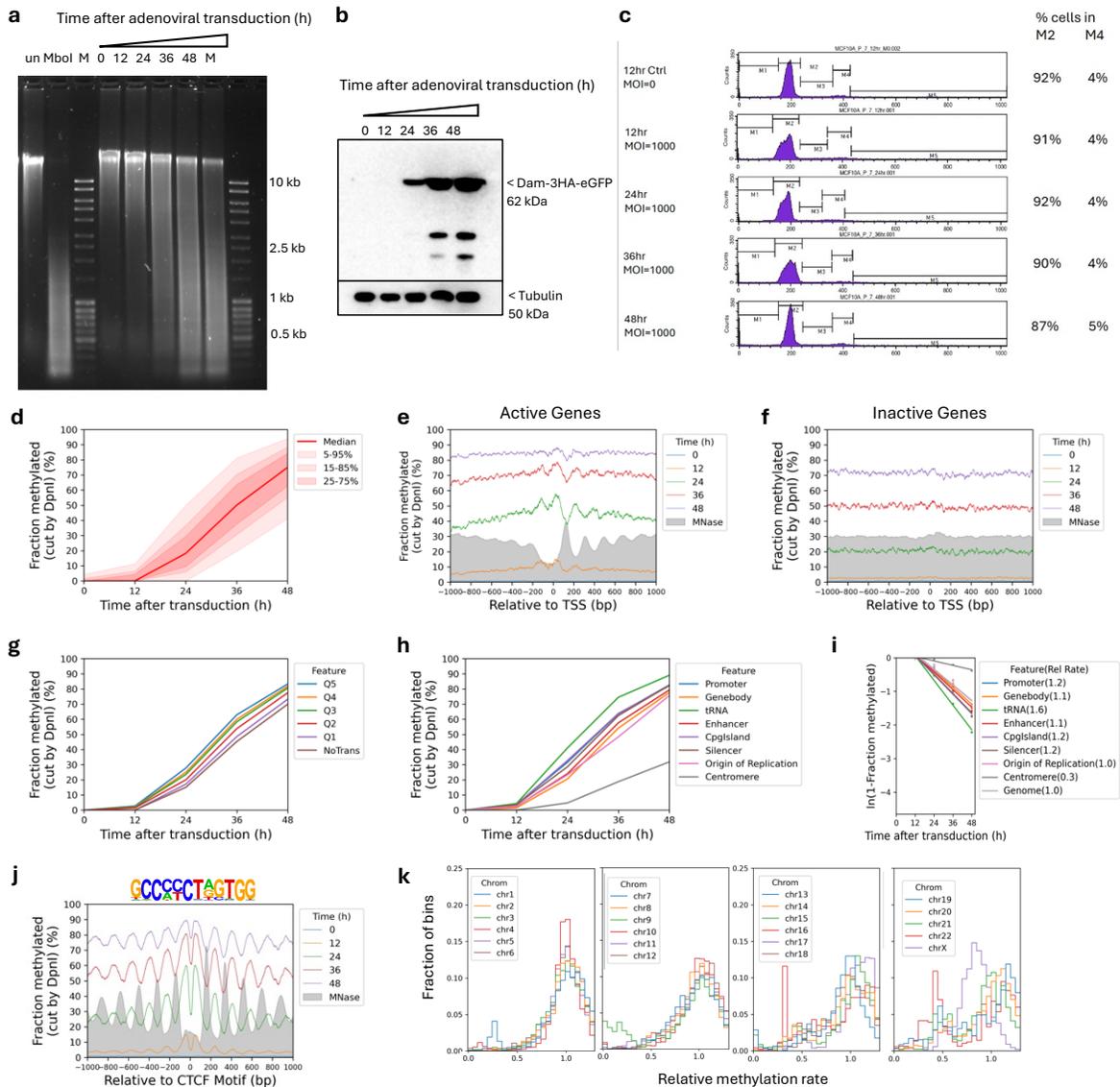
Extended Data Fig.2 | Dam methylation of the various centromeric satellite repeats in MCF7 cells. **a**, Methylation time courses for the median GATC site and **b**, methylation rates for the various centromeric elements relative to the genomic average site. Data for Replicate 2 (see Fig. 2c,d for Replicate 1). **c**, Methylation time courses for the median GATC site for the various centromeric elements. Red line: median GATC site; shading: data range as indicated. **d**, Methylation time courses for the median GATC site and **e**, relative methylation rates for the various active α -satellite supra-chromosomal families. Data for Replicate 2 (see Fig. 2e,f for Replicate 1). **f**, Methylation time courses for the median GATC site for the various active α -satellite supra-chromosomal families. Red line: median GATC site; shading: data range as indicated.



Extended Data Fig. 3 | Dam methylation of heterochromatin and euchromatin in living MCF7 cells and in MCF7 nuclei. **a**, Methylation time courses for the median GATC site and methylation rates for regions marked by histone modifications associated with euchromatin (H3K4me1, H3K4me3, H3K27ac or H3K36me3) or heterochromatin (H3K9me3 or H3K27me3), relative to the genomic average site. ChIP-seq data from ³⁸. Data for Replicate 2 (see Fig. 3a,b for Replicate 1). **b**, Methylation time courses for the median GATC site and relative methylation rates for the 15 epigenetic states defined by our ChromHMM model (see Fig. 3b). Data for Replicate 2 (see Fig. 3c,d for Replicate 1). **c**, Methylation of the median GATC site in nuclei for regions marked by histone modifications associated with euchromatin or heterochromatin. **d**, Dam methylation in isolated nuclei of the different chromatin states specified by the ChromHMM model (see Fig. 3b).

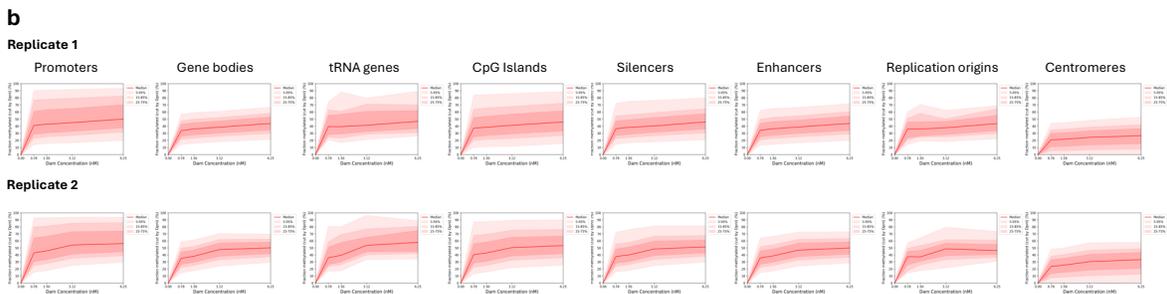
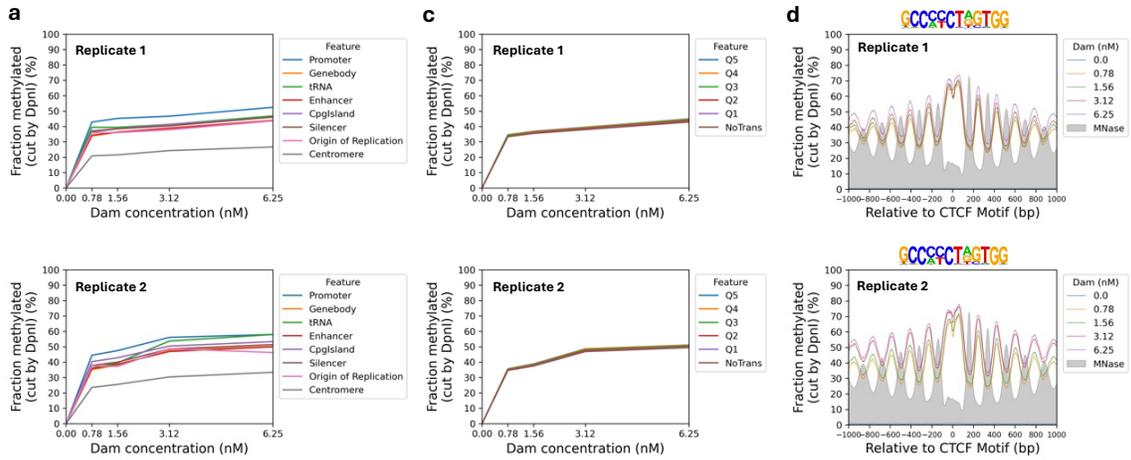


Extended Data Fig.4 | The human genome is globally accessible in live MCF10A cells. **a**, Agarose gel electrophoresis of DpnI-digested genomic DNA purified from MCF10A cells as a function of time of adenovirus treatment. 'un', undigested genomic DNA; 'MboI', DNA from non-transduced cells digested with MboI; 'M', DNA size marker. **b**, Anti-HA immunoblot to detect Dam-3HA-eGFP expression in MCF10A cells. **c**, Almost complete methylation of GATC sites in MCF10A cells after transduction. Red line and shading: median GATC site methylation with data range indicated. **d,e**, Nucleosome phasing with respect to the TSS for active and inactive genes, as defined by ATAC-seq data for MCF10A cells⁵⁵. Grey profile: nucleosome dyad distribution in nuclei (MNase-seq data for MCF7 cells arbitrarily normalised to 30%). **f**, The effect of transcriptional activity on methylation rate. Active genes were divided into quintiles Q1 to Q5 based on increasing transcriptional activity (Q5 is the highest) using RNA-seq data for MCF10A cells from (Dorgham et al. 2023); methylation of the median GATC site in each quintile is shown. Inactive genes are treated as a single separate group ('NoTrans'). **g**, Median GATC methylation for various genomic regions using annotations from the hg38 genome. **h**, Relative median GATC site methylation rates for various genomic regions. **i**, Heat map showing the variation in methylation rate at the chromosomal level in MCF10A cells. The average methylation rate was calculated for all GATC sites in each 100 kb window in the T2T genome by plotting 'ln(1 - fraction methylated)' against time after adenovirus transduction, and then normalised to the genomic average rate to obtain relative rates. Red rectangles: centromeric regions. **j**, Relative methylation rates for the various centromeric elements. **k**, Nucleosome phasing around CTCF motifs in MCF10A cells using the motif shown. **l**, Relative methylation rate data derived as in 'i' for each of the 23 chromosomes are separated into four separate plots for ease of comparison. Histograms of the fraction of 100-kb windows having a given relative methylation rate.

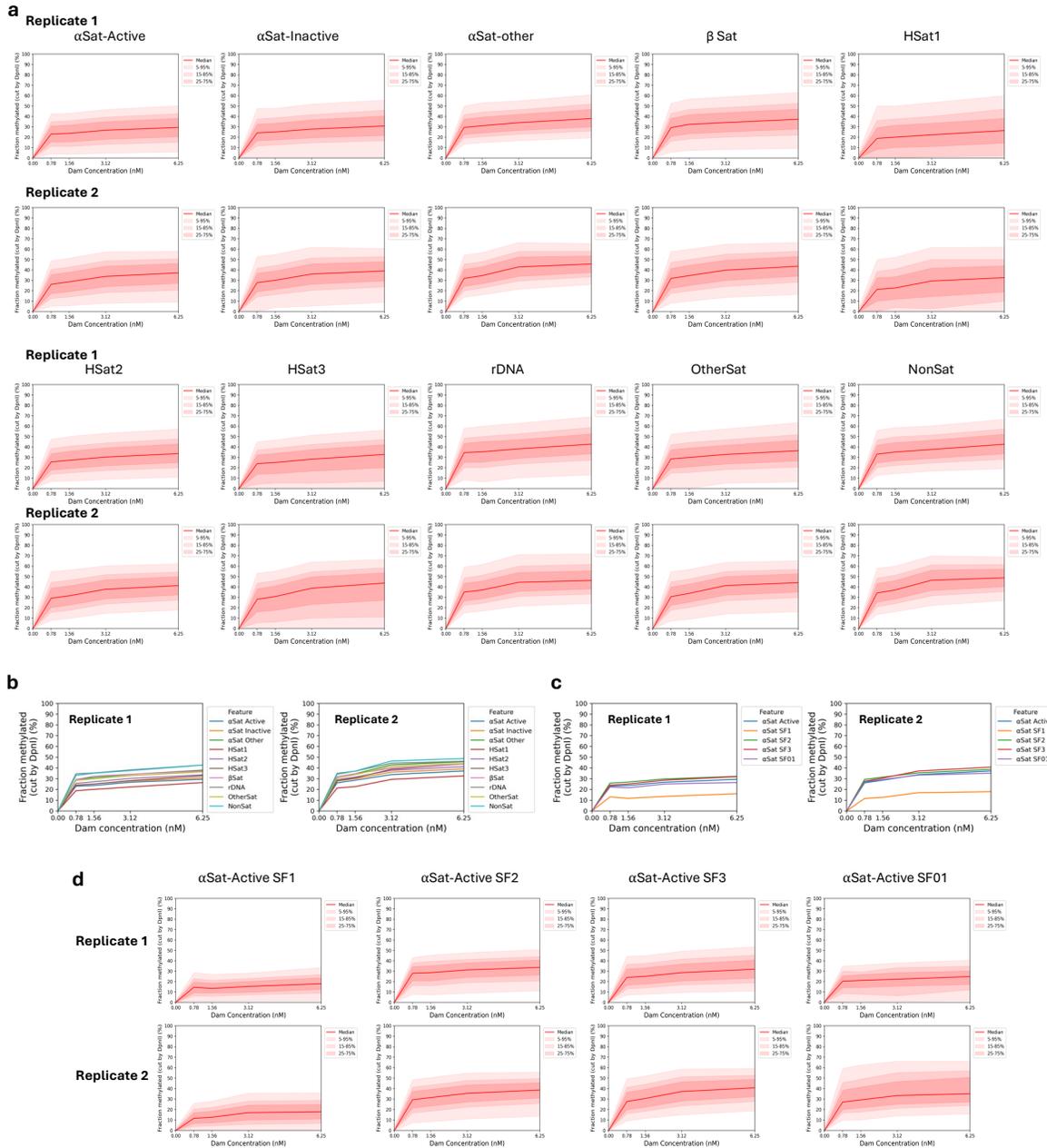


Extended Data Fig.5 | The human genome is globally accessible in confluent MCF10A cells.

a, Agarose gel electrophoresis of DpnI-digested genomic DNA purified from confluent MCF10A cells as a function of time of adenovirus treatment. 'un', undigested genomic DNA; 'MboI', DNA from non-transduced cells digested with MboI; M, DNA size marker. **b**, Anti-HA immunoblot to detect Dam-3HA-eGFP expression in MCF10A cells. **c**, FACS analysis confirms that the confluent MCF10A cells are arrested in G1 during the time course after adenovirus transduction. **d**, Time course of Dam methylation of all GATC sites in MCF10A cells after transduction. Red line and shading: median GATC site methylation with data range indicated. **e, f**, Nucleosome phasing with respect to the TSS for active and inactive genes, as defined by ATAC-seq data for MCF10A cells⁵⁵. Grey profile: nucleosome dyad distribution in nuclei (MNase-seq data for MCF7 cells arbitrarily normalised to 30%). **g**, The effect of transcriptional activity on methylation rate. Active genes were divided into quintiles Q1 to Q5 based on increasing transcriptional activity (Q5 is the highest) using RNA-seq data for MCF10A cells (Dorgham et al. 2023); methylation of the median GATC site in each quintile is shown. Inactive genes are treated as a single separate group ('NoTrans'). **h**, Median GATC methylation for various genomic regions using hg38 genome annotations. **i**, Relative median GATC site methylation rates for various genomic regions. **j**, Nucleosome phasing around CTCF motifs in MCF10A cells using the motif shown. **k**, Histograms of the fraction of 100-kb windows having a given relative methylation rate for each of the 23 chromosomes are separated into four separate plots for ease of comparison.

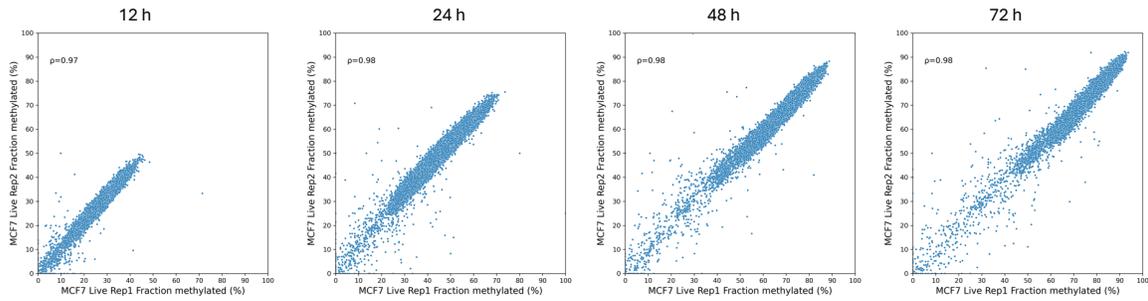


Extended Data Fig.6 | Limited genome accessibility in isolated MCF7 nuclei. Nuclei were treated with increasing concentrations of Dam. **a**, Comparison of the methylation of the median GATC site in various genomic regions as a function of Dam concentration. **b**, Separate plots for methylation of the median GATC site in various genomic regions as a function of Dam concentration. Red line and shading: median GATC site methylation with data range indicated. **c**, The effect of transcriptional activity on methylation rate in nuclei. Active genes were divided into quintiles Q1 to Q5 based on increasing transcriptional activity (Q5 is the highest) using RNA-seq data for MCF7 cells³²; methylation of the median GATC site in each quintile is shown. Inactive genes are treated as a single separate group ('NoTrans'). **d**, Nucleosome phasing in nuclei relative to CTCF motifs detected by Dam methylation using the motif shown. Methylation data for GATC sites at each Dam concentration are plotted relative to each CTCF site (smoothed with a 21-bp window). Grey profile: nucleosome dyad distribution in nuclei (MNase-seq data normalised to 30%).

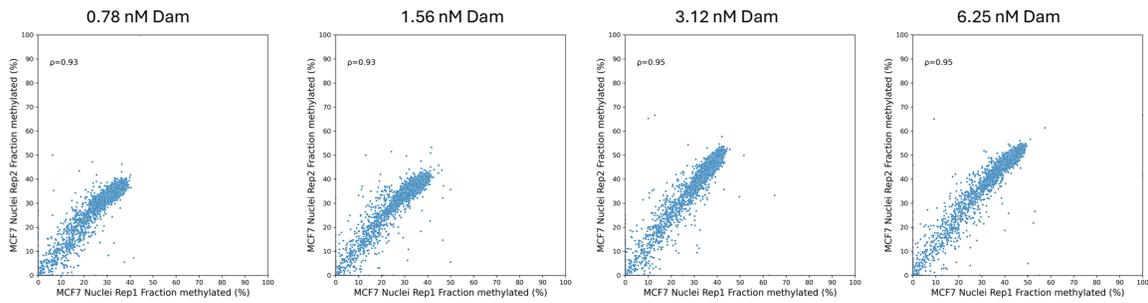


Extended Data Fig.7 | Methylation of centromeric elements is limited in nuclei. a, Separate plots showing methylation of the median GATC site in various centromeric elements as a function of Dam concentration. Red line and shading: median GATC site methylation with data range indicated. **b**, Comparison of the methylation of the median GATC site in the various centromeric elements as a function of Dam concentration. **c**, Comparison of the methylation of the median GATC site in the various active α -satellite supra-chromosomal families as a function of Dam concentration. **d**, Separate plots showing methylation of the median GATC site in various active α -satellite supra-chromosomal families as a function of Dam concentration.

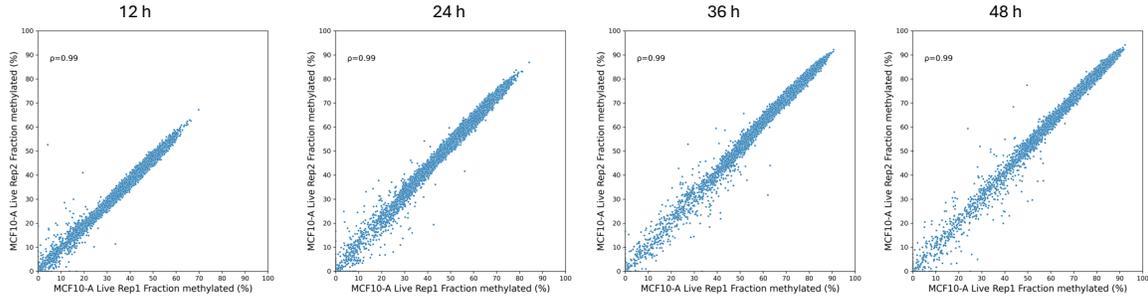
a MCF7 live cell experiments



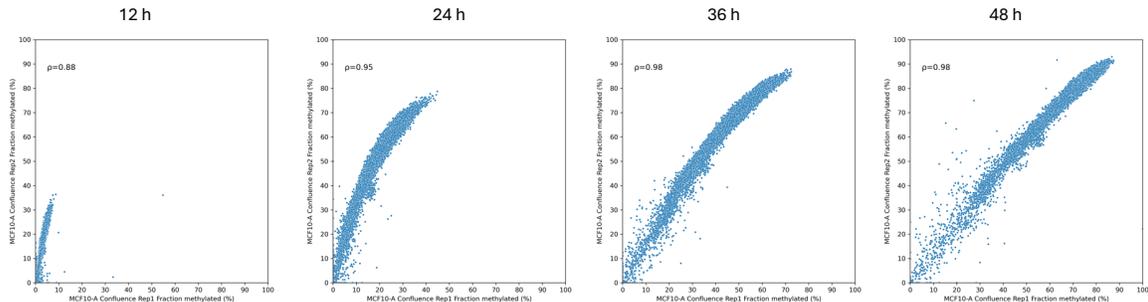
b MCF7 nuclei experiments



c MCF10A live cell experiments



d MCF10A confluent live cell experiments



Extended Data Fig.8 | Comparison of biological replicate experiments at the chromosomal level. Pearson correlations for the average % methylated for all GATC sites in each 100 kb window for each time point (live cells) or Dam concentration (nuclei). **a**, MCF7 cells, **b**, MCF7 nuclei, **c**, Dividing MCF10A cells, **d**, Confluent MCF10A cells.