



HAL
open science

G4access identifies G-quadruplexes and their associations with open chromatin and imprinting control regions

Cyril Esnault, Talha Magat, Amal Zine El Aabidine, Encar Garcia-Oliver, Anne Cucchiarini, Soumya Bouchouika, David Lleres, Lutz Goerke, Yu Luo, Daniela Verga, et al.

► To cite this version:

Cyril Esnault, Talha Magat, Amal Zine El Aabidine, Encar Garcia-Oliver, Anne Cucchiarini, et al.. G4access identifies G-quadruplexes and their associations with open chromatin and imprinting control regions. *Nature Genetics*, 2023, 55 (8), pp.1359-1369. 10.1038/s41588-023-01437-4 . hal-04285615

HAL Id: hal-04285615

<https://hal.science/hal-04285615>

Submitted on 20 Nov 2023

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.



Distributed under a Creative Commons Attribution 4.0 International License

1 **G4access identifies G-quadruplexes and their associations with open**
2 **chromatin and imprinting control regions**

3
4 Cyril Esnault^{1*}, Talha Magat^{1*}, Amal Zine El Aabidine^{1*}, Encar Garcia-Oliver^{1*}, Anne
5 Cucchiarini², Soumya Bouchouika¹, David Lleres¹, Lutz Goerke¹, Yu Luo^{2,3}, Daniela
6 Verga³, Laurent Lacroix⁴, Robert Feil¹, Salvatore Spicuglia⁵, Jean-Louis Mergny²,
7 Jean-Christophe Andrau^{1#}

8
9 ¹ Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS-
10 UMR 5535, 1919 Route de Mende, 34293 cedex 5, Montpellier, France.

11 ² Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, Inserm, Institut
12 Polytechnique de Paris, Palaiseau, France.

13 ³ Université Paris-Saclay, Institut Curie, Orsay, France.

14 ⁴ Institut de Biologie de l'École Normale Supérieure, ENS, CNRS UMR8197, Inserm
15 U1024, Paris, France.

16 ⁵ Aix-Marseille University, INSERM, TAGC, UMR 1090, Marseille, France. Equipe
17 Labellisée Ligue Contre le Cancer.

18
19
20 # Corresponding author: jean-christophe.andrau@igmm.cnrs.fr

21 * These authors contributed equally.

22
23
24 **Abstract:** Metazoan promoters are enriched in secondary DNA structure forming
25 motifs, such as G-quadruplexes (G4s). Here we describe 'G4access', an approach to
26 isolate and sequence G4s associated with open chromatin via nuclease digestion.
27 G4access is antibody- and crosslinking-independent and enriches for computationally
28 predicted G4s (pG4s), most of which are confirmed *in vitro*. Using G4access in human
29 and mouse cells, we identify cell-type-specific G4 enrichment correlated with
30 nucleosome exclusion and promoter transcription. G4access allows measurement of
31 variations in G4 repertoire usage following G4 ligand treatment, HDAC and G4
32 helicases inhibitors. Applying G4access to cells from reciprocal hybrid mouse crosses
33 suggests a role for G4s in the control of active imprinting regions. Consistently, we also
34 observed that G4access peaks are unmethylated, while methylation at pG4s correlates
35 with nucleosome repositioning on DNA. Overall, our study provides a novel tool for
36 studying G4s in cellular dynamics and highlights their association with open chromatin
37 , transcription, and their antagonism to DNA methylation.

40 Introduction

41 Eukaryotic promoters encompass a wide range of sequence, but tend to have modest
42 conservation in evolution. For example, while mammalian promoters tend to be GC-
43 rich (>70%), yeast and Drosophila promoters are AT-rich. Despite these pronounced
44 differences, eukaryotic promoters harbor similar properties in their ability to recruit the
45 transcriptional machinery and to exclude and position nucleosomes¹. We previously
46 showed that mammalian CpG islands (CGIs) intrinsically exclude nucleosomes,
47 independently of transcription². Our more recent work also emphasizes that within
48 CGIs, and more generally in human and mouse promoters, G-quadruplex (G4) forming
49 sequences are likely to play a crucial role in nucleosome exclusion both in cells and *in*
50 *vitro*³. Unimolecular G4s are DNA secondary structures well characterized *in vitro*⁴.
51 Their investigation in living cells is more recent and they are suggested to play
52 essential roles in transcription, replication, genome stability and homeostasis⁵. They
53 can also be predicted by computational algorithms^{6,7} such as G4Hunter that calculates
54 a robust likelihood score of genomic sequences forming G4 structures at fixed window
55 sizes.

56 Various experimental techniques have been developed to characterize G4 formation
57 in cells. The recent development of G4 ChIP has enabled the identification of
58 thousands of G4-forming genomic sequences at promoters and elsewhere⁸⁻¹⁰.
59 However, G4 ChIP is highly dependent on the selectivity of G4 recognition by the BG4
60 nanobody¹¹, which may stabilize unstable structures *in vitro*^{12,13}, introducing a potential
61 bias. Finally, previous observations report that it is difficult to apply G4 ChIP to all cell
62 types, in particular non-cancerous primary cells⁸. Therefore, orthogonal methods are
63 needed to identify G4s formed in the chromatin context.

64
65 Here, we describe 'G4access', an Ab- and crosslinking-independent method coupled
66 to high-throughput sequencing, that identifies G4 forming sequences (G4FS)
67 associated with open chromatin in cells. Taking advantage of the sequence preference
68 of Micrococcal nuclease, we isolated G4-enriched fractions of chromatin following
69 enzymatic titration. G4access yields cell-specific G4 patterns that are enriched for
70 accessible chromatin at promoters and other genomic loci. We validated a large
71 fraction of G4access sequences as forming G4 structures using multiple, large-scale
72 *in vitro* assays. Enriched G4access loci not only correlate with open chromatin but also
73 associate with repositioned nucleosomes, and are tightly linked to the presence of
74 initiating/paused RNA Polymerase (Pol) II. However, G4access signals are only
75 moderately impaired by transcriptional inhibitors, suggesting that they are not
76 dependent on active transcription. Unexpectedly, cell treatment with a G4-stabilizing
77 ligand yielded strong G4 dynamics *in vivo* associated with gain of regions with lower
78 G4 potential. Knockdown of the G4 helicases DHX36 and WRN resulted in the specific
79 increase of G4access signal at strong G4 containing-promoters. Moreover, applying
80 G4access to reciprocally-crossed hybrid mESCs shows that increased allelic G4
81 potential correlates with gene expression, suggesting a link between G4 formation and
82 transcription. We also describe an antagonism between apparent G4 formation and
83 DNA methylation, providing a possible mechanism for this observation. Finally, we
84 applied our procedure to genomes from other species with lower densities of predicted
85 G4-forming sequences and consistently find an association with open chromatin, albeit
86 to a lower extent than in mammalian cells.

87 Results

G4access, a method to enrich G4 forming sequences *in vivo*

We previously showed that CGIs, enriched at mammalian promoters, tend to exclude nucleosomes intrinsically². We then searched for motifs associated with this property, by analyzing sequences associated with the deepest point of apparent nucleosome depletion upstream of annotated transcription start sites (TSSs). Our motif search led us to G-rich motifs, including several G stretches that are characteristic of G4 structure formation. Furthermore, predicted G4s (pG4s) densities at promoters confirmed an association between pG4 and the lowest nucleosome density (Fig. 1a, Extended Data Fig. 1a), consistent with previous description¹⁴. Based on this observation, we hypothesized that we could map G4s in living cells based on their nucleosome exclusion potential. For this purpose, we developed a technique to isolate subnucleosomal (~147bp) DNA fragments that are protected by DNA secondary structures using Micrococcal Nuclease (MNase) digestion.

First, we reasoned that subnucleosomal fragments might be enriched in G4FS given the observed nucleosome depletion at promoters and other locations of the genome. Second, MNase possesses both endo and exonuclease activity and has been reported to have a cutting preference before G-stretches^{15,16}, while G4s are also resistant to λ -exonuclease¹⁷. Thus, we assumed that G4s should be enriched in the genomic sequences targeted by low levels of MNase digestion (Fig. 1b).

We performed MNase titration in three human cell lines from different tissues, and optimized recovery of G4 sequences at a known model G4 by qPCR (Fig. 1c and Extended Data Fig. 1b). As expected, we observed that low/moderate MNase digestion (around 30% mononucleosome) consistently yielded subnucleosomal fractions enriched for our model G4. Because MNase digestion was performed in suboptimal ionic conditions for G4 formation, we checked that control G4s could form *in vitro* under ionic and temperature conditions used in our procedure (Extended Data Fig. 1c-d). We finally subjected the purified and size-selected DNA to library preparation and high-throughput sequencing. Experimental signal was reproducibly correlated ($r > 0.76$) within each of the three cell lines (Extended Data Fig. 1e). A closer investigation of the data indicated that the signals are cell-specific with both common and specific locations in the three cell types (Fig. 1d and Extended Data Fig. 2a-b). Because G4FS are found enriched in open chromatin areas, we further dubbed this technique “G4access”. Interestingly, G4access widely overlaps with Pol II at promoters (see below) and is consistent with G4 ChIP profiles^{8,10} (Fig. 1d and Extended Data Fig. 2c). However, peaks called in G4access data are more sharply resolved than in G4 ChIP (Fig. 1e). Furthermore, pG4 scores and the enrichment of G4Hunter-predicted G4s were comparable across the 2 techniques (Fig. 1f-g and Extended Data Fig. 2d). Shuffling nucleotides while keeping base composition constant confirmed the methods’ specificity (Extended Data Fig. 2e). Importantly, G4access allows for >50% enrichment of sequences with a G4Hunter prediction score >1.5, which have a high likelihood of forming a G4 *in vitro*⁶. We confirmed this observation by scoring G4 subtype structures from motif predictions in individual cell lines. We found that 75 to 90% of the sequences fit one or another pG4 category (Extended Data Fig. 3a). When considering peaks common to the 3 cell types, this fraction climbed to 96% of the sequences (Fig. 1h). Additionally, G4access genomic locations are over-represented at TSSs and at 5’ regions of genes (Extended Data Fig. 3b), with 15-40% of the peaks located within promoters, comparable to G4 ChIP data sets. We also noted that G4access can yield

140 substantial enrichment of CTCF motif in K562 cells. From a sequence point of view,
141 ChIP and G4access display comparable numbers of G-stretches (Extended Data Fig.
142 3c) while numbers of G per track show more cell-type variability in ChIP (Extended
143 Data Fig. 3d). As expected, their GC and CpG content at promoters is also higher than
144 average, consistent with a strong association with CpG islands (Extended Data Fig.
145 3e). We also observed G4access-enriched promoters at genes expressed in multiple
146 cell-types, with the possible exception of DNA repair-genes, which may represent a
147 more G4-specific class (Extended Data Fig. 3f-g). Finally, a sequence search clearly
148 indicates G4 compatible-motifs in all human G4access peaks, with a prevalence at
149 TSSs (Extended Data Fig. 4).

150
151 Based on the above findings, we concluded that G4access enables the identification
152 of cell-specific enrichment of G4FS, with a prevalence at TSSs.

153

154 **G4access enriched areas can form G-quadruplexes *in vitro***

155

156 G4Hunter-predicted G4s were previously validated *in vitro* on a set of mitochondrial
157 sequences⁶. We therefore asked if G4access-identified G4FS could, at large scale, be
158 validated using three independent *in vitro* assays. Firstly, we identified 4743 common
159 G4access peaks in all three cell lines, and then selected 596 representative 30nt
160 regions with G4Hunter scores >0.5 (see methods; Fig. 2a). We next assessed their
161 abilities to form G4 structures by performing FRET-MC¹⁸, Thioflavin T (Th-T) and N-
162 methyl mesoporphyrin IX (NMM) fluorescent ligand assessment (Extended Data Fig.
163 5a). FRET-MC validated that ~80% of the tested sequences form G4 structures *in vitro*.
164 Moreover, we found G4 formation for 97% of sites with G4Hunter score >1.35 (Fig. 2b
165 and Extended Data Fig. 5b), using ThT and NMM assays (Fig. 2c and Extended Data
166 Fig. 5c-d). Strikingly, FRET-MC, which may be the most robust method, validated 95%
167 of G4access sequences with a score of 1.2 and above (Fig. 2d). This high level of
168 validation exceeds that was previously observed for mitochondrial genome and
169 suggests that G4access further enriches for G4FS.

170

171 In sum, our *in vitro* analyses confirm that G4access strongly enriches for G4FS, in line
172 with our genomic observations.

173

174 **G4access can monitor G4 dynamics in living cells**

175

176 To ascertain whether the G4access procedure can be used to analyze G4 dynamics
177 in cells, we performed experiments in which G4s are expected to change in the
178 chromatin context via small molecule treatments and knocking down G4 helicases.

179

180 G4s can be targeted by ligands that stabilize them *in vitro*. *In vivo* ligands' mode of
181 action remains relatively enigmatic although it appears that they generate double
182 strand breaks at various genome locations¹⁹. To get further insights into G4 ligand's
183 action, we treated the cells with Pyridostatin (PDS), a well known G4 ligand¹⁹, for a
184 short time (30 min), to avoid indirect effects, and performed G4access before and after
185 treatment. We found that although many hits are conserved, G4access regions are
186 relatively dynamic after PDS treatment (Extended Data Fig. 6a-b). Many peaks
187 observed, including at promoters, redistributed globally or locally. Surprisingly, we
188 found that the average G4Hunter scores of the PDS-induced G4s were lower, while
189 G4s with high scores tended to decrease (Extended Data Fig. 6c). An example of this

190 is shown in Extended Data Fig. 6a in which the G4access signals relocate from a
191 strong to a weaker G4FS at the *ATP2B4* promoter. Overall, these results suggest that
192 while strong G4s are only moderately affected by the drug since they are stable, the
193 weaker ones become more enriched in open chromatin areas due to ligand
194 stabilization. However, we could not exclude that the loss of strong G4s might be due
195 to sequencing biases. To address this possibility, we analysed previously published
196 genome-wide *in vitro* data using G4seq experiments with or without PDS²⁰. G4seq
197 maps G4s based on error rates incorporation during DNA amplification on purified
198 genomic DNA and thus out of the chromatin context. We found that average scores of
199 G4seq are higher than with G4access and preferential stabilization of the weaker G4s
200 is also observed, although to a lesser extent (Extended Data Fig. 6d), suggesting that
201 weak G4s indeed become preferentially stabilized by brief PDS treatment.
202

203 Next, we investigated the impact of known G4 helicases disruption by siRNA knock-
204 down experiments of the WRN and DHX36 helicases, previously described as
205 unwinding G4s^{21,22}. As shown in Fig. 3a-b, siRNA knockdown almost fully reduced
206 protein expression 72 h after transfection. We then performed G4access in mock- and
207 siRNA-treated cells. Differential analysis identified both increased and decreased G4
208 signals but, strikingly, more G4s were up-regulated at promoters (~98%), suggesting
209 that G4s normally resolved by WRN and DHX36 helicases form specifically at these
210 locations. Examples of such increased G4access signals at promoters are shown in
211 Fig. 3c-d. Of note, the promoters of the up-regulated G4access peaks also had higher
212 G4 scores (Fig. 3e-f). These results suggest that WRN and DHX36 function as G4
213 unwinding DNA helicases with preferential activity at promoters.
214

215 Overall, we conclude that G4s are dynamic *in vivo* in response to a G4 ligand or upon
216 helicase disruption, and that G4access can identify these dynamics.
217

218 **G4access hallmarks nucleosome exclusion and transcription**

219
220 Given the strong association of G4access-identified loci with promoters and open
221 chromatin, we subsequently investigated the association of G4 regions with
222 nucleosome positioning, exclusion and transcription at promoters and non-promoter
223 regions of the genome. To improve the accuracy of our analysis, we selected
224 G4access peaks associated with medium stringency G4 annotations (see methods).
225 As shown in Fig. 4a, at all locations, G4access regions were associated with
226 nucleosome exclusion, Pol II binding and nucleosome positioning around the
227 G4access summits (examples are shown in Fig. 4b).
228

229 We next ranked non-promoter G4access regions by increasing nucleosome density
230 (MNase signal) and analyzed nucleosome positioning, Pol II and G4access signal (Fig.
231 4c and Extended Data Fig. 7a) and defined 4 classes of nucleosome positioning and
232 depletion patterns. We observed that nucleosome positioning is generally conserved
233 in the $\frac{3}{4}$ of the regions (classes 2 and 3), while a minority of regions in which G4s
234 deplete or position more accurately nucleosomes (class 1 and 4, respectively) do not
235 show such phasing. We speculate that class 1 represents regulatory regions with
236 promoter-like properties while class 4 represents areas in which predicted G4s cannot
237 open chromatin and thus do not affect positioning of their surrounding nucleosomes.
238 In support of this, class 1 regions show stronger H3K4me3 signals as compared to

239 H3K4me1, consistent with histone mark signature of promoter activity²³ (Extended
240 Data Fig. 7a).

241
242 Since G4access also isolates non-G4FS, we wondered whether non-G4s showed
243 similar properties. In contrast to the predicted G4-containing sequences, these non-
244 G4s areas do not associate with open chromatin, show fuzzy nucleosome positioning
245 and reduced Pol II association (Extended Data Fig. 7b), indicating that G4FS in
246 G4access display specific properties. We also interrogated whether G4access
247 specifically yielded open chromatin regions associated with G4s, or if any observed
248 NDR would similarly yield G4-enriched regions. To address this, we isolated all NDRs
249 and plotted G4Hunter scores for those with and without G4access signal. Our analyses
250 revealed that G4access NDRs were enriched in high scores as compared to other
251 NDRs, supporting the specificity of G4access for G4FS (Extended Data Fig. 7c).

252
253 Noting the association with transcription, we then asked if G4 formation in open
254 chromatin would be affected by transcriptional perturbation. It has been proposed in
255 the past that transcription induces a torsional stress that may stabilize G4s *in vitro*. Our
256 recent observations suggested the opposite, i.e. that G4s would promote chromatin
257 opening and Pol II recruitment^{8,24,25}. To further address this question, we treated the
258 cells with transcription inhibitors specific for initiation or elongation (triptolide and
259 KM05283, respectively) and analysed if this resulted in loss of G4access signal. As
260 previously reported, we found that triptolide treatment stripped Pol II signal across all
261 genic features while KM05283 only removed Pol II from within gene bodies²⁷ (Fig. 4d,
262 left). Interestingly, at promoter locations elongation blockade did not reduce G4
263 signals, whereas initiation inhibition reduced G4access levels by half (Fig. 4d, right).
264 At genic locations, distal to the promoter, both inhibitors reduced G4access by half. An
265 example of such variations is shown in Fig. 4e. We conclude that transcriptional
266 inhibition neither abolishes G4 formation, nor its association to open chromatin, but
267 instead reduces it. This suggests a model in which G4 formation precedes Pol II
268 recruitment and becomes further stabilized by the ensuing transcriptional activity.
269 Recent work described similar observations in K562 cells²⁸ although in this case ChIP-
270 seq did not identify changes in G4 signal in the presence of inhibitors. This difference
271 might originate from the cell types, the techniques used or bioinformatic set-up of the
272 analyses.

273
274 Next, we compared G4access and G4 ChIP in their abilities to identify regions of open
275 chromatin, as well as to generate strong nucleosome positioning. As shown in
276 Extended Data Fig. 8, both sets of technique-specific regions have comparable ATAC-
277 seq signal, while the G4access-specific showed more apparent NDRs and better
278 nucleosome positioning in MNase-seq data. We also note that common ChIP-
279 G4access areas are the most open, possibly because they are more enriched for
280 promoter regions.

281
282 To investigate how global chromatin perturbation might impact G4access signals, we
283 treated Raji cells with TSA, a HDAC inhibitor. While TSA globally relaxes chromatin
284 through histone hyperacetylation (Extended Data Fig. 9a), it can reduce ATAC-seq
285 signals at open chromatin regions^{29,30}. We confirmed this effect in MNase-seq following
286 24h TSA treatment where nucleosome densities increased at NDR locations. At both
287 promoter and non-promoter G4access locations, we also observed consistent
288 reduction of the signals where predicted G4s are present, indicating that nucleosomes

289 tend to re-position to G4 sites, thus presumably reducing their formation (Extended
290 Data Fig. 9b-c). We note that these results are in contrast with effects observed with
291 another HDAC inhibitor^{8,28}. However, when growth condition such as hypoxia result in
292 chromatin compaction at pG4 sites, G-quadruplex formation is also reduced²⁷.

293
294 Together, our analyses show that G4 structures found in open chromatin regions are
295 associated with transcription as well as nucleosome positioning. While transcriptional
296 activity does not appear essential for G4 formation, it might stabilize their structure.

297

298 **G4s associate with active alleles in hybrid mES cells**

299

300 Because of their G-rich content, predicted G4s are over-represented at CGIs. We find
301 that the same is true for Differentially Methylated Regions (DMRs) at imprinted gene
302 domains, many of which also correspond to CGIs as well³¹. Because of their parental
303 DNA methylation imprints, DMRs are stably repressed on their methylated allele, and
304 are transcriptionally active on the non-methylated allele³¹. They constitute thus an
305 attractive model to compare repressed and active alleles within the same nuclear
306 context.

307

308 We hypothesized that G4 formation in open chromatin might occur specifically at the
309 active, unmethylated alleles of imprinted DMRs. To address this question, we
310 performed G4access in hybrid mouse embryonic stem cells (ESCs) obtained through
311 reciprocal crosses between *M. m. domesticus* C57BL/6J and *M. m. molossinus* JF1³²
312 strains (BJ and JB; Fig. 5a). These strains are genetically divergent, allowing efficient
313 discrimination of the parental alleles.

314

315 We assessed a total of 31 well-characterised mouse imprinted DMRs (Supplementary
316 Table 1, Methods). We observed differential allelic G4access signals at 7 of them, out
317 of 11 DMRs showing signal (pvalue<0.05; Fig. 5b). Interestingly, at both paternally and
318 maternally imprinted DMRs, G4access signals were much stronger on the expressed,
319 unmethylated allele, suggesting that G4 formation is related to the unmethylated state
320 (Fig. 5b). For example, the promoter-DMR of the imprinted *Meg3* gene is methylated
321 and repressed on its paternally-inherited allele, and is exclusively expressed from the
322 maternal allele (Fig. 5c). G4access displays similar allelic asymmetry, with signal
323 coming virtually only from the unmethylated copy of the DMR (Fig. 5c). This suggests
324 that G4 formation is associated with the allelic expression of *Meg3*. We confirmed this
325 observation at another imprinted locus, *Peg13*, which comprises a maternally
326 methylated DMR (Fig. 5d). At this imprinted DMR, G4access signal is again strong at
327 the expressed and unmethylated paternal allele. These experiments suggest that G4
328 formation and repressive DNA methylation are mutually exclusive, at least at imprinted
329 DMRs. However, they do not directly address whether methylation of G4 DNA could
330 be used as a mechanism to destabilize them in favour of nucleosome association.

331

332 **DNA methylation antagonizes G4 formation and open chromatin**

333

334 To gain further insight into this question, we analyzed methylation at G4Hunter-
335 predicted regions with or without G4access signal (Fig. 6a-b). These analyses were
336 performed at medium G4 predictions (G4H1.2) that contained at least 2 CpG
337 nucleotides. We found that experimentally-identified G4s are associated with a loss of
338 CpG methylation, consistent with previous observations¹⁰. Yet, we also found that the

339 G4s corresponding to weak/medium predictions tend to be more methylated when
340 compared to neighbouring DNA. These results suggest that G4 formation favours open
341 chromatin formation, while pG4 DNA methylation antagonizes this process. To further
342 address this issue at the most basic level, in the context of naked chromatin, we re-
343 analyzed data from *in vitro* nucleosome assembly using human granulocyte genomic
344 DNA³³. We ranked the *in vitro* nucleosome densities at pG4s by increasing signal (Fig.
345 6c) and analysed corresponding signals for *in vivo* nucleosomes and DNA methylation.
346 This analysis indicates a clear relationship between methylation levels at pG4s and
347 nucleosome formation *in vitro* and *in vivo* (Fig. 6d). These results propose that high
348 methylation of G4FS results in a sequence context favourable to intrinsic nucleosome
349 formation and G4 structure destabilisation.

350

351 **G4access enriches for weaker G4s in lower eukaryotes**

352

353 While our data suggest an important role of G4s found in open chromatin in mammalian
354 transcription and imprinting control, we wondered whether the G4access procedure
355 could also isolate G4FS in lower complexity eukaryotes. We chose *D. melanogaster*
356 and *S. cerevisiae* as models as they carry 1.5 and 12-fold less G4FS at the genomic
357 scale as compared to humans, respectively (Extended Data Fig. 10a). However, and
358 interestingly, the relative G4Hunter density at promoters suggests a selective pressure
359 for G4FS in human cells that is absent in *Drosophila*. In budding yeast, a bias towards
360 the stronger scores is also observed at promoters although to a lesser extent than in
361 humans.

362

363 We adapted our G4access protocol in S2 *drosophila* cells and in a S288C yeast strain,
364 following an adjusted workflow (see methods). We compared the procedure's
365 performance to isolate G4FS in the 3 organisms by plotting the G4access's prediction
366 scores (Extended Data Fig. 10b). As in mammalian cells, we found an enrichment of
367 G4s in the sequenced fragments. However, and as expected from the genomic
368 densities of G4 predictions, we observed lower G4Hunter scores in *drosophila* and far
369 lower in yeast. Further investigation of the DNA motifs present in G4access data
370 indicated G-rich and GAGA motifs in *drosophila* and A/T or G/C-rich sequences in yeast
371 (Extended Data Fig. 10c). Moreover, 60% and 34% of potential G4 subtypes were
372 found in *Drosophila* and in yeast, respectively (Extended Data Fig. 10d). When
373 browsing the G4access locations in the genomes, our data confirmed that G4FS were
374 more enriched at promoters in yeast as compared to *Drosophila*. As expected, in both
375 cases, G4access locations correlated with open chromatin ATAC-seq signal
376 (Extended Data Fig. 10e).

377

378 We conclude that G4s can be identified in yeast and fly chromatin using G4access,
379 albeit less efficiently than in mammals, likely reflecting a lower association of G4s with
380 open chromatin in these organisms. We note an enrichment of strong G4 sequences
381 in yeast that may reflect a specific role for the few genes described to display G4s at
382 promoters^{14,34}.

383

384 **Discussion**

385

386 In this study, we introduce G4access, an efficient procedure for scoring G4 forming
387 sequences associated with open chromatin in cells. Because G4 formation requires
388 opening of the DNA double-helix, making DNA incompatible with a stable nucleosome

389 locally, we propose that G4access-identified pG4s reflect structures that are formed *in*
390 *vivo*. Based on our *in vitro* validation, it appears that G4access loci that are conserved
391 across different cell lines have a robust potential to form G4 structures in various
392 assays, even at low prediction scores. We speculate that these areas will define critical
393 gene control regions that display a broad spectrum of expression in different cell types.
394 Compared to existing methods such as ChIP or Cut&Run, G4access thus appears as
395 a useful orthogonal approach with comparable performance. Future improvements of
396 the method may take advantage of combining assays but also optimizing K+
397 concentration during extraction procedure, closer to physiological conditions.

398
399 Our investigations suggest a role for G4s in chromatin opening, nucleosome
400 positioning, Pol II transcription and imprinting control, expanding the possible functions
401 of these secondary DNA structures in the genome. Among these properties, it is
402 remarkable that G4s harbor the apparent ability to position nucleosomes with a
403 phasing comparable to that of insulators^{3,35}. We further show that G4s are maintained
404 without active transcription, albeit reduced. This indicates that G4s may be formed prior
405 to transcription and that transcription further stabilizes their structure, reconciling
406 previous conflicting observations^{8,28}. This result is also consistent with the idea that
407 negative supercoiling upstream of the transcription front increases G4 stability²⁴.

408
409 Our study also finds that G4s might contribute to imprinted gene expression. Amongst
410 the DMRs analyzed, all those displaying an allele-specific expression or active
411 unmethylated state were shown to display an allele-specific G4access signal. This
412 suggests that G4s play an activating role in DMRs/CGIs that are not methylated. Our
413 data also indicate that while G4s formed in a given cell antagonize methylation, local
414 CpG methylation also disfavours G4 formation and their related nucleosome exclusion
415 property, providing a possible model for paternal or maternal allele inactivation at DMR.
416 In agreement with this, it was previously shown that CpG methylation results in loss of
417 chromatin opening and phasing at CGIs³⁶, both features associated with G4FS in the
418 present study. Nevertheless, we note that methylation on G4 oligonucleotides does not
419 clearly modify the properties of G4 formation *in vitro*³⁷⁻³⁹. This could originate from the
420 difference of topology between oligonucleotides and dsDNA embedded in nucleosome
421 *in vivo*. In the latter, topological constraints could result in more sensitivity to DNA
422 methylation on G4 structure, a possibility that requires further investigation.

423
424 Up to now, the precise mechanism of action of G4 ligands *in vivo* has remained elusive.
425 It is proposed that they stabilize G4s in the genome, but the spectrum of their DNA
426 target sites has never been explored extensively. Unexpectedly, our results suggest
427 that G4s with low-to-medium *in silico* prediction scores are preferentially stabilized by
428 PDS. Although we cannot completely rule out sequencing bias in our data, we favor
429 the possibility that weaker G4s are preferential ligand targets. Future development of
430 this idea will require exploring the genomic effects of other G4 ligands.

431
432 In summary, the G4access technique will enable future investigations to better
433 understand the role of G4s in transcription control but also other genomic processes
434 such as replication⁴⁰, DNA repair and the role of helicase in genome stability⁴¹, as well
435 as deepening the mechanistic understanding of various G4-targeting drugs, including
436 some that are believed to display an anti-cancer potential⁴².

437
438 **Acknowledgements:**

439 This work was supported in the JCA lab by grants from the (ANR-20-CE12-0023), FRM
440 (AJE20130728183), INCA PLbIO (N°2020-117) and CNRS 80prime 2021 (DeciphG4).
441 This project has received financial support from the CNRS through the MITI
442 interdisciplinary programs. CE was supported in part by an ARC grant (retour postdoc).
443 We thank Beatrice Loriod and the Transcriptomics and Genomics Marseille-Luminy
444 (TGML) platform for sequencing the G4access samples. We are grateful to David
445 Monchaud for providing us with G4 interfering molecules in exploratory experiments.
446 TGML is a member of the France Génomique consortium (ANR-10-INBS-0009). EGO,
447 TM and SB were supported by grants from the Epigenesys Labex of excellence and
448 EGO in part by ANR-18-CE12-0019. We acknowledge the financial support from the
449 France Génomique National Infrastructure, funded as part of "Investissement d'Avenir"
450 program managed by the Agence Nationale de la Recherche (contract ANR-10-INBS-
451 09) for the MGX sequencing platform facility in Montpellier. We are also grateful to the
452 Genotoul Bioinformatics Platform Toulouse Midi-Pyrenees for computing and storage
453 resources. The funders had no role in study design, data collection and analysis,
454 decision to publish or preparation of the manuscript. We also thank Pablo Navarro and
455 Eric Kremer for critical reading of the manuscript and the Raman-Livaja lab for help in
456 the yeast extracts preparation.

457

458 **Author contribution statement**

459 CE, TM, AZEA and EGO contributed equally to the work. More specifically, CE and
460 EGO performed important initial set-up experiments for G4access that were further
461 optimized by CE and TM. AZEA performed most of the bioinformatic analyses with the
462 help of CE and TM, and LL developed and provided the modified G4Hunter scripts
463 used in our analyses. AC contributed most of the G4 biophysical assays described in
464 the manuscript under the supervision of JLM and with the help of YL and DV in the
465 initial stages of the project. SB and TM contributed the helicases, HDAC and
466 transcription inhibition while EGO and CE performed the PDS treatment. TM and LG
467 performed and analyzed the G4access in yeast and drosophila with the help of AZEA
468 for some bioinformatic analyses. DL and RF provided support for the mouse hybrid ES
469 model including by providing the ES cells, advices on cell culture and by establishing
470 the list of DMR regions listed in Supplementary Table 1. SS helped and advised us for
471 sequencing some of the G4access samples at the TGML. JCA conceived the frame of
472 the study, supervised the work and wrote the manuscript. All authors read, corrected
473 and proofread the manuscript.

474

475 **Competing Interests statement**

476 The authors declare no competing interests.

477

478

479 **Figure legends**

480 **Figure 1: G4access principle and validation.**

481 **a-** G4FS at promoters and motifs associated to open regions upstream of TSSs. The
482 graph shows nucleosome and G4H2.0 densities in Raji cells (top 20% of active
483 promoters, all promoters are shown in Extended Data Fig. 1a). Motifs are shown for all
484 promoters, top and bottom 20% of active promoters. **b-** Principle of G4access.
485 Chromatin is digested by MNase and subnucleosomal fractions are purified at
486 moderate digestion prior library preparations. **c-** Initial set-up, optimization and quality
487 controls of G4access. Top left shows gel electrophoresis of a representative MNase
488 titration (out of $n > 10$) and the subnucleosomal purified DNA (red square) used for

489 initial set-up. Bottom left is the fraction of mononucleosomes in the titration curve
490 expressed as percent of all DNA (Mean and SD of n=2 biological replicates are
491 displayed). The 2nd point (30% mononucl.) is considered as the optimal for G4
492 enrichment in our experimental frame. The right panel indicates mono- (146 bp), di-
493 (320 bp) and multi-nucleosomal Bioanalyzer profiles. **d-** Genome browser view (Chr7:
494 7.100.000-7.900.000) of G4access signal and corresponding G4-ChIP^{8,10} in indicated
495 cell lines. G4H2.0 predictions are indicated below. **e-** Peak size distribution in G4-ChIP
496 and G4access. Vertical dashed lines indicate the maximum (bulk) sizes of fragments
497 in the sequenced samples. **f-** Observed/expected G4Hunter predictions in
498 G4access/ChIP in human cell lines. **g-** G4Hunter prediction scores in G4access and
499 equivalent selection of random genomic DNA fragments. Around 75% of G4access
500 peaks are >G4H1.2, which represents a likelihood >85% of forming a G4 *in vitro*⁶. The
501 difference of distribution is highly significant (p -value < $2.2E^{-16}$, two-sided Wilcoxon
502 test). **h-** Repartition of the G4 subtypes in G4access peaks in Raji cells. The various
503 categories are 'Loop size' 1–3, 4–5 and 6–7: sequences with at least one loop of the
504 respective length; simple bulge: sequences with a G4 with a bulge of 1–7 bases in one
505 G-run or multiple 1-base bulges; 2-tetrads/Complex bulge: sequences with a G4s with
506 two G-bases per G-run or several bulges of 1–5 bases; and other: other G4 types that
507 do not fall into the former categories (see methods).

508

509 **Figure 2: Experimental validation of G4 structure in G4access selected**
510 **sequences.**

511 **a-** Strategy for *in vitro* validation of G4access sequences common to the 3 cell lines as
512 indicated. **b-** FRET-MC assessment of the G4 structures on the 596 30nt
513 oligonucleotides. All oligonucleotides with a S factor < 0.5 are considered as G4s.
514 Positive and negative controls are shown on the side (see Supplementary Table 2). **c-**
515 Validated (yellow) and non-validated (blue) G4access sequences analyzed by the
516 indicated methods (n=596). Sequences are ranked by decreasing G4Hunter scores
517 from top to bottom. **d-** Percentage of validated G4 structures in indicated G4Hunter
518 ranks for FRET-MC experiments. The percentage of sequences in each rank is also
519 indicated (red line).

520

521 **Figure 3: G4access measures G4 dynamics in response to siRNA targeting G4**
522 **helicases.**

523 **a-** siRNA targeting DHX36 efficiently reduces the helicase expression level and
524 promotes G4access signal at promoters. (left) Representative Western blots of total
525 H3 (loading controls) and of DHX36 are shown (n=2 biological replicates; full blot scans
526 of biological replicates are shown in 'Source data'). (right) DESeq was used to identify
527 differential G4access signal from control to DHX36 knock-down cells. The resulting
528 volcano plot, performed on sequences with G4H>1.2, is displayed, (red) differential
529 promoter signal, (blue) differential non TSS signal, (grey) unaffected signals (DESeq,
530 p -value < 0.05). Majority (97.7%) of differential promoters show an increased
531 G4access signal (422/432) (n=2 biological replicates of G4access). **b-** siRNA targeting
532 WRN efficiently reduces the helicase expression level and promotes G4access signal
533 at promoters. (left) Representative Western blots of total H3 and of WRN are shown
534 (n=2 biological replicates; full blot scans of biological replicates are shown in 'Source
535 data'). (right) DESeq volcano plot, performed on sequences with G4H>1.2, result is
536 displayed, (red) differential promoter signal, (blue) differential non TSS signal, (grey)
537 unaffected signals (DESeq, p -value < 0.05). Majority (97.8%) of differential promoters
538 show an increased G4access signal (1591/1626) (n=2 biological replicates of

539 G4access). **c-** Representative example of G4access increasing signal upon DHX36
540 knockdown at the *SIN3B* promoter (Chr19: 16.938.400-16.942.400). **d-** Representative
541 example of G4access increasing signal upon WRN knockdown at the *PDDC1* promoter
542 (Chr11: 775.000-779.000). G4H predictions with score > 1.5 are shown. **e-** DHX36
543 knockdown leads to increase G4access signal at high confidence/strong pG4s.
544 G4Hunter score density plots of unaffected, decreased and increased G4access
545 peaks. **f-** WRN knockdown leads to an increased G4access signal at high
546 confidence/strong pG4s. G4Hunter score density plots of unaffected, decreased and
547 increased G4access peaks.

548

549 **Figure 4: G4s are associated to open chromatin, nucleosome positioning and**
550 **Pol II and are partially independent on transcription.**

551 **a-** Nucleosome positioning and Pol II recruitment centered on G4access sites
552 overlapping G4 predictions at promoter (TSS, n=4619) and non-promoter sites (non
553 TSS, n=11807). See also methods. **b-** Examples of G4access peaks, nucleosome
554 depletion and positioning at promoter (upper panel, Chr7: 116.498.400-116.507.000)
555 and non-promoter (lower panel, Chr1: 155.061.700-155.065.500) sites in Raji cells.
556 G4H regions with score > 2.0 are shown below. **c-** G4access peaks associate to strong
557 nucleosome depletion, Pol II recruitment and define highly positioned nucleosomes at
558 most genomic G4s. Heat maps are shown ranked by decreasing MNase signals
559 around G4access peaks overlapping weak/moderate G4 annotations (see methods).
560 4 groups were defined based on the features of the individual MNase profiles as
561 follows: I (844): strong depletion and low positioning, II (6215): depletion and strong
562 positioning, III (2073): moderate depletion and positioning, IV (2675): no depletion and
563 weak positioning (see Extended Data Fig. 7 for individual profiles). **d-** Transcription
564 inhibition results in moderate G4access decrease. Average profiles of Pol II over genes
565 (composite profile, n=1808) or G4access at promoters (n=1808) and gene bodies
566 (n=349) following triptolide and KM05283 treatment. **e-** Examples of transcription
567 inhibition effects on Pol II and G4access signal over the CD19 gene (Chr16:
568 28.941.930- 28.951.789) in Raji cells. G4H regions with score > 2.0 are shown below.

569

570 **Figure 5: Differential G4access signal at active and inactive imprinted genomic**
571 **loci in mouse hybrid ES cells.**

572 **a-** Reciprocal crosses of hybrid mouse strains (JF1 and C57/Bl6) used in this study for
573 maternal and paternal imprinting (2i-medium derived cell lines). **b-** Allele-specific
574 G4access and RNA signals in BJ and JB ESCs at DMRs with differential G4access
575 signals, only considering the regions containing a strain specific SNP with interpretable
576 signals. In each case the imprinted methylated allele is indicated by B^m or J^m depending
577 on the strain it originates from (see Supplementary Table 1). Individual datasets, mean
578 with SEM are presented (n=2 for G4access, n=2 for RNA-seq for each cross BJ and
579 JB). **c, d-** Example of G4access enrichment at the TSS-proximal DMR of the *Meg3*
580 (chr12: 110.775.000-110.825.000) and *Peg13* (chr15: 72.635.500-72.642.500)
581 imprinted locus. G-stretches involved in predicted G4 formation (G4Hunter score 1.76
582 for *Meg3* and 1.52 for *Peg13* genes, orange bars) within the DMR are highlighted in
583 red below the track. Allelic G4 and RNA signals (normalized read count) are indicated
584 in the dashed rectangle for each of the strains and the maximum corresponds in each
585 case to the active allele.

586

587 **Figure 6: DNA methylation is antagonistic with G4access signal and with G4**
588 **associated nucleosome exclusion *in vivo* and *in vitro*.**

589 **a-** G4access signal is found at unmethylated DNA sites in K562 cells. Heatmaps
590 centered on G4 medium predictions (G4H1.2) containing at least 2 CpG are separated
591 in 2 groups: (Top) G4access peaks overlapping pG4s ranked by increasing G4access
592 signal n=16820; (Bottom) randomly selected pG4s (G4H>1.2) at sites that do not show
593 any G4access peak n=16820. Heatmaps show G4access and DNA methylation
594 (Whole Genome Bisulfite Sequencing, WGBS) signals. **b-** Metaprofiles of G4access
595 and methylation signals at pG4s (G4H1.2) overlapping or not with G4access peaks in
596 a window of +/-5kb. **c-** G4-associated nucleosome depletion is antagonistic with DNA
597 methylation. Heatmaps of nucleosome mapping by MNase-seq from *in vitro*
598 reconstituted chromatin, from native samples (granulocyte) and methylation (WGBS)
599 signals centered on pG4 annotations (G4H2, +/- 50bp) that contain at least 2 CpGs
600 are shown (n=52854). They are ranked by increasing *in vitro* MNase-seq signals. Six
601 groups of equal sizes (n=8809) were defined (right, colors are displayed). **d-**
602 Corresponding metaprofiles of the heatmaps and the 6 groups defined in **c** are shown.
603 A zoom-in at 100bp of the DNA methylation signals is also displayed.

604
605

606 References

607

- 608 1 Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances
609 through genomics. *Nat Rev Genet* **10**, 161-172, doi:10.1038/nrg2522 (2009).
- 610 2 Fenouil, R. *et al.* CpG islands and GC content dictate nucleosome depletion in a
611 transcription-independent manner at mammalian promoters. *Genome Res* **22**,
612 2399-2408, doi:10.1101/gr.138776.112 (2012).
- 613 3 Esnault, C. *et al.* G-quadruplexes are promoter elements controlling nucleosome
614 exclusion and RNA polymerase II pausing. *bioRxiv*, 2023.2002.2024.529838,
615 doi:10.1101/2023.02.24.529838 (2023).
- 616 4 Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability
617 and function of G-quadruplex structures. *Nat Rev Genet* **13**, 770-780,
618 doi:10.1038/nrg3296 (2012).
- 619 5 Hansel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in
620 the human genome: detection, functions and therapeutic potential. *Nat Rev Mol*
621 *Cell Biol* **18**, 279-284, doi:10.1038/nrm.2017.3 (2017).
- 622 6 Bedrat, A., Lacroix, L. & Mergny, J. L. Re-evaluation of G-quadruplex propensity
623 with G4Hunter. *Nucleic acids research* **44**, 1746-1759, doi:10.1093/nar/gkw006
624 (2016).
- 625 7 Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human
626 genome. *Nucleic acids research* **33**, 2908-2916, doi:10.1093/nar/gki609 (2005).
- 627 8 Hansel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory
628 chromatin. *Nat Genet* **48**, 1267-1272, doi:10.1038/ng.3662 (2016).
- 629 9 Li, C. *et al.* Ligand-induced native G-quadruplex stabilization impairs
630 transcription initiation. *Genome Res* **31**, 1546-1560, doi:10.1101/gr.275431.121
631 (2021).
- 632 10 Mao, S. Q. *et al.* DNA G-quadruplex structures mold the DNA methylome. *Nat*
633 *Struct Mol Biol* **25**, 951-957, doi:10.1038/s41594-018-0131-8 (2018).
- 634 11 Ray, S. *et al.* Custom DNA Microarrays Reveal Diverse Binding Preferences of
635 Proteins and Small Molecules to Thousands of G-Quadruplexes. *ACS Chem Biol* **15**,
636 925-935, doi:10.1021/acscchembio.9b00934 (2020).

- 637 12 Tran, P. L. T. *et al.* Folding and persistence times of intramolecular G-
638 quadruplexes transiently embedded in a DNA duplex. *Nucleic acids research* **49**,
639 5189-5201, doi:10.1093/nar/gkab306 (2021).
- 640 13 van Holde, K. & Zlatanova, J. Unusual DNA structures, chromatin and
641 transcription. *Bioessays* **16**, 59-68, doi:10.1002/bies.950160110 (1994).
- 642 14 Hershman, S. G. *et al.* Genomic distribution and functional analyses of potential G-
643 quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic acids*
644 *research* **36**, 144-156, doi:10.1093/nar/gkm986 (2008).
- 645 15 Dingwall, C., Lomonosoff, G. P. & Laskey, R. A. High sequence specificity of
646 micrococcal nuclease. *Nucleic acids research* **9**, 2659-2673,
647 doi:10.1093/nar/9.12.2659 (1981).
- 648 16 Horz, W. & Altenburger, W. Sequence specific cleavage of DNA by micrococcal
649 nuclease. *Nucleic acids research* **9**, 2643-2658, doi:10.1093/nar/9.12.2643
650 (1981).
- 651 17 Foulk, M. S., Urban, J. M., Casella, C. & Gerbi, S. A. Characterizing and controlling
652 intrinsic biases of Lambda exonuclease in nascent strand sequencing reveals
653 phasing between nucleosomes and G-quadruplex motifs around a subset of
654 human replication origins. *Genome Res*, doi:10.1101/gr.183848.114 (2015).
- 655 18 Luo, Y., Granzhan, A., Verga, D. & Mergny, J. L. FRET-MC: A fluorescence melting
656 competition assay for studying G4 structures in vitro. *Biopolymers*, ebp23415,
657 doi:10.1002/bip.23415 (2020).
- 658 19 Rodriguez, R. *et al.* Small-molecule-induced DNA damage identifies alternative
659 DNA structures in human genes. *Nat Chem Biol* **8**, 301-310,
660 doi:10.1038/nchembio.780 (2012).
- 661 20 Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex
662 structures in the human genome. *Nature biotechnology* **33**, 877-881,
663 doi:10.1038/nbt.3295 (2015).
- 664 21 Sun, H., Karow, J. K., Hickson, I. D. & Maizels, N. The Bloom's syndrome helicase
665 unwinds G4 DNA. *J Biol Chem* **273**, 27587-27592, doi:10.1074/jbc.273.42.27587
666 (1998).
- 667 22 Vaughn, J. P. *et al.* The DEXH protein product of the DHX36 gene is the major
668 source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell
669 lysates. *J Biol Chem* **280**, 38117-38120, doi:10.1074/jbc.C500348200 (2005).
- 670 23 Natoli, G. & Andrau, J. C. Noncoding transcription at enhancers: general principles
671 and functional models. *Annu Rev Genet* **46**, 1-19, doi:10.1146/annurev-genet-
672 110711-155459 (2012).
- 673 24 Xia, Y. *et al.* Transmission of dynamic supercoiling in linear and multi-way
674 branched DNAs and its regulation revealed by a fluorescent G-quadruplex torsion
675 sensor. *Nucleic acids research* **46**, 7418-7424, doi:10.1093/nar/gky534 (2018).
- 676 25 Esnault, C. *et al.* G-quadruplexes are promoter elements controlling nucleosome
677 exclusion and RNA polymerase II pausing. *Manuscript in review to Cell and*
678 *Molecular Cell* (2023).
- 679 26 Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its
680 interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**, e02407,
681 doi:10.7554/eLife.02407 (2014).
- 682 27 Medlin, J. *et al.* P-TEFb is not an essential elongation factor for the intronless
683 human U2 snRNA and histone H2b genes. *Embo j* **24**, 4154-4165,
684 doi:10.1038/sj.emboj.7600876 (2005).

- 685 28 Shen, J. *et al.* Promoter G-quadruplex folding precedes transcription and is
686 controlled by chromatin. *Genome Biol* **22**, 143, doi:10.1186/s13059-021-02346-7
687 (2021).
- 688 29 Cusack, M. *et al.* Distinct contributions of DNA methylation and histone
689 acetylation to the genomic occupancy of transcription factors. *Genome Res* **30**,
690 1393-1406, doi:10.1101/gr.257576.119 (2020).
- 691 30 Vaid, R., Wen, J. & Mannervik, M. Release of promoter-proximal paused Pol II in
692 response to histone deacetylase inhibition. *Nucleic acids research* **48**, 4877-4890,
693 doi:10.1093/nar/gkaa234 (2020).
- 694 31 Kelsey, G. & Feil, R. New insights into establishment and maintenance of DNA
695 methylation imprints in mammals. *Philos Trans R Soc Lond B Biol Sci* **368**,
696 20110336, doi:10.1098/rstb.2011.0336 (2013).
- 697 32 Sanli, I. *et al.* Meg3 Non-coding RNA Expression Controls Imprinting by
698 Preventing Transcriptional Upregulation in cis. *Cell Rep* **23**, 337-348,
699 doi:10.1016/j.celrep.2018.03.044 (2018).
- 700 33 Valouev, A. *et al.* Determinants of nucleosome organization in primary human
701 cells. *Nature* **474**, 516-520, doi:10.1038/nature10002 (2011).
- 702 34 Capra, J. A., Paeschke, K., Singh, M. & Zakian, V. A. G-quadruplex DNA sequences
703 are evolutionarily conserved and associated with distinct genomic features in
704 *Saccharomyces cerevisiae*. *PLoS Comput Biol* **6**, e1000861,
705 doi:10.1371/journal.pcbi.1000861 (2010).
- 706 35 Fu, Y., Sinha, M., Peterson, C. L. & Weng, Z. The insulator binding protein CTCF
707 positions 20 nucleosomes around its binding sites across the human genome.
708 *PLoS Genet* **4**, e1000138, doi:10.1371/journal.pgen.1000138 (2008).
- 709 36 Lay, F. D. *et al.* The role of DNA methylation in directing the functional
710 organization of the cancer epigenome. *Genome Res* **25**, 467-477,
711 doi:10.1101/gr.183368.114 (2015).
- 712 37 Ozcan, K. A., Ghaffari, L. T. & Haeusler, A. R. The effects of molecular crowding and
713 CpG hypermethylation on DNA G-quadruplexes formed by the C9orf72 nucleotide
714 repeat expansion. *Sci Rep* **11**, 23213, doi:10.1038/s41598-021-02041-4 (2021).
- 715 38 Stevens, A. J. *et al.* G-quadruplex structures and CpG methylation cause drop-out
716 of the maternal allele in polymerase chain reaction amplification of the imprinted
717 MEST gene promoter. *PLoS One* **9**, e113955, doi:10.1371/journal.pone.0113955
718 (2014).
- 719 39 Wang, Z. F. *et al.* Cytosine epigenetic modification modulates the formation of an
720 unprecedented G4 structure in the WNT1 promoter. *Nucleic acids research* **48**,
721 1120-1130, doi:10.1093/nar/gkz1207 (2020).
- 722 40 Valton, A. L. & Prioleau, M. N. G-Quadruplexes in DNA Replication: A Problem or a
723 Necessity? *Trends Genet* **32**, 697-706, doi:10.1016/j.tig.2016.09.004 (2016).
- 724 41 Mendoza, O., Bourdoncle, A., Boule, J. B., Brosh, R. M., Jr. & Mergny, J. L. G-
725 quadruplexes and helicases. *Nucleic acids research* **44**, 1989-2006,
726 doi:10.1093/nar/gkw079 (2016).
- 727 42 Asamitsu, S., Obata, S., Yu, Z., Bando, T. & Sugiyama, H. Recent Progress of
728 Targeted G-Quadruplex-Preferred Ligands Toward Cancer Therapy. *Molecules* **24**,
729 doi:10.3390/molecules24030429 (2019).

730
731
732
733

Methods

734 Our research complies with French/European ethical policies and did not require a
735 specific board to approve our study.

736

737 **Cell culture and treatments**

738 K562 (human myelogenous leukemia, gift of E. Soler) and Raji (human Burkitt's
739 lymphoma, gift from D. Eick) cell lines were cultured in RPMI 1640 medium (Thermo
740 Fisher 11875085) supplemented with 10% heat-inactivated fetal bovine serum (Sigma
741 12103C), 100 units/mL penicillin, and 100 µg/mL streptomycin (Thermo Fisher
742 15140122) at 37°C in a humidified 5% CO₂ incubator. Human Raji cells, grown at an
743 approximative density of 10⁶ cells/ml, were treated with 10 µM pyridostatin (PDS;
744 Sigma SML0678) or H₂O (control) for 30 minutes, with 2 µM TSA (Sigma, T8552) or
745 DMSO (control) for 24 hours for G4access experiment. TSA efficiency was tested by
746 western-blot on H3K9ac histone acetylation (Ab Millipore, CS200583). For
747 transcription inhibition experiments cells were treated for 2 hours with DMSO 0.1%
748 (control), 1 µM triptolide (inhibiting initiation stage; Sigma Merck) or 100 µM KM05283
749 (inhibiting elongation stage; Thermo Fisher Scientific) prior collection for G4access or
750 Pol II ChIP-seq. HaCaT (spontaneously immortalized human keratinocyte line, gift from
751 the IRCM screening platform in oncology) was cultured in DMEM-high glucose medium
752 (Thermo Fisher 11965084) supplemented with 10% heat-inactivated fetal bovine
753 serum (Sigma 12103C), 100 units/mL penicillin, and 100 µg/mL streptomycin (Thermo
754 Fisher 15140122) at 37°C in a humidified 5% CO₂ incubator.

755 HeLa cells (gift from E. Bertrand) used for siRNA knockdown assays were maintained
756 in DMEM supplemented with 10% fetal calf serum, penicillin/streptomycin (100 units/L)
757 and glutamin (2.9 mg/L) at 37°C and 5% CO₂. Knockdown experiments were
758 performed using Invitrogen siRNA Silencer® Select (siDHX36 and siWRN, 4392420)
759 control siRNA (Dharmacon, [ON-TARGETplus Non-targeting Pool](#)) and lipofectamine
760 RNAiMAX (Thermo Fisher Scientific) following manufacturer instructions. Knockdown
761 efficiencies were analysed by westernblot using specific antibodies (DHX36: Abcam,
762 ab70269, WRN: Sigma, W0393).

763

764 The 2i-medium-derived ESCs hybrid between *M. m. domesticus* strain C57BL/6J and
765 *M. m. molossinus* strain JF1 were derived recently³². The two chosen lines, BJ (full
766 laboratory name BJ-WT3) and JB (full laboratory name JB-WT2), are both male, both
767 with a normal karyotype. They were cultured for cell harvesting on 0.1% gelatin-coated
768 dishes (Merck-Millipore SF008) in serum-free ESGRO Complete PLUS medium
769 (Millipore, with LIF and GSK3 inhibitor, SF001) at 37°C in a humidified 5% CO₂
770 incubator.

771

772 Drosophila S2 cells (gift from J. Tazi) were cultured in Schneider's S2 Drosophila
773 medium (Dominique Dutscher, L0207-500) supplemented with 10% heat-inactivated
774 fetal bovine serum (Sigma, 12103C), 100 units/mL penicillin, and 100 µg/mL
775 streptomycin (Thermo Fisher, 15140122) at 27°C and harvested at 2x10⁶ cells/mL.
776 Yeast S288C (BY4741, gift from M. Radman-Livaja) haploid cells were grown up to
777 and collected at mid-log phase (OD₆₀₀=0.5) in 100mL of YPD (2% glucose).

778

779 **G4access**

780 For adherent HeLa, HaCaT and mES cells, cells were first trypsinized and then
781 pelleted, while suspension (K562, Raji) cells were directly pelleted by centrifugation at
782 413 rcf 10 min at 4 °C. For S2 semi-adherent cells, cells were resuspended by pipetting
783 up and down multiple times. All cell pellets were rinsed twice in phosphate-buffered

784 saline buffer (PBS). HaCaT cells were further subjected to mechanical
785 permeabilization in 5 mL of permeabilization buffer (150 mM of sucrose, 80 mM KCl,
786 5 mM KH_2PO_4 , 5 mM MgCl_2 , 0.5 mM CaCl_2 and 35 mM HEPES pH 7.4) and
787 homogenized using a dounce (40 strokes). HeLa cells tend to clump under such
788 condition. Therefore, nuclei were first purified under sucrose cushion as described⁴³.
789 For this, cells were resuspended on ice-cold nuclei buffer#1 (15 mM Tris-HCl pH 7.5,
790 300 mM sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl_2 , 0.1 mM ethylene glycol-bis
791 N,N,N',N'-tetra-acetic acid (EGTA), 0.5 mM dithiothreitol (DTT), 0.1 mM
792 phenylmethylsulfonyl fluoride (PMSF), 3.6 $\mu\text{g}/\text{ml}$ aprotinin) before addition of ice-cold
793 buffer#2 (15 mM Tris-HCl pH 7.5, 300 mM sucrose, 60 mM KCl, 15 mM NaCl, 5 mM
794 MgCl_2 , 0.1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 3.6 $\mu\text{g}/\text{ml}$ aprotinin, 0.4% IGEPAL
795 CA-630) and then were carefully layered on ice-cold nuclei buffer #3 (15 mM Tris-HCl
796 pH 7.5, 1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl_2 , 0.1 mM EGTA, 0.5 mM
797 DTT, 0.1 mM PMSF, 3.6 $\mu\text{g}/\text{ml}$ aprotinin). Cells were then centrifuged at 10.000g for
798 20 min at 4 °C. The supernatants were discarded and the nuclei-containing pellets
799 were resuspended in MNase digestion buffer for later steps (see below). For all
800 mammalian cell lines, 5×10^6 cells per titration points were then re-suspended in 50 μL
801 of prewarmed permeabilization buffer supplemented with 0.2% (v/v) NP40 and
802 incubated for 5 minutes at 37°C prior digestion. The same procedure was performed
803 for 50×10^6 cells for S2 Drosophila cells. MNase digestions, were then performed by
804 adding a volume of 500 μL of prewarmed MNase reaction buffer (150 mM sucrose, 50
805 mM Tris-HCl pH 8, 50 mM NaCl and 2 mM CaCl_2) supplemented with either 3, 6, 12,
806 25 or 50U of MNase (Merck, 10107921001). Digestions were incubated at 37°C for 10
807 min and stopped on ice and by adding 11 μL of 500 mM EDTA to each reaction
808 followed by addition of 550 μL of SDS lysis buffer (1% (v/v) SDS, 10 mM EDTA and
809 50 mM Tris.HCl pH 8). Samples were then incubated 10 minutes on ice. Before DNA
810 purification, 1 mL of water was added to dilute the SDS and the samples were
811 incubated with 5 μL of RNase A (ThermoFisher, EN0531) at 37 °C for 2 hours and with
812 8 μL of proteinase K (Euromedex, 09-0911) at 56 °C for 2 hours to complete the lysis.
813 To then quality control the MNase digestions: 125 μL of each sample were cleaned-up
814 using QIAquick PCR Purification Kit (QIAGEN, 28106) and assessed by agarose gel
815 and Bioanalyzer 2100, using High-sensitivity DNA Chips (Fig. 1c). At this step, for
816 efficient G4access, samples should present ~30% (+/-5%) of mono-nucleosomes (Fig.
817 1c). Importantly, this assessment should be performed on purified DNA that does not
818 contain the subnucleosomal fraction, using a bioanalyzer equipment. The remaining of
819 the samples was then purified by phenol-chloroform and ethanol precipitation for
820 subsequent steps. We recommend that, when implementing this method, a wide range
821 of MNase concentrations shall be tested in a first round of preparative experiments to
822 narrow the condition in which the critical fraction of 30% of mononucleosome, as
823 compared to the whole nucleosomal fraction, shall be obtained. Our experiences
824 showed this fraction is on average optimal for best G4 sequence recovery. We also
825 recommend that the bioanalyser and agarose gel QCs shall be used to assess that
826 MNase digestion and lysis are as complete as possible. Whenever chromatin is not
827 properly digested a smear occurs on top of the digestion pattern (high molecular
828 weights from non-permeabilized nuclei) that does not get resolved by increasing
829 digestion time or MNase concentration. Such samples are typically discarded and it is
830 preferable in this case to check that cell lysis is as complete as possible under the
831 microscope so that all nuclei can be digested similarly by the enzyme.
832

833 For yeast, a distinct procedure was used. Cells were pelleted and rinsed twice in
834 phosphate-buffered saline buffer (PBS). Pellets from 100mL culture were suspended
835 in 600 μ l of cell breaking buffer (20% glycerol, 100 mM Tris pH 7.5), 600 μ l zirconia
836 beads (0.5mm), 10 μ l of 100x protease inhibitors (Roche, 06538282001). Beads
837 beating was performed in 1.5mL screw cap tubes in a Bullet Blender (Next advance)
838 for 4 x 3min at a strength of 8 at 4°C. Cell suspensions were recovered by
839 centrifugation; a tube was punctured (23 gauge syringes) and the nuclei were collected
840 in collection tube by centrifugation at 170 rcf. The nuclei fraction was then centrifuged
841 for 5min at 20G and the pellets which contained the nuclei were resuspended in 300 μ l
842 of prewarmed NP buffer (50 mM NaCl, 10 mM Tris pH 7.4, 5 mM MgCl₂, 1 mM CaCl₂,
843 0.2% NP40 (v/v), 0.5 mM spermidine (Sigma, S0266-1G), 0.007% β -Mercaptoethanol
844 (Sigma, M3148-100ML). Digestion took place by adding a volume of 300 μ L prewarmed
845 NP buffer supplemented with either 60u, 30u, 15u, 7.5u, 3.75u, 1.9u, or 1u of MNase
846 (Merck 10107921001). Digestions were stopped on ice and by adding 150 μ L of stop
847 buffer (5% SDS, 50 mM EDTA). Before DNA purification, digestions were incubated
848 with 5 μ L of RNase A (ThermoFisher, EN0531) at 37 °C for 2 hours and with 10 μ L of
849 proteinase K (Euromedex, 09-0911) at 56°C for 2 hours. Purification was performed
850 with two consecutive phenol and one chloroform steps followed by ethanol and linear
851 acrylamide precipitation. Importantly, purified DNAs were once again incubated with
852 5 μ l RNase A (ThermoFisher, EN0531) to get rid of persistent RNA contaminations. As
853 for other cells described above, a criterion of 30% mononucleosome was used for the
854 choice of the subnucleosomal fraction to be sequenced. We note that the absence of
855 K⁺ ions in this yeast-specific procedure might impact an optimal G4 recovery.

856
857 The phenol-chloroform purified DNAs were subjected to size selection to select
858 fragments under 100 bp. For that, 1 μ g of each digestion product was migrated in a 4-
859 20% polyacrylamide Novex™ TBE gel (ThermoFisher, EC6225BOX) at 100 V for 60
860 min. The gels were stained with Syber® Gold (ThermoFisher, S11494) for 30 min.
861 Fragments of 0-100 bp were cut out from the gel and transferred to 0.5-mL Eppendorf
862 tubes, previously punctured in the bottom with a 0.45 μ m needle. These tubes were
863 inserted into 1.5-mL collection tubes and centrifuged 10 min at 15.300 rcf to collect the
864 gel through the hole, generating gel beads. To elute the DNA from the beads, 700 μ L
865 of water was added and the tubes were incubated overnight at 55 °C in a thermomixer
866 at 1500 rpm. DNA was purified by transferring the samples (DNA eluate and gel beads)
867 to the top of a 0.22 μ m spin filter (Agilent 5185-5990). Spin filters were centrifuged 2
868 min at 15.300 rcf to recover the DNA eluate. DNA was precipitated with isopropanol
869 and linear acrylamide. Size-selection was verified by Bioanalyzer. Relative amount of
870 targeted G4s was also evaluated by qPCR (Extended Data Fig. 1b). The titration points
871 showing a percentage of mononucleosomal fraction of 30% of the total DNA (excluding
872 subnucleosomal fraction) gave the best qPCR enrichment of targeted G4s after size
873 selection. This observation was further confirmed when sequencing the corresponding
874 libraries.

875
876 **G4access library preparation**
877 The 0-100 bp size-selected fragments from MNase digestions that have ~30% of
878 mono-nucleosomes were subjected to DNA library preparation. In parallel, genomic
879 DNA libraries were sonicated by Bioruptor® Pico sonicator (Diagenode) to obtain DNA
880 fragments of ~150 bp to be used later as reference data sets for bioinformatic analyses.
881 Paired-end libraries were constructed using NEBNext® Ultra™ II DNA Library Prep Kit
882 for Illumina (New England Biolabs, E7645S) using a starting material of 50 ng. DNA

883 fragments were treated with end-repair, A-tailing and ligation of Illumina-compatible
884 adapters. Clean-up of adaptor-ligated DNA was performed by using CleanNGS beads
885 (CNGS-0050) with a bead:DNA ratio of 2:1. The purified products were amplified with
886 8 cycles of PCR. Finally, samples were cleaned up with a bead:DNA ratio of 0.8:1 to
887 remove the free sequencing adapters. Human and mouse libraries were sequenced
888 on the Illumina NextSeq-500 Sequencer using paired 75 bp reads (Raji), or paired 50-
889 30 bp reads (K562, mES cells) or an Illumina HS4000 using paired 76bp reads (HaCaT
890 cells). Libraries from HeLa, Raji cells treated with TSA, Drosophila and Yeast cells
891 were sequenced on Novaseq 6000 Sequencer in paired-end (50-50bp) sequencing
892 runs.

893

894 **ChIP-seq**

895 Fifty million cells were used to prepare extracts for Pol II ChIP-seq experiments. Cells
896 were crosslinked for 10 min at 20°C with the crosslinking solution (10 mM NaCl, 0.1
897 mM EDTA pH 8, 0.05 mM EGTA pH 8, 5 mM HEPES pH 7.8 and 1% formaldehyde).
898 The reaction was stopped by adding glycine to reach a final concentration of 250 mM.
899 After 5 min of formaldehyde quenching, cells were washed twice with cold PBS and
900 resuspended in cold 2.5 mL LB1 (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA
901 pH 8, 10% glycerol, 0.75% NP-40, 0.25% Triton X-100) at 4°C for 20 min on a rotating
902 wheel. Nuclei were pelleted down by spinning at 1350 rcf in a refrigerated centrifuge
903 and washed in 2.5 mL LB2 (200 mM NaCl, 1 mM EDTA pH 8, 0.5 mM EGTA pH 8, 10
904 mM Tris pH 8) for 10 min at 4°C on a rotating wheel followed by centrifugation to collect
905 nuclei. Nuclei were then resuspended in 1mL LB3 (1 mM EDTA pH 8, 0.5 mM EGTA
906 pH 8, 10 mM Tris pH 8, 100 mM NaCl, 0.1% Na-Deoxycholate, 0.5% N-
907 lauroylsarcosine) and sonicated using Bioruptor Pico (Diagenode) in 15-mL tubes for
908 24 cycles of 30 s ON and 30 s OFF pulses in 4°C bath. All buffers (LB1, LB2 and LB3)
909 were complemented with EDTA free Protease inhibitor cocktail (Roche), 0.2 mM PMSF
910 just before use. After sonication, Triton X-100 was added to a final concentration of 1%
911 followed by centrifugation at 20000 rcf and 4°C for 10 min to remove particulate matter.
912 After taking aside a 50 µL aliquot to serve as input and to analyze fragmentation,
913 chromatin was aliquoted and snap-frozen in liquid nitrogen and stored at - 80°C until
914 use in ChIP assays. Input aliquots were mixed with an equal volume of 2X elution
915 buffer (100 mM Tris pH 8.0, 20 mM EDTA, 2% SDS) and incubated at 65°C for 12
916 hours for reverse-crosslinking. An equal volume of TE buffer (10 mM Tris pH 8 and 1
917 mM EDTA pH 8) was added to dilute the SDS to 0.5% followed by treatment with
918 RNase A (0.2µg/mL) at 37°C for one hour and Proteinase K (0.2 µg/L) for two hours at
919 55°C. DNA was isolated by phenol:chloroform:isoamylalcohol (25:24:1 pH 8)
920 extraction followed by Qiaquick PCR Purification (QIAGEN, Germany). Purified DNA
921 was then analyzed on a 1.5% agarose gel and on Bioanalyzer (Agilent, USA) using a
922 High Sensitivity DNA Assay.

923 For Pol II ChIP, Protein-G coated Dynabeads were incubated at 4°C in blocking
924 solution (0.5% BSA in PBS) carrying Pol II F12 (Santa Cruz 10 sc-55492, lot H2019)
925 specific antibodies. Sonicated chromatin (10x10⁶ Raji cells equivalent and 5x10⁶
926 drosophila S2 cells equivalent (2:1 spike-in ratio)) was added to pre-coated beads (250
927 µL) and the mix was incubated overnight at 4°C on a rotating wheel. After incubation
928 with chromatin, beads were washed 7 times with Wash buffer (50 mM HEPES pH 7.6,
929 500 mM LiCl, 1 mM EDTA pH 8, 1% NP-40, 0.7% Na-Deoxycholate, 1X protease
930 inhibitor cocktail) followed by one wash with TE-NaCl buffer (10 mM Tris pH 8 and 1
931 mM EDTA pH 8, 50 mM NaCl) and a final wash with TE buffer (10 mM Tris pH 8 and

932 1 mM EDTA pH 8). Immunoprecipitated chromatin was eluted by two sequential
933 incubations with 50 μ L Elution buffer (50 mM Tris pH 8, 10 mM EDTA pH 8, 1% SDS)
934 at 65°C for 15 min. The two eluates were pooled and incubated at 65°C for 12 hours
935 to reverse-crosslink the chromatin followed by treatment with RNase A and Proteinase
936 K and purification of DNA, as described above for input samples. Both input and ChIP
937 samples were subjected to Bioanalyzer analysis to check that the major bulk of isolated
938 DNA was in the 250 bp size range.

939 For ChIP-seq experiments in Raji cells, purified DNA was quantified with Qubit DS
940 DNA HS Assay (ThermoFisher Scientific, USA). 1 ng of ChIP DNA were used to
941 prepare sequencing libraries with NEBNext® Ultra™ II DNA Library Prep Kit for
942 Illumina (New England Biolabs, E7645S). After end-repair and adaptor ligation, library
943 fragments were amplified by 13 cycles of PCR. Barcoded libraries from different
944 samples were pooled together and sequenced on Illumina NextSeq500 Sequencer in
945 paired-end (50-30bp) sequencing runs.

946

947 **Nascent Chr-RNA-seq**

948 Chromatin-associated RNAs (ChrRNAs) presented in ED Figure 3g were either
949 analyzed from GSE90238 or isolated from 2×10^7 Raji cells as follows. Nuclear
950 fractionation was performed by incubating cells in 4mL of Buffer I (10mM Tris-HCl pH
951 7.5, 10 mM NaCl, 2.5mM MgCl₂, 0.5% IGEPAL CA-630) on ice for 5 minutes. Next,
952 we carefully underlaid 1 mL of Buffer II (10mM Tris-HCl pH 7.5, 10 mM NaCl, 2.5mM
953 MgCl₂, 0.5% IGEPAL CA-630, 10% sucrose) before harvesting the nuclear fraction at
954 1400 rcf for 5 minutes at 4°C. Nuclei were resuspended with 125 μ L of NUN1 buffer
955 (20mM Tris-HCl pH 8, 75mM NaCl, 0.5mM EDTA, 50% Glycerol) followed by 1.2mL of
956 NUN2 buffer (20mM HEPES-KOH pH 7.6, 300mM NaCl, 0.2mM EDTA, 7.5mM MgCl₂,
957 1% IGEPAL CA-630, 1M Urea). After 15 minutes of vigorous vortexing, the chromatin
958 was centrifuged at 15000 rcf for 10 minutes at 4°C and washed with 500 μ L of NUN2
959 buffer. After discarding the supernatant, the chromatin was resuspended in 500 μ L of
960 Trizol. At this stage, the chromatin pellet is very tight and needs to be dissolved in
961 Trizol by repeated pipetting with decreasing volume tips (1mL-200 μ L-10 μ L) then
962 pushing through very small syringe needles. Then RNA was extracted from chromatin
963 according to the Trizol manufacturer's protocol and resuspended in 50 μ L of nuclease-
964 free water (Invitrogen, ref#AM9906) followed by TurboDNase (Invitrogen,
965 ref#AM2238) treatment. Trizol RNA extraction and TurboDNase were performed two
966 more times and RNA was resuspended in 20 μ L of nuclease-free water. Purified RNAs
967 were quantified by Qubit and quality was assessed using the RNA Assay kit (Agilent
968 RNA 6000 Pico reagents, ref#1567-1513) with Bioanalyzer 2100 (Agilent
969 Technologies, USA). ChrRNAs were then subjected to library preparation using the
970 True-seq stranded total RNA library prep gold kit (Illumina, ref#220599) using 1 μ g of
971 ChrRNA, with a total of 15 cycles of amplification and following the manufacturer's
972 instructions (including ribo-depletion).

973

974 **Quantitative PCR analysis of targeted G4s**

975 For the relative quantification of targeted G4s, human cells Raji, K562 and HaCaT cells
976 were used. Briefly, the DNA was isolated by phenol-chloroform and ethanol
977 precipitation after MNase treatment and size selection. The relative amount of targeted

978 G4s related to its corresponding non-size selected sample was evaluated by qPCR by
979 using the delta-Cq method using the following primers for the G4 at the human MAZ
980 locus (chr16): G4_Maz_F ACTGAGCGCAGGATTGTAAATA and
981 CCTCATGCTTCGGCTTCC and control primers at the KRAS locus (chr12):
982 Control_NEG-1_F TAAACCAGGGCTGCTGTTCT and Control_NEG-1_R
983 TGACCGCAAAGCTGTTACAC. Quantitative PCR reaction was performed using the
984 Platinum® Taq DNA polymerase (ThermoFisher 11304011) following the
985 manufacturer's instructions on a LightCycler® 480 system. qPCR reactions were
986 performed in triplicates. Cycling conditions were 95°C for 10 min., then 40 cycles of
987 95°C for 30 sec./65°C for 30 sec./72°C for 15 sec., followed by melting curve analysis.
988 Results presented on Extended Data Fig. 1c are displayed as the ratio of enrichment
989 of the G4 to non G4 regions.

990

991 **FRET melting profiles**

992 FRET melting profile assays were performed on 6 G-quadruplexes and one hairpin to
993 test their stabilities in MNase digestion conditions. FRET buffers were lithium
994 cacodylate supplemented with 140mM KCl and 10mM KCl with 45mM NaCl for
995 physiological and NMase conditions, respectively. The sequences tested,
996 F21CTAT (GGGCTAGGGCTAGGGCTAGGG),
997 EBR1 (GGGCAGGGGGTGATGGGGAGGAGCCAGGG),
998 F21T (GGGTTAGGGTTAGGGTTAGGG),
999 F25cebT (AGGGTGGGTGTAAGTGTGGGTGGGT),
1000 FAG3AT (AGGGAAGGGAAGGGAAGGGA),
1001 FmycT (TTGAGGGTGGGTAGGGTGGGTAA), and
1002 FdxT (TATAGCTAT-PEG-TATAGCTATA) were double labeled with FAM and TAMRA
1003 to follow their unfolding by fluorescence. All G4s are highly stable in the MNase buffer
1004 condition at room temperature or at 37 degrees. Most differences in their stabilities are
1005 observed at higher temperatures.

1006

1007 **FRET-MC, Th-T and NMM massive G4 validation assays *in vitro***

1008 Target selection for the design of 596 oligonucleotides was performed by first
1009 overlapping human G4access peaks common to HaCaT, K562 and Raji cell lines. This
1010 allowed the isolation of 4743 regions of various sizes (Extended Data Fig. 2B) from
1011 which we extracted the maximum G4Hunter score, using a window of 30 nt within the
1012 G4Hunter algorithm. To perform our large-scale *in vitro* assays, we subsequently
1013 selected 596 fragments from this list, with a score distribution comparable to that of the
1014 initial pool (not shown). The list of peaks with genomic coordinates and
1015 oligonucleotides is presented in Supplementary Tables 2 and 3. The library of
1016 oligonucleotides was synthesized and purchased at Eurogentec (Seraing, Belgium)
1017 with RP cartridge gold purification, and further used for FRET-MC, ThT and NMM.
1018 FRET-MC assay was performed in 96-well plates and the fluorescence of the labeled
1019 oligonucleotide F21T was recorded using a CFX96 qPCR instrument (Biorad).
1020 Oligonucleotides were annealed at 7.5 μM strand concentration (95°C, 5min) in FRET
1021 buffer (10 mM KCl, 10 mM lithium cacodylate, 90 mM LiCl, pH 7.2). The tested
1022 oligonucleotides and the F21T were added to each well (final concentration of 3 μM
1023 and 0.2 μM, respectively), which were incubated with or without 0.4 μM of phenDC3 in
1024 a final volume of 25 μL. The microplate was incubated at 25°C for 5 min, after which
1025 the temperature was increased by increments of 0.5°C per minute to reach 95°C. The
1026 collected signal was normalized to 1 and the melting temperature (T_m) was defined
1027 when the normalized signal was 0.5. ΔT_m corresponds to the difference of T_m between

1028 the oligonucleotide with and without PhenDC3. Each sequence was tested in an intra-
1029 day duplicate.

1030
1031 The fluorescence assay was performed in 96-well plates, using a plate reader M1000
1032 Pro (TECAN). Fluorescence emission was read at 490 and 610 nm after excitation at
1033 420 and 380 nm for thioflavin T (ThT) and NMM, respectively. Oligonucleotides were
1034 annealed at 7.5 μ M in K100 buffer (100 mM KCl, 10 mM LiCaco, pH 7.2). To each well
1035 in a 96 well plate, 3 μ M of oligonucleotide and 2 μ M of fluorescence ligands were
1036 added, reaching a total volume of 100 μ L. The plate was then shaken and incubated
1037 for 10 min. Each oligonucleotide was tested in an intra-day duplicate. The threshold of
1038 positive or negative sequences were determined using different controls (G4s with
1039 different topologies, duplex, singles strands). All fluorescence results (including the
1040 controls) were normalized to the number of nucleotides. A summary of the results of
1041 all experiments is presented in Supplementary Table 1.

1042 1043 **Bioinformatic procedures**

1044 1045 ***Genomic data set processing, peak calling and differential analysis***

1046 For both native and re-analyzed published data sets considered in this study, raw
1047 sequencing reads were aligned using Bowtie2⁴⁴ (version 2.1.0) to the human (hg19),
1048 mouse (mm9), drosophila (dm6) and yeast (sc3) genomes. Aligned reads were
1049 elongated *in silico* using the DNA fragment size inferred using an in-house developed
1050 PASHA (version 0.99.21) R (version 3.3.1) pipeline⁴⁵ or using MACS2⁴⁶ (version 2.1.2)
1051 which also allows peak calling for G4-ChIP and G4access (peaks were considered
1052 below a *p*-value of 10^{-10} from the narrow peaks table). PASHA was used for ChIP-seq
1053 and MNase-seq datasets, using drosophila spike-in for ChIP normalization⁴⁷ (Fig. 4d),
1054 and MACS2 was used for G4access and G4 ChIP for the sake of consistency in
1055 comparison with previously published G4 ChIP analyses. MACS2 was run using input
1056 DNA as control and with recommended settings⁴⁶. Bedgraph files generated by
1057 MACS2 were then converted to wig files (bin10) and scaled using the sequencing
1058 depth with PASHA. Wiggle files representing average enrichment score every 10bp
1059 were generated. For nucleosome positioning analyses (midpoints), to determine the
1060 average nucleosome positions, wiggle files representing the central nucleotides of
1061 DNA fragments were also generated (Fig. 1a, 4a and Extended Data Fig. 6a, 7a, 9c).
1062 Finally, for nucleosome densities representation and analyses (MNase-seq), we
1063 smoothed the signal by replacing each 10bp bin by the average of the 5 surrounding
1064 bins on each side and using an in-house script.

1065 To assess differences in G4access peak intensities, DESeq⁴⁸ (version 1.26.0) was
1066 used with the MACS2 peak definition as genomic references. Differences between the
1067 conditions (control versus treated cells) were called at a *p*-value below 0.05.

1068 RNA-seq datasets analyzed or re-analyzed in this study were processed by aligning
1069 sequencing reads to mouse genome (mm9) or human genome (hg19) using TopHat2
1070 for RNA-seq and Chr-RNA-seq. Alignment files were then treated using Cufflinks
1071 (v2.2.1) to generate RPKM used in ED Fig. 3g. SNP analyses for mouse RNA-seq data
1072 from Fig. 5a are described further below.

1073 ***Motif search analysis and NDRs determination***

1074 To analyze motifs associated to open chromatin/NDRs at promoters, we focused on
1075 CGI-containing promoters since those display an established intrinsic property to

1076 exclude nucleosomes². NDRs and deepest NDR points were defined using an in-
1077 house script. First by creating an inversed track of the MNase-seq signal ($y=1/x$, 0
1078 values were replaced by the minimal value found in the region -300/+100 of annotated
1079 TSSs in the sequencing signals). Then, peaks and peak summits corresponding to
1080 NDRs and their deepest points were determined. The peak-calling was performed
1081 using wigpeakcaller⁴⁹ (<https://doi.org/10.7554/eLife.02105.022>) fixing a threshold
1082 based on the peak height and the gap between two adjacent signals (Fig. 1a and
1083 Extended Data Fig. 7c). *De novo* motif discovery analyses were thus performed at the
1084 lowest nucleosome density (deepest point of MNase-seq signal) at promoters
1085 overlapping CGIs between -500 to TSS (Fig. 1a) and on G4access or G4-ChIP
1086 (Extended Data Fig. 4) datasets using MEME-ChIP⁵⁰ (default settings, version 5.0.2).
1087 Fragments from -30 to +30 bp centered on the deepest point of MNase-seq signal or
1088 peak summits for the other datasets were used and tested using the jaspardb 2020 core
1089 non redundant database. Fragments overlapping the annotated TSS region (-200 bp
1090 to the TSS) were defined using Bedtools (version 2.21.0). For yeast datasets
1091 (Extended Data Fig. 10a), TSS determined by CAGE were used
1092 (<http://www.yeastss.org/download/>). The First 3 motifs are displayed ranked by site
1093 numbers (Fig. 1a, Extended Data Fig. 4 and 10c).

1094 ***Allelic data analysis of mESC***

1096 Raw sequencing reads were aligned strictly to the mm9 or JF1⁵¹ genomes using
1097 Bowtie2. Reads overlapping distinguishable SNPs
1098 (<https://molossinus.brc.riken.jp/mogplus/#JF1>) between the two genomes were
1099 considered to assess allelic signals of G4 access and RNA-seq datasets. At these loci,
1100 reads overlapping SNP from each allele were determined and counted using HTSeq-
1101 count (version 0.6.1p1) at G4access peaks (for differential G4access analysis) and
1102 within gene annotations (for differential gene expression analysis). Assessment of
1103 statistical analyses of differential signals were performed using DESeq⁴⁸ (version
1104 1.26.0) (Fig. 5b-d; Supplementary Table 1).

1105 The list of 31 murine imprinted DMRs, all corresponding to CGIs, was in-house curated
1106 by the Feil laboratory and is largely based on previous reports by others⁵²⁻⁵⁴ together
1107 with Feil laboratory's own compilation of imprinted DMRs. Mm9 was used as a
1108 reference genome. The list of these annotations with genomic coordinates is provided
1109 in Supplementary Table 1.

1110 ***G4 motif and enrichment analysis.***

1111 Peak size for all G4 ChIP and G4access peaks detected by MACS2 were analyzed
1112 and the distribution plots were generated using R v3.3.1 (Fig. 1e). G4 scores of all
1113 peaks were determined using G4Hunter^{6,55} (see also below) and their distributions
1114 plotted using R (Fig. 1g, 3d, 3h, Extended Data Fig. 2d-e, 6c-d, 7c, 10b). To compare
1115 G4access and G4-ChIP, all peaks were resized at 90 bp (+/- 45 bp from peak summits)
1116 before G4Hunter score determination (Extended Data Fig. 2d-e).

1117
1118 G4Hunter scores of all experiments were tested against fragments of the same size
1119 distribution randomly selected in the genomes (10 permutations), thus allowing the
1120 evaluation of the observed versus expected from random selection scores (Fig. 1f).
1121 Finally, G4 motifs were analyzed using a published code to assign G4 classes⁸: Loop
1122 size 1–3, 4–5 and 6–7; sequences with at least one loop of the respective length; long

1123 loop: sequences with a G4 with any loop of length >7 (up to 12 for any loop and 21 for
1124 the middle loop); simple bulge: sequences with a G4 with a bulge of 1–7 bases in one
1125 G-run or multiple 1-base bulges; 2-tetrads/Complex bulge: sequences with a G4s with
1126 two G-bases per G-run or several bulges of 1–5 bases; and other: other G4 types that
1127 do not fall into the former categories (Fig. 1h, Extended Data Fig. 3a, Extended Data
1128 Fig. 10d). In this analysis random sequences with the same size distribution were used
1129 as control.
1130

1131 ***Computation of G4Hunter scores***

1132 G4Hunter scores were computed using a previously published principle⁵⁵ with specific
1133 functions (see provided R scripts). First, G4Hunter hits were extracted from the
1134 reference genomes (hg19, mm9, dm6 and sc3; with a window of 25 and a minimal
1135 score of 0.5). Each hit is characterized by its genomic coordinates and a "max_score"
1136 reflecting the score of the highest G4Hunter within this window. These scores reflect
1137 the relative G4 propensity of the peaks. Peaks with no overlapping G4Hunter hit have
1138 a score of 0. Random genomic regions of same size distribution were used as control.

1139 To assess GC richness effects in G4Hunter score distributions, all peaks were resized
1140 at 90bp around their summits. Shuffled DNA sequences using the same pool of
1141 nucleotides and the same size distribution were generated and compared to G4access
1142 or G4-ChIP datasets using G4Hunter analysis at 25 bp window settings and selecting
1143 the best scores in the fragments⁶ (Extended Data Fig. 2e).

1144 ***GC and CpG contents, G-tracks and number of G per track analyses***

1145 For number of G-tracks and number of G per G-track analyses, all peaks were resized
1146 at 90bp around their summits (Extended Data Fig. 3c-d). To assess GC and CpG
1147 contents, G or C and dinucleotide CG were determined in the specified windows. For
1148 promoters, windows from -200bp to gene TSSs were used (Extended Data Fig. 3e).
1149

1150 ***Gene expression and gene ontology analyses***

1151 Genes exhibiting G4access peaks in their promoters (within -200bp and their TSS)
1152 were determined using Bedtools⁵⁶. Fragments Per Kilobase per Million (FPKM) of all
1153 genes were determined using Cufflinks⁵⁷ (version 2.2.1) (Extended Data Fig. 3g).
1154 Charts were drawn from all expressed genes defined as genes with FPKM>0. (RNA-
1155 seq dataset used are indicated in Supplementary Table 4). Gene ontology was
1156 performed using DAVID webtool⁵⁸ (<https://david.ncifcrf.gov/>) on the top 3000
1157 promoters ranked by G4access signals (Extended Data Fig.3f).
1158

1159 ***Methylation analysis***

1160 WGBS datasets were analyzed using Bismarck (version 0.22.3)⁵⁹ for Fig. 6a-b (K562
1161 cells) and the computed beta files provided in ref⁶⁰ (GSE186458) for Fig. 6c-d (human
1162 granulocytes) . A Selection on G4H1.2 exhibiting at least 2CpG within the annotation
1163 and overlapping with G4access was performed (Fig. 6). A subset of 100000 randomly
1164 selected G4H1.2 which do not overlap with G4access peaks was used as control (Fig.
1165 6b).

1166 **Average binding profiles and heatmaps**

1167 To generate average binding or G4Hunter profiles (Fig. 1a, 4a, 4d, Extended Data Fig.
1168 1a, 7a-b, 8, 9c), R scripts were developed and used for retrieving bin scores in defined
1169 regions from 10 bp bin sized wiggle files⁴⁵. Heatmaps were generated, viewed and
1170 color-scaled according to sample read depth using Java TreeView⁶¹ (version 1.2.0-
1171 osx) (Fig. 4c, 6a,c). Regions were defined as centered on the G4access peak summits
1172 (from the narrow peak table of MACS2 at p -value $< 10^{-10}$). For Fig.4a and 4c, and
1173 because we applied a filter of G4access peaks with weak/moderate G4 predictions,
1174 G4access peaks (from the narrow peak table of MACS2, p -value $< 10^{-8}$) overlapping
1175 G4 predictions G4H1.2 or QP longloops were considered (n=10018). The longloop
1176 predictions were generated using the Quad-Parser consensus¹⁴ QP3-7, allowing one
1177 loop of a maximum length of 30 nt. To generate average binding profiles of Pol II (Fig.
1178 4d left), hg19 Ensembl gene annotations were used to extract values from wiggle files
1179 associated with the selected genes. Bin scores inside these annotations and in a
1180 region of 5kb before the TSSs and after 5kb of annotated termination sites were
1181 determined. Based on the gene list selections, bin scores from wiggle files were used
1182 to re-scale values between TSSs and transcription termination sites (gene body) of all
1183 genes using linear interpolation. In total, 1000 points were interpolated for the gene
1184 body of each selected gene in all average profiles presented.

1185
1186 Metaprofiles of G4access at gene bodies were performed as above at sites that do not
1187 overlap with H3K27ac ChIP-seq signals to avoid enhancers and unannotated
1188 promoters (3d right). Finally, Deeptools⁶² (version 3.3.0) was used to generate
1189 metaprofiles in Fig. 6.

1191 **Frequency analysis of predicted G4 in the human, mouse and yeast genomes**

1192 Predicted G4s were defined by G4Hunter using a window of 25 nt and thresholds
1193 ranging from G4H0.25 to G4H2.0^{6,55}. Number of hits per kb of the sequenced genome
1194 were then determined. For TSSs, predicted G4 densities found in the TSS area (-
1195 200bp to TSS) were calculated and expressed as predicted G4s/kb (Extended Data
1196 Fig. 10a).

1198 **Analysis of genomic locations of predicted G4s or G4access peaks**

1199 Predicted G4s were defined by G4Hunter using a window of 25 nt at thresholds of 1.2,
1200 1.5 and 2.0. Genomic locations of predictions and G4access peaks detected by
1201 MACS2 (see above) were defined as follows (Extended Data Fig. 3b): Promoter (-
1202 200bp to TSS), 5' Gene Body (TSS to +400bp), Gene Body (from +400 bp after the
1203 TSS to -200bp upstream off the Transcription End Site (TES), TES (-200 bp to +200
1204 bp of TES), all other locations were defined as intergenic (using Ensembl annotations).

1206 **Analysis of signal variation between replicates**

1207 All peaks found in the replicates of the same experiments were merged (using
1208 bedtools) and signal from individual samples were extracted for each sample. Results
1209 were then plotted using R as scatterplots (x-axis: signal from individual replicates; y-
1210 axis: merged G4access signals; Extended Data Fig. 1e)

1212 **Statistics and reproducibility**

1213 *Sample size*

1214 No statistical method was used to predetermine sample size. In the case of Fig. 2
1215 analysis, 596 G4access peaks were randomly selected within the 4743 G4access
1216 peaks that were found in common in the 3 cell lines (Hela, Raji, K562). Margin of error
1217 of this selection is 3.68% for a Confidence Level of 95% following the central limit
1218 theorem.

1219
1220 *Data exclusion*

1221 For Fig. 4a and 4c, and because we applied a filter of G4access peaks with
1222 weak/moderate G4 prediction scores, G4access peaks (from the narrow peak table of
1223 MACS2, p-value < 10⁻⁸) overlapping G4 predictions G4H1.2 and/or QP longloops were
1224 considered. The longloop predictions were generated using the Quad-Parser
1225 consensus.

1226 In the gene expression analysis presented in Extended Data Fig. 3g, genes that are
1227 not expressed (RPKM=0) were not included.

1228 In MNase-seq heatmaps, saturated or absence of signals expanding all along the
1229 displayed genomic areas were considered as artifacts or outliers and removed.

1230
1231 *Replication*

1232 Experiments were repeated in replicates as indicated in the manuscript, (Between 2
1233 and 4 times). All replicates were successful.

1234 Please note that Extended Data Fig. 1b has been performed as technical replicates
1235 only.

1236
1237 *Randomization*

1238 Randomization of genome sequences were repeated 10 times for enrichment
1239 analyses. In Fig. 1g, 6a, Extended Data Fig. 2d-e,6c-d,7c and 10a-b genomic
1240 sequences were randomly selected and compared to experimental datasets of same
1241 sizes.

1242
1243 *Blinding*

1244 596 G4access peaks were blindly tested using in vitro assays (Fig. 2) using coded
1245 oligonucleotides.

1246

1247 **Data availability**

1248 The genomic data sets published in this study are available under GSE187007. All data
1249 used from previously published study are referenced in Supplementary Table 4.

1250 **Code availability**

1251 Code generated and used for this study has been deposited in the linked Zenodo
1252 repository: <https://zenodo.org/record/7912528> (ref. ⁶³). The G4Hunter algorithm
1253 version code and functions used in this article are included as supplementary files
1254 (G4Hunter.r and G4HunterAccess_function.r).

1255
1256 **Methods-only references**

1257
1258

1259 43 Esnault, C., Magat, T., García-Oliver, E. & Andrau, J. C. Analyses of Promoter ,
1260 Enhancer, and Nucleosome Organization in Mammalian Cells by MNase-Seq.
1261 *Methods Mol Biol* **2351**, 93-104, doi:10.1007/978-1-0716-1597-3_5 (2021).

1262 44 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat*
1263 *Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

1264 45 Fenouil, R. *et al.* Pasha: a versatile R package for piling chromatin HTS data.
1265 *Bioinformatics* **32**, 2528-2530, doi:10.1093/bioinformatics/btw206 (2016).

1266 46 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137,
1267 doi:10.1186/gb-2008-9-9-r137 (2008).

1268 47 Orlando, D. A. *et al.* Quantitative ChIP-Seq normalization reveals global
1269 modulation of the epigenome. *Cell Rep* **9**, 1163-1170,
1270 doi:10.1016/j.celrep.2014.10.018 (2014).

1271 48 Anders, S. & Huber, W. Differential expression analysis for sequence count data.
1272 *Genome Biol* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).

1273 49 Descostes, N. *et al.* Tyrosine phosphorylation of RNA polymerase II CTD is
1274 associated with antisense promoter transcription and active enhancers in
1275 mammalian cells. *Elife* **3**, e02105, doi:10.7554/eLife.02105 (2014).

1276 50 Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic*
1277 *acids research* **37**, W202-208, doi:10.1093/nar/gkp335 (2009).

1278 51 Anvar, Z. *et al.* ZFP57 recognizes multiple and closely spaced sequence motif
1279 variants to maintain repressive epigenetic marks in mouse embryonic stem cells.
1280 *Nucleic acids research* **44**, 1118-1132, doi:10.1093/nar/gkv1059 (2016).

1281 52 Morison, I. M., Ramsay, J. P. & Spencer, H. G. A census of mammalian imprinting.
1282 *Trends Genet* **21**, 457-465, doi:10.1016/j.tig.2005.06.008 (2005).

1283 53 Schulz, R. *et al.* WAMIDEX: a web atlas of murine genomic imprinting and
1284 differential expression. *Epigenetics* **3**, 89-96, doi:10.4161/epi.3.2.5900 (2008).

1285 54 Xie, W. *et al.* Base-resolution analyses of sequence and parent-of-origin
1286 dependent DNA methylation in the mouse genome. *Cell* **148**, 816-831,
1287 doi:10.1016/j.cell.2011.12.035 (2012).

1288 55 Lacroix, L. G4HunterApps. *Bioinformatics* **35**, 2311-2312,
1289 doi:10.1093/bioinformatics/bty951 (2019).

1290 56 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
1291 genomic features. *Bioinformatics* **26**, 841-842,
1292 doi:10.1093/bioinformatics/btq033 (2010).

1293 57 Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals
1294 unannotated transcripts and isoform switching during cell differentiation. *Nature*
1295 *biotechnology* **28**, 511-515, doi:10.1038/nbt.1621 (2010).

1296 58 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis
1297 of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57,
1298 doi:10.1038/nprot.2008.211 (2009).

1299 59 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for
1300 Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572,
1301 doi:10.1093/bioinformatics/btr167 (2011).

1302 60 Loyfer, N. *et al.* A human DNA methylation atlas reveals principles of cell type-
1303 specific methylation and identifies thousands of cell type-specific regulatory
1304 elements. *bioRxiv*, 2022.2001.2024.477547, doi:10.1101/2022.01.24.477547
1305 (2022).

1306 61 Saldanha, A. J. Java Treeview--extensible visualization of microarray data.
1307 *Bioinformatics* **20**, 3246-3248, doi:10.1093/bioinformatics/bth349 (2004).

1308 62 Humayun, M. S., Rady, A. M. & Soliman, G. M. Obstructive jaundice secondary to
1309 intra-biliary rupture of hepatic hydatid cyst. *Int Surg* **74**, 4-6 (1989).
1310 63 Makrini, A., Esnault, C., Andrau, J.C., Magat, T. Scripts and codes for G4access
1311 analysis. <https://zenodo.org/record/7912528> (2023).
1312

Supplementary information

Extended data Figure legends

ED Figure 1: G4access signal optimization and characterization.

a- G4FS at all promoters are associated to open regions upstream of TSSs. The graph shows nucleosome and G4H2.0 densities in Raji cells, as well as the location of upstream and downstream peak's locations of G4 and nucleosome deep of all CGI containing promoters. **b-** qPCR quantification of a model G4 (human MAZ locus) in G4access preparations, normalized to a KRAS region negative control (KRAS_neg). This G4 containing fragment is enriched in the 3 cell lines at various digestion levels of MNase as indicated. MNase activity was controlled by measuring the level mononucleosome fractions (see Fig. 1b). **c-** FRET melting profiles for comparison of physiological (red) and MNase (black) digestion conditions. Fluorescence level reflects denaturation of the G4 structure. **d-** Table of test sequences and G4Hunter scores. T_m and ΔT_m are indicated for all sequences except Myc, because of complex melting and very high stability. Note that all G4s are highly stable in the MNase buffer at room temperature or 37°C (blue bar). **e-** Correlation plots of G4access merged signals (of 2 biological replicates) compared to individual biological replicates.

ED Figure 2: G4access comparison to G4-ChIP and G4Hunter

a- Comparison of G4access signal and G4-ChIP at a selected area of the genome (KRAS locus, (chr12: 25.330.000-25.560.000)) **b-** Venn diagram of overlapping G4access peaks in the 3 model cell lines. (Fisher exact tests of 3 the overlaps $<1E-4$) **c-** Venn diagram of overlapping G4access and G4-ChIP peaks in the HaCaT and K562 cell lines. (Fisher exact tests of all overlaps $<1E-4$). **d-** G4Hunter prediction scores in G4access performed in 3 human cell lines and comparison to published G4 ChIP-seq in 2 of these cell lines. For the sake of comparison, all fragments were resized at 90 bp in G4-ChIP, G4access peaks and genomic DNA (40.000 annotations; see methods). All distributions are highly significant compared to random selections (not shown) using a two-sided Wilcoxon test (p -value $<2.E-16$). **e-** G4Hunter prediction scores compared to shuffled sequences of same sizes and same nucleotide compositions and to random sequences (see methods; all differences in the distributions of G4access associated scores are highly significant compared to random and shuffled selections using a two-sided Wilcoxon test, p -value $<2E-16$).

ED Figure 3: G4access genomic localization, sequence characterization and association to gene expression programs.

a- G4 subtypes identified in the 3 cell lines (see methods). The various categories are 'Loop size' 1–3, 4–5 and 6–7: sequences with at least one loop of the respective length; simple bulge: sequences with a G4 with a bulge of 1–7 bases in one G-run or multiple 1-base bulges; 2-tetrads/Complex bulge: sequences with a G4s with two G-bases per G-run or several bulges of 1–5 bases; and other: other G4 types that do not fall into the former categories (see methods). **b-** Compared partition of G4access and G4 ChIP regions in the human genome. The control bars represent the genomic distribution of G4FS at various stringencies (G4Hunter scores of 1.2, 1.5 and 2.0). TES represent transcription end sites at gene units. **c-** Analyses of number GG or GGG tracks found in G4-ChIP or G4access peak datasets ($n=11563, 44412, 12216, 13320$ and 9031). **d-** Number of Gs found in the G-tracks of the predicted G4s in the G4-ChIP or G4access

datasets, with at least 2 G per track. **e-** GC and CpG contents distributions at promoters associated to G4access peaks (K562 n=8343, HaCaT n=4090, Raji n=4465, all genes n=20314). **f-** Gene ontology analyses using DAVID database of the genes associated to promoter with G4access peaks in K562, Raji and HaCaT cells (DAVID, $-\log_{10}$ of modified Fisher Exact p-value.). **g-** Gene expression level analysis expressed as Fragment per kb per million (FPKM) in chromatin RNA-seq datasets in K562 and Raji cells (n=4660, 8569, 32355, 31779, 4659, 8601, 32753 and 31434). Box plots represent minimal and maximal values, first and third quartiles and the median value.

ED Figure 4: Motifs associated to G4access and G4 ChIP peaks in the 3 model cell lines (2 in the case of ChIP) at TSS and all sites as indicated.

The sequence logos and statistics associated to this analysis were generated using the MEME algorithm. Presented motifs are ranked by occurrence (top 3). MEME-ChIP e-value are displayed.

ED Figure 5: *In vitro* G4 characterization and validation (relates to Fig. 2).

a- Principle of the ThT and NMM G4 determination. **b-** Cumulative percentage of validated regions in FRET-MC above a given threshold of G4Hunter of G4access selection sequences. **c-** Experimental fluorescence for NMM experiments. G4 threshold is indicated at 125 (a. u). **d-** Experimental fluorescence for ThT experiments. G4 threshold is indicated at 200 (a. u).

ED Figure 6: G4access measures G-quadruplex dynamics in response to cell treatments with a G4 ligand.

a- Genome browser view illustrating Pyridostatin (PDS, 10 μ M for 30 min) effect on G4access peaks dynamics in Raji cells (Chr1: 203.500.000-205.500.000). In the zoom area is shown the promoter ATP2B4, in which the main G4access signal redistributes from strong to weak G4FS. **b-** DESeq analysis of G4access signal following 30 min of treatment by PDS. The promoter-proximal (TSS) and non-promoter G4s are indicated in red and blue respectively (DESeq, p-value <0.05). **c-** G4access score density is shifted toward weaker G4s following PDS treatment. **d-** G4seq score density is shifted toward weaker G4s following PDS treatment, although to a lesser extent than for G4access.

ED Figure 7: Nucleosome and Pol II features at G4access peaks, with or without strong G4 predictions.

a- average profiles of G4access regions depending of their nucleosome depletion level (relates to Fig. 4). Metaprofiles of MNase-seq (Nucleosome midpoints), G4access and Pol II ChIP-seq centered on G4 access summits in the 4 groups defined in Fig. 4c. The corresponding signals for the H3K4me3 and H3K4me1 ChIP-seq in Raji cells are also shown (right panels), for which the relative high amount of H3K4me3/me1 is indicative of a promoter feature, as seen for group 1 and, to a lesser extent, group 2. **b-** Features of signals below G4 formation threshold in G4access signal. G4access signals were selected above (G4Hunter >1.2; n=9047 regions) or below (<1.0; n=3492 regions) threshold for G4 formation in all genomic locations and analyzed for nucleosome positioning/density, G4access signals and Pol II loading. G4-forming sequences are strongly associated with nucleosome depletion and positioning **c-** G4Hunter prediction

scores in nucleosome depleted regions (NDRs, see methods) associated or not to G4access peaks. A random selection of genomic area of same size is indicated in light grey. While distributions of scores at G4access associated NDRs are highly significant compared to random selections (using a two-sided Wilcoxon test, p -value $<2.2E-16$), distributions of G4Hunter scores at other NDRs are not significantly different to random selections.

ED Figure 8: Nucleosome depletion at experimentally defined G-quadruplexes.

a- G4Hunter (G4H1.2) and chromatin landscape (ATAC-seq and MNase-seq density or positioning) profiling in K562 cells at sites with common or specific G4access and G4-ChIP peaks as indicated. **b-** G4Hunter (G4H1.2) and chromatin landscape (ATAC-seq) profiling in HaCaT cells. Groups were defined as in Extended Data Fig. 2c and genomic datasets used are listed in the Supplementary Table 1.

ED Figure 9: G4access dynamics in response to nucleosome perturbation by the HDAC inhibitor TSA.

a- TSA treatment for 24 hours leads to H3K9acetylation increase. Western-blot of VCP and total H3 (loading controls) and of H3K9ac in 3 independent replicates are shown. **b-** Representative examples of G4access decrease associated to NDR closure at the MFSD2A promoter (chr1: 40.418.000- 40.424.000) and the chr16: 19.503.827-19.506.304 genomic region. **c-** TSA treatment for 24 hours leads to a global decrease of chromatin accessibility at NDRs associated to G4access decrease signal. Metaprofiles of G4access (left) and MNase-seq density and positioning (right) are shown at all TSSs (up) and non TSS (bottom) sites.

ED Figure 10: Application of the G4access procedure in organisms with less genomic G4 densities.

a- Comparison of G4Hunter prediction frequencies per kb (higher table) and densities (lower table and graph in the right panel) in 3 distinct organisms (Human, *D. melanogaster* and *S. cerevisiae*). **b-** G4 prediction scores in G4access and equivalent selection of random DNA fragments in the 3 organisms. **c-** Motif search (MEME) at promoter and non-promoter sites, ranked by occurrence in flies and yeast. MEME-ChIP e-value are displayed. **d-** Repartition of the G4 subtypes in G4access peaks in flies and yeast as for Fig. 1h. In yeast, the majority of G4access peaks are non-forming G4 sequences. **e-** Examples of G4access, ATAC-seq and Pol II ChIP-seq signals in *Drosophila* (chr3L: 18.755.000-18.772.500) and Yeast (chrIV: 766.800-771.500). The isolated peaks for G4access and ATAC, and the G4H1.2 annotations are indicated below the signal tracks.

Tables

Supplementary Table 1 includes 3 spreadsheets describing DMR summary, randomly selected G4access peaks used for biophysical assays and a description of all genomic data sets used in this study.

Supplementary Table 2 describes oligonucleotides used for G4 structure determination of human genomic fragments isolated in G4access experiments.

Figure 1

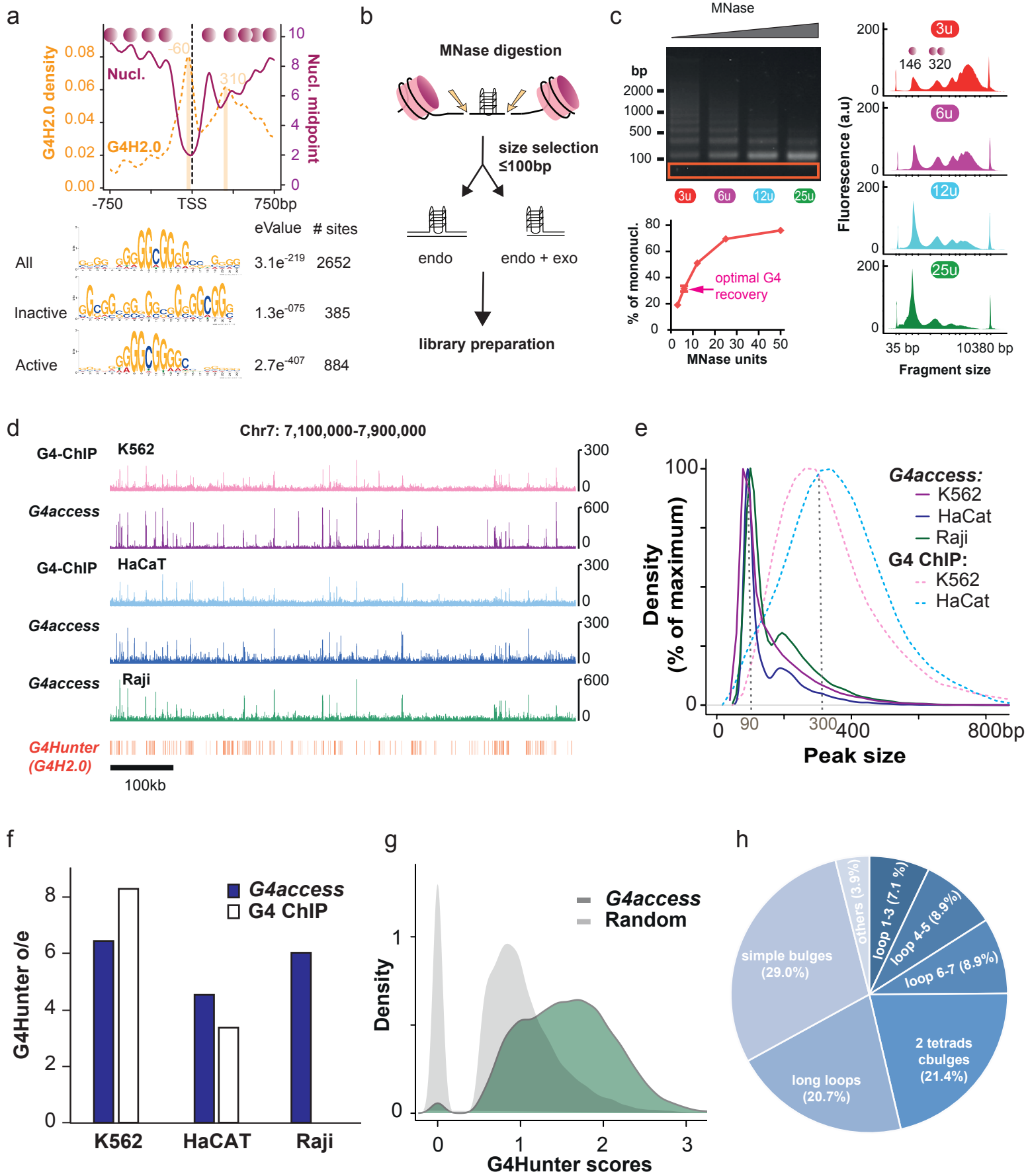


Figure 2

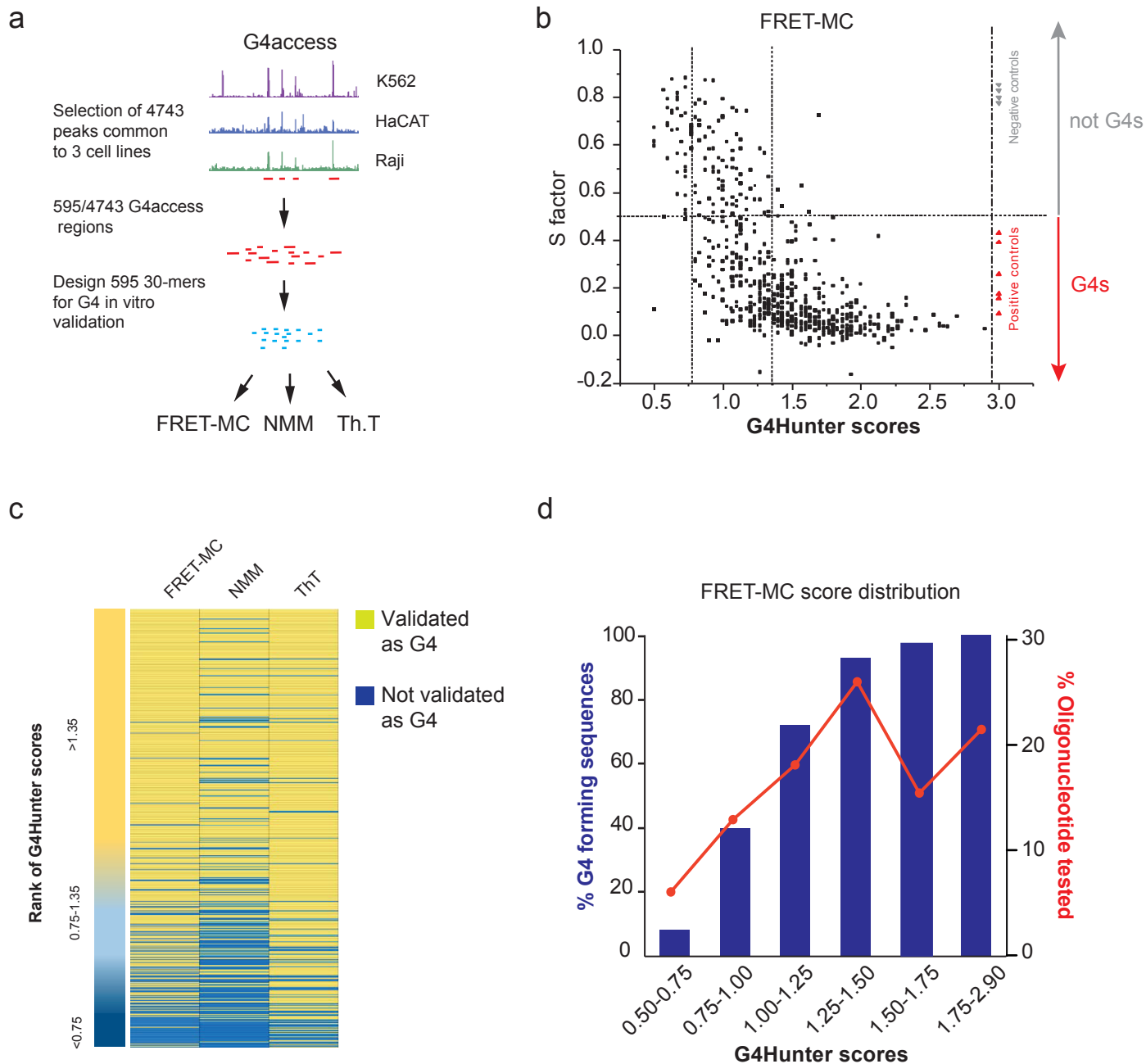


Figure 3

siDHX36 vs siNT

siWRN vs siNT

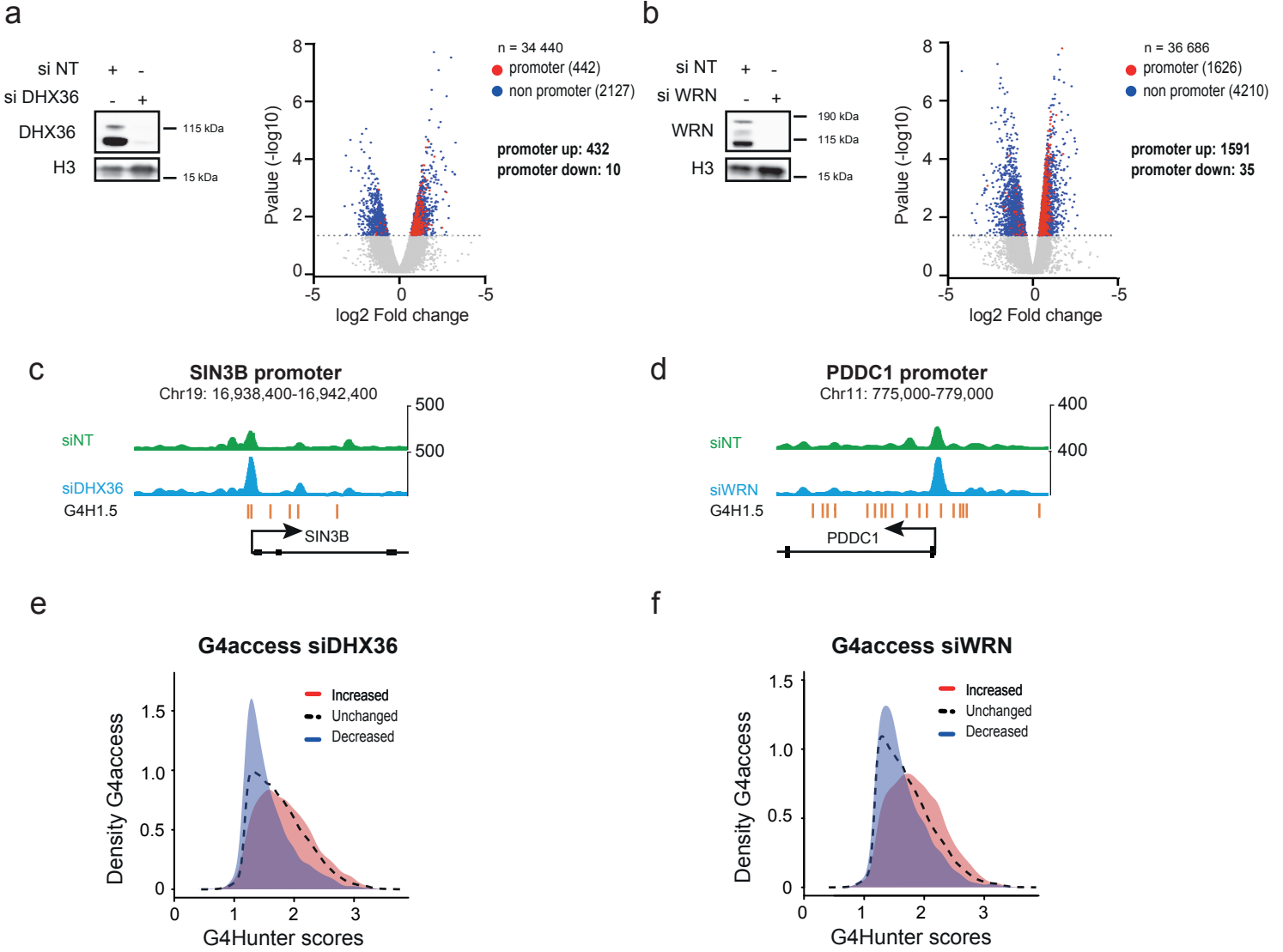


Figure 4

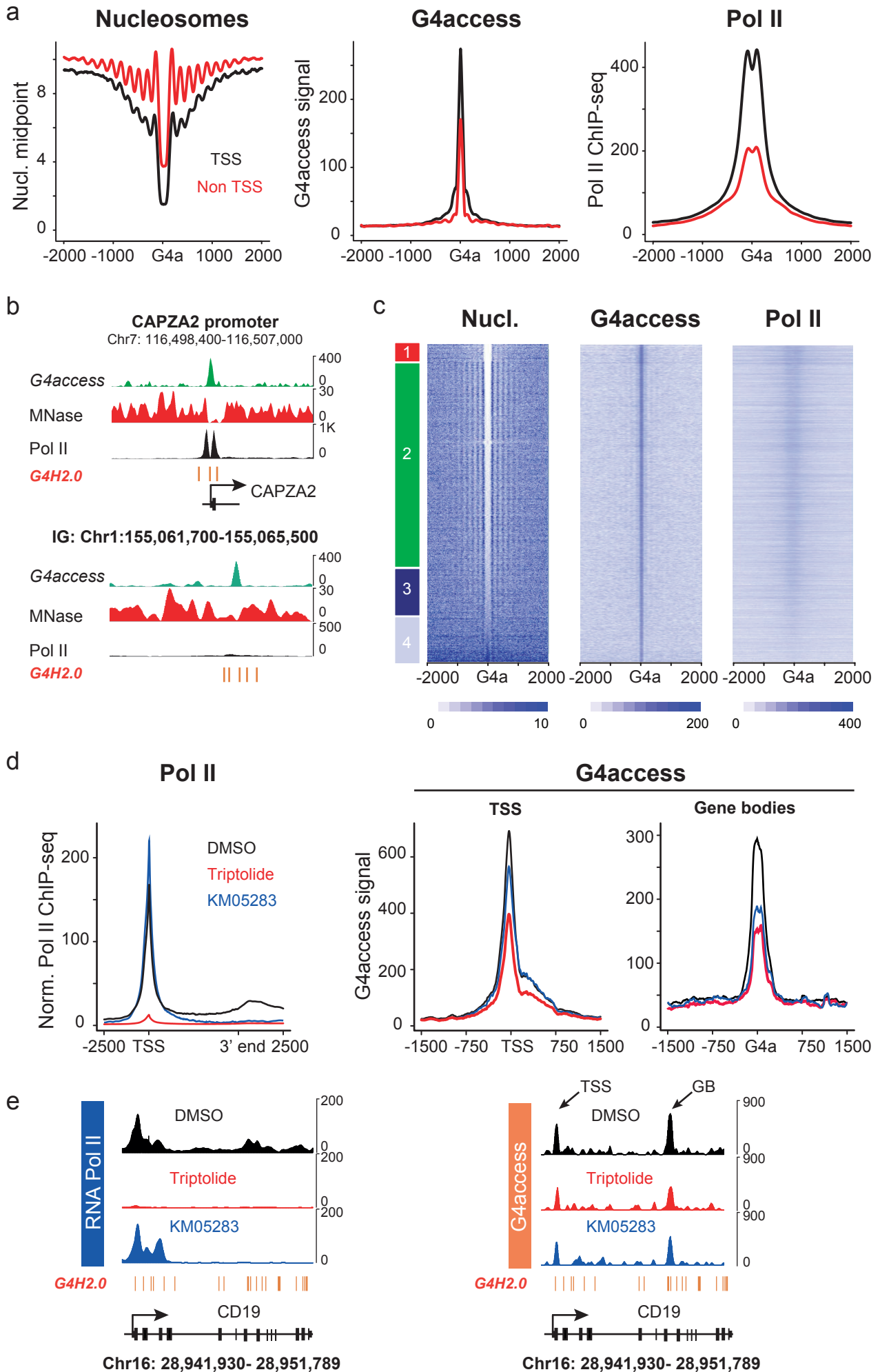


Figure 5

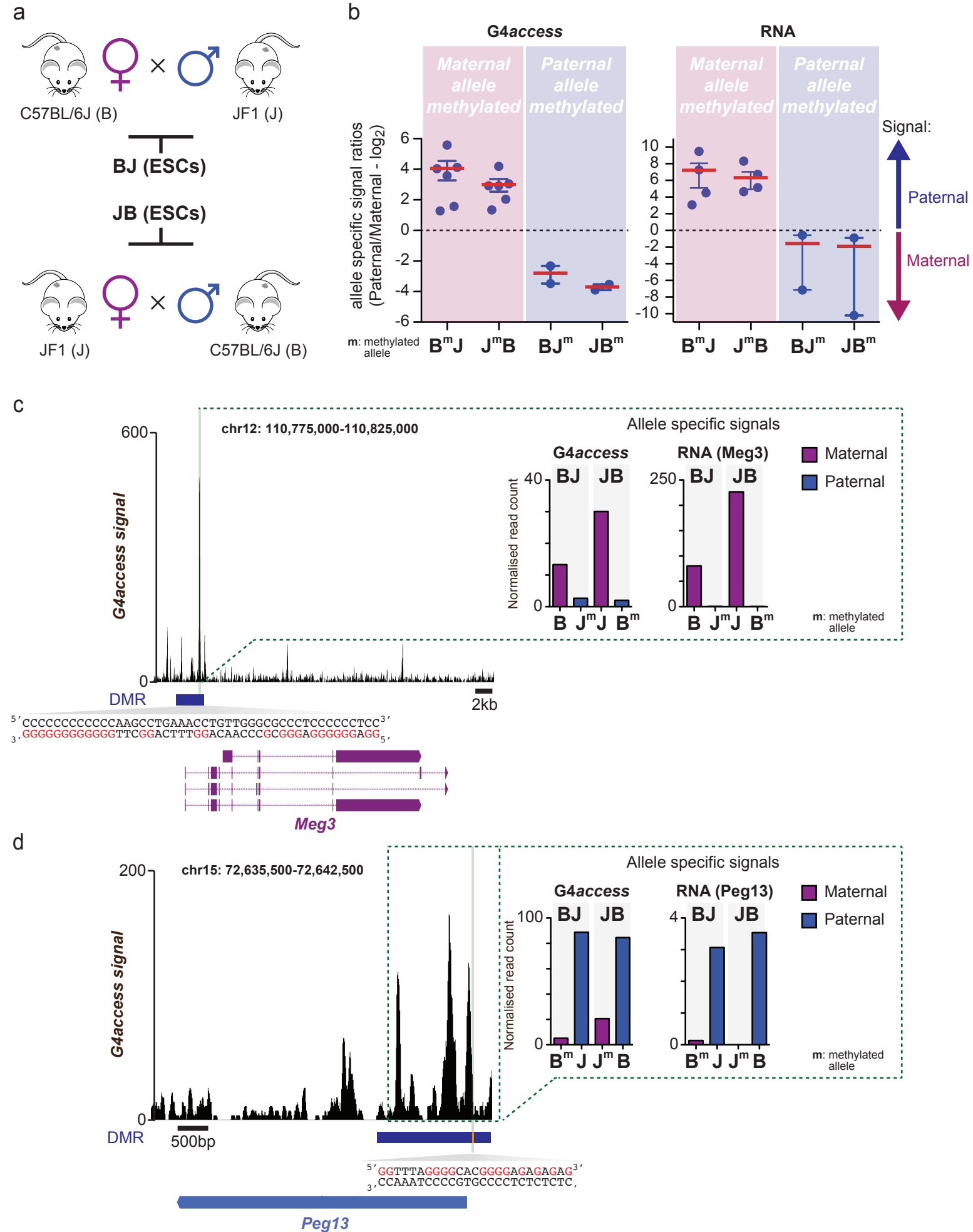
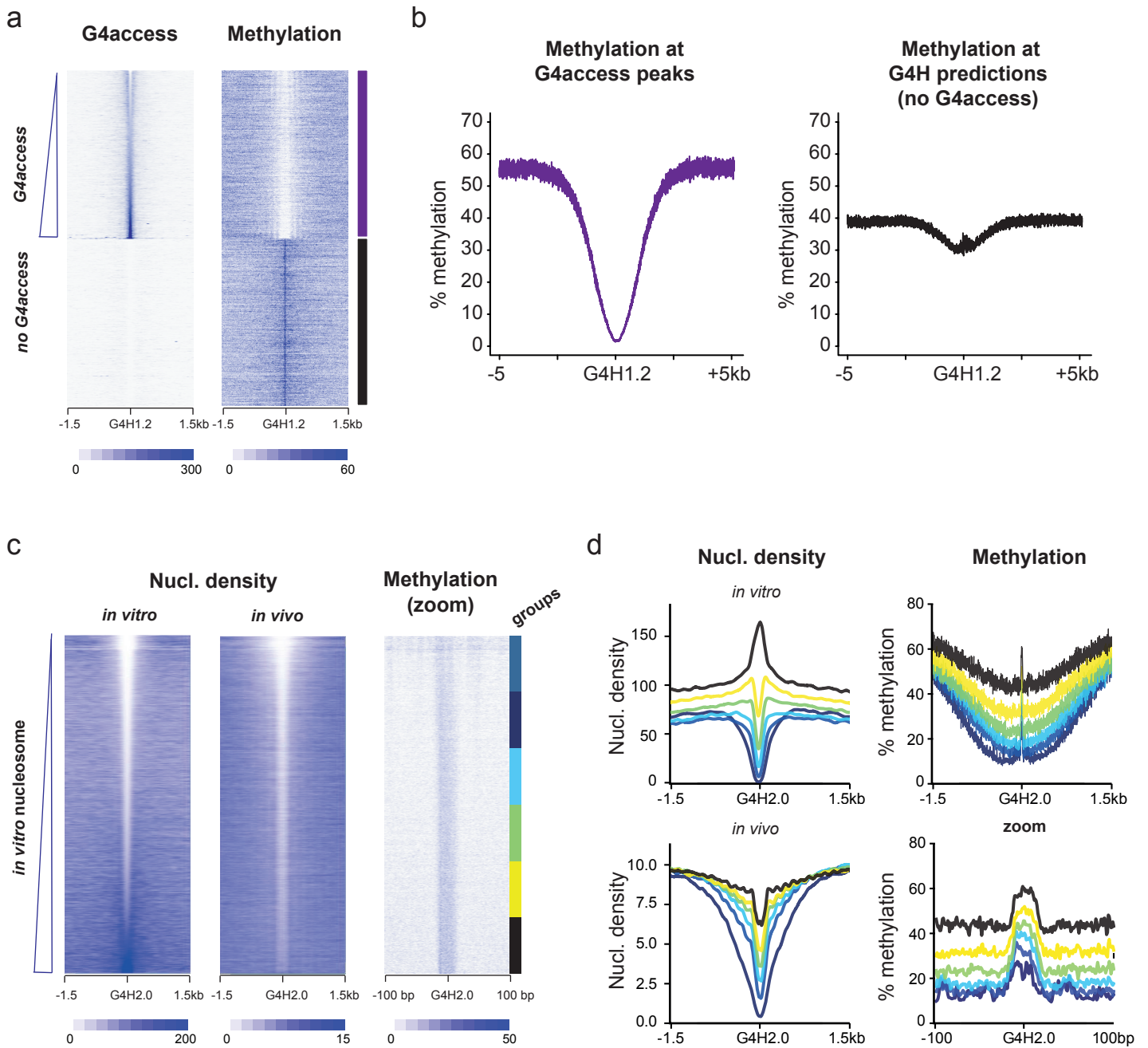
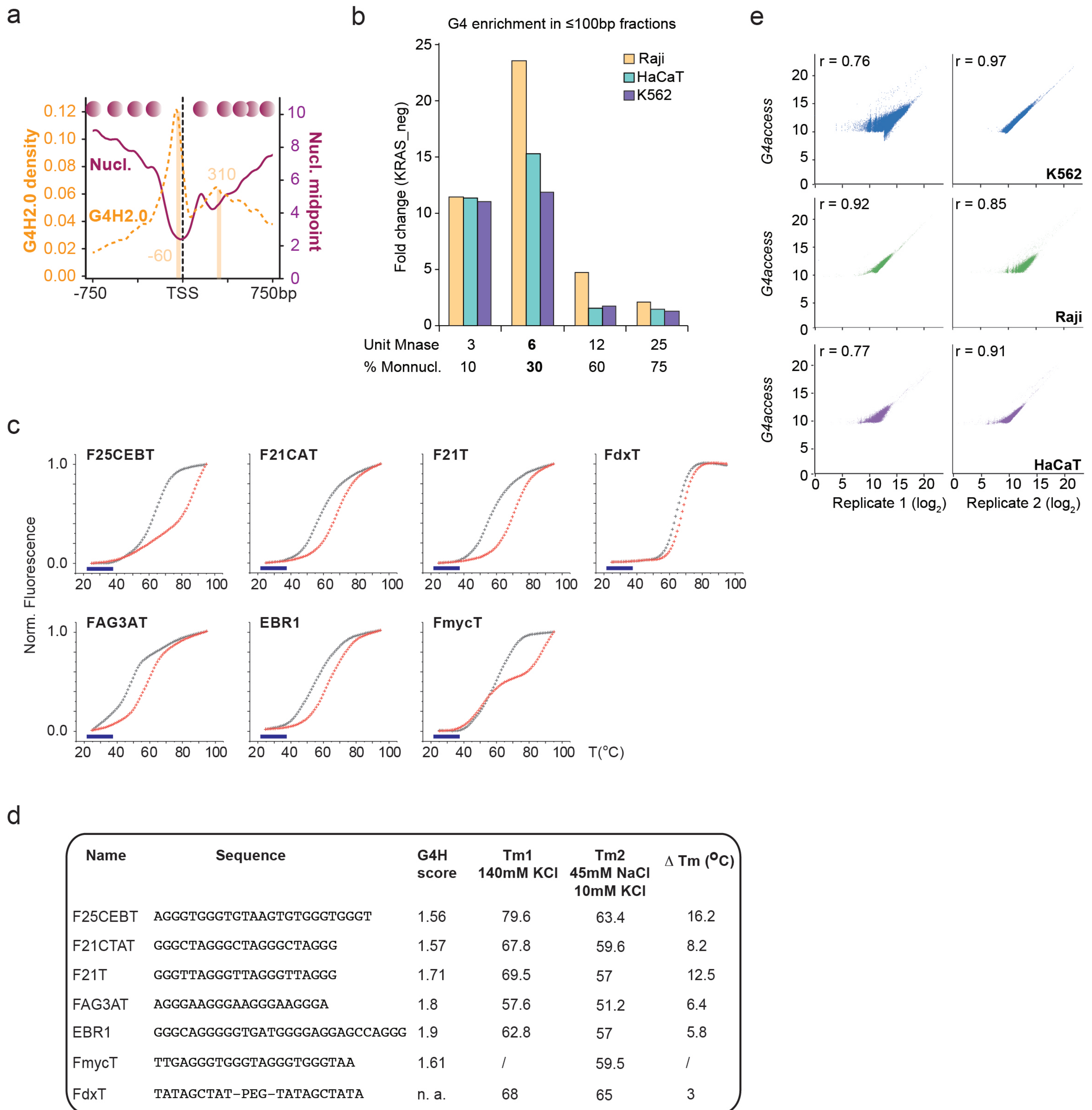


Figure 6

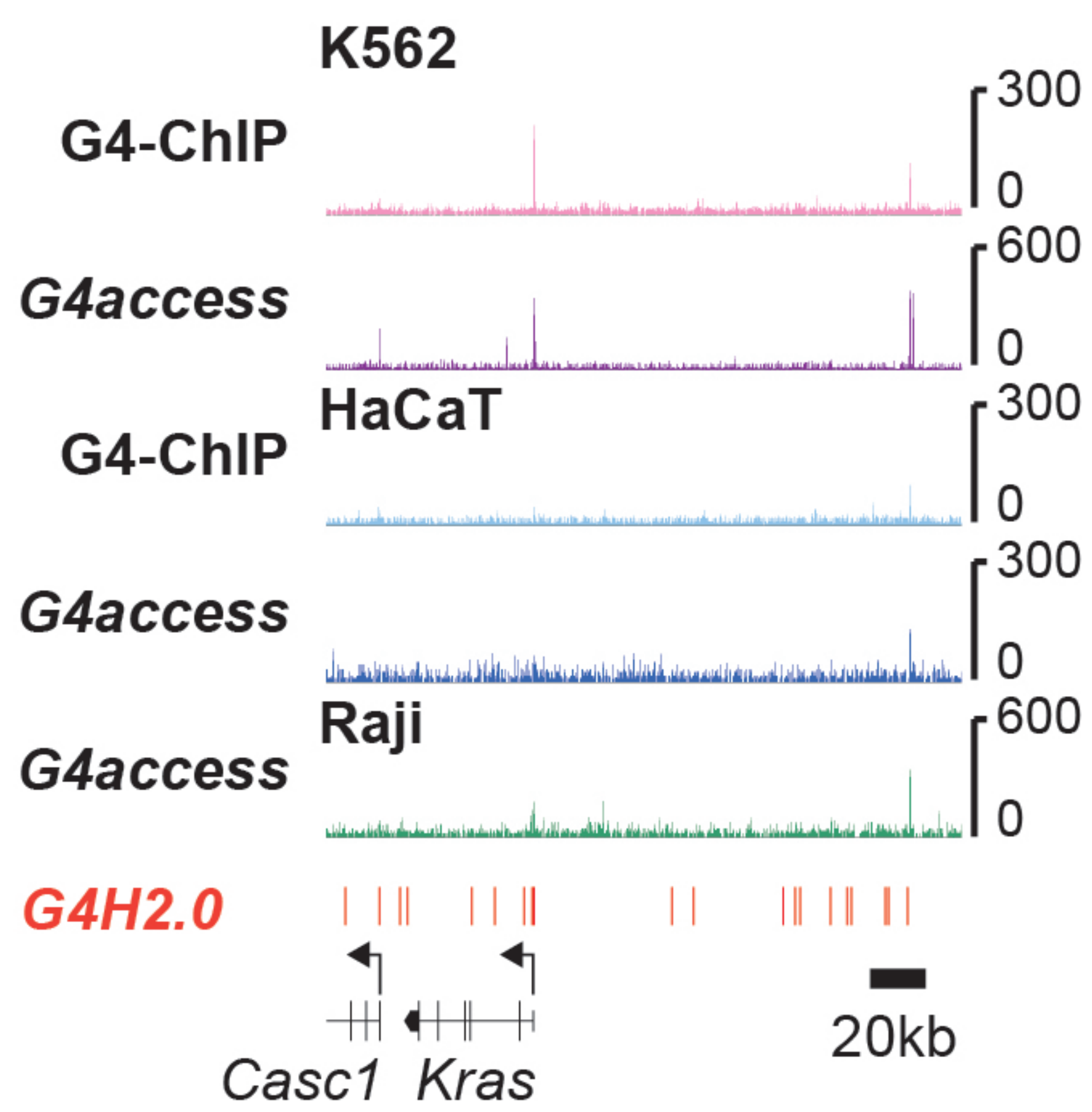


ED Figure 1

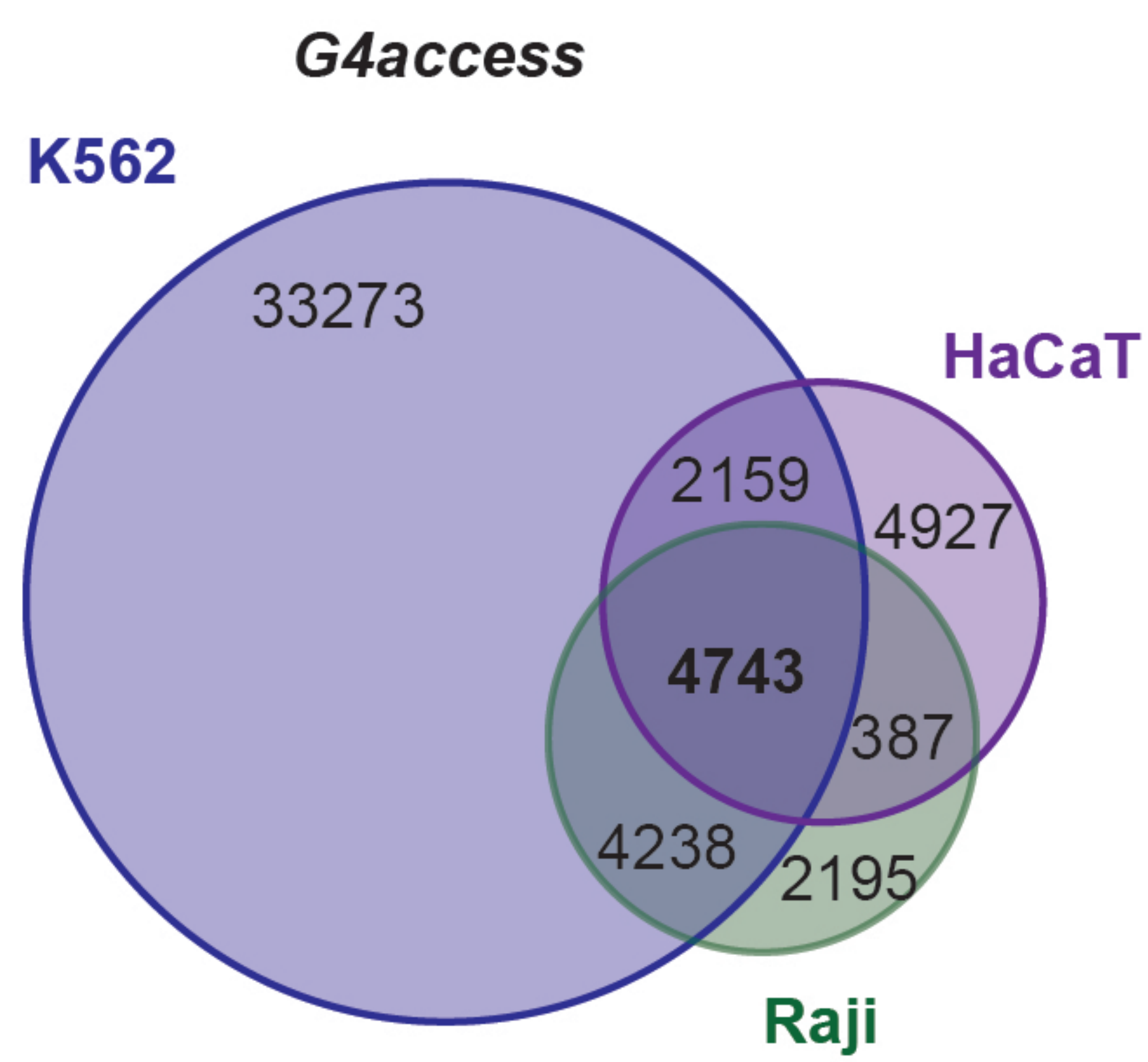


ED Figure 2

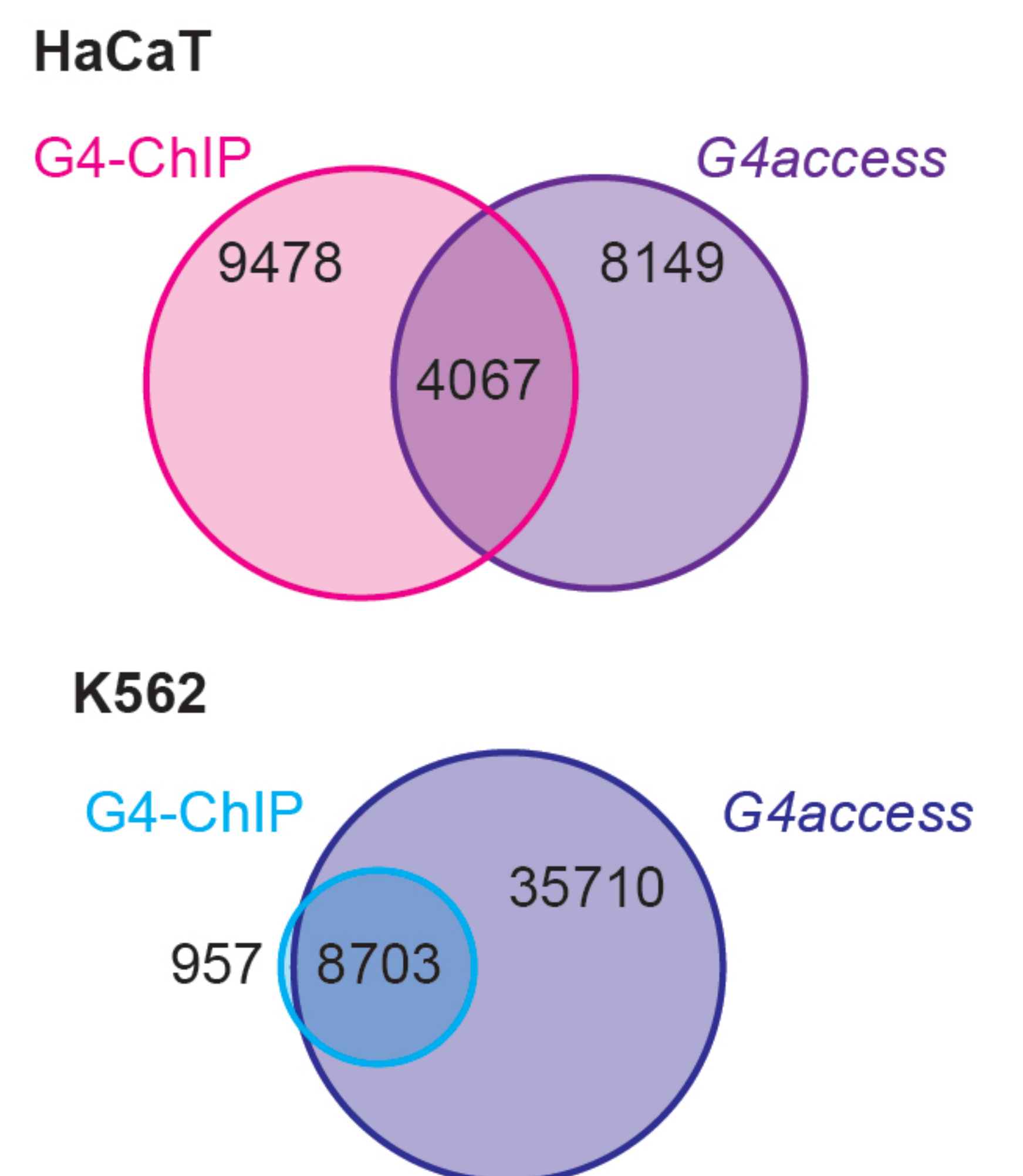
a chr12: 25,330,000-25,560,000



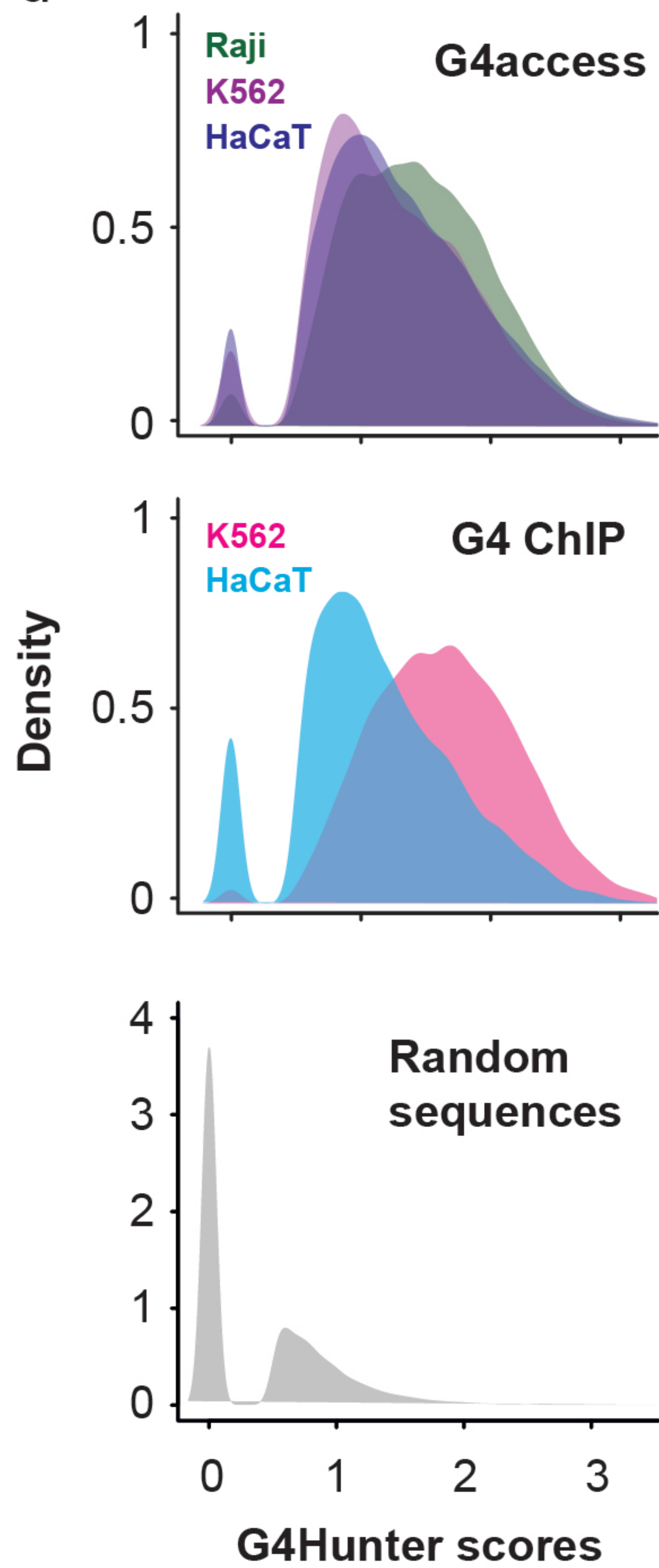
b



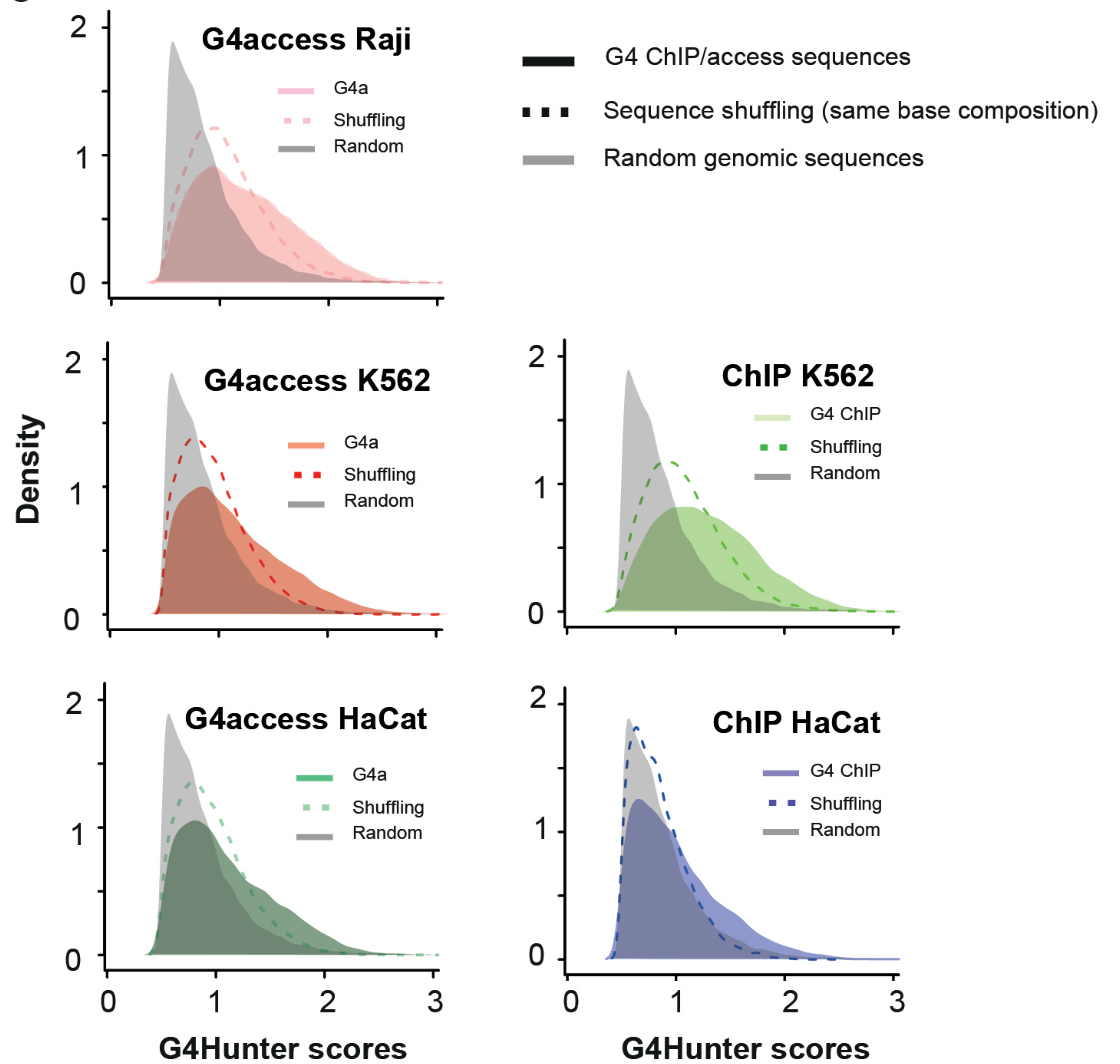
c



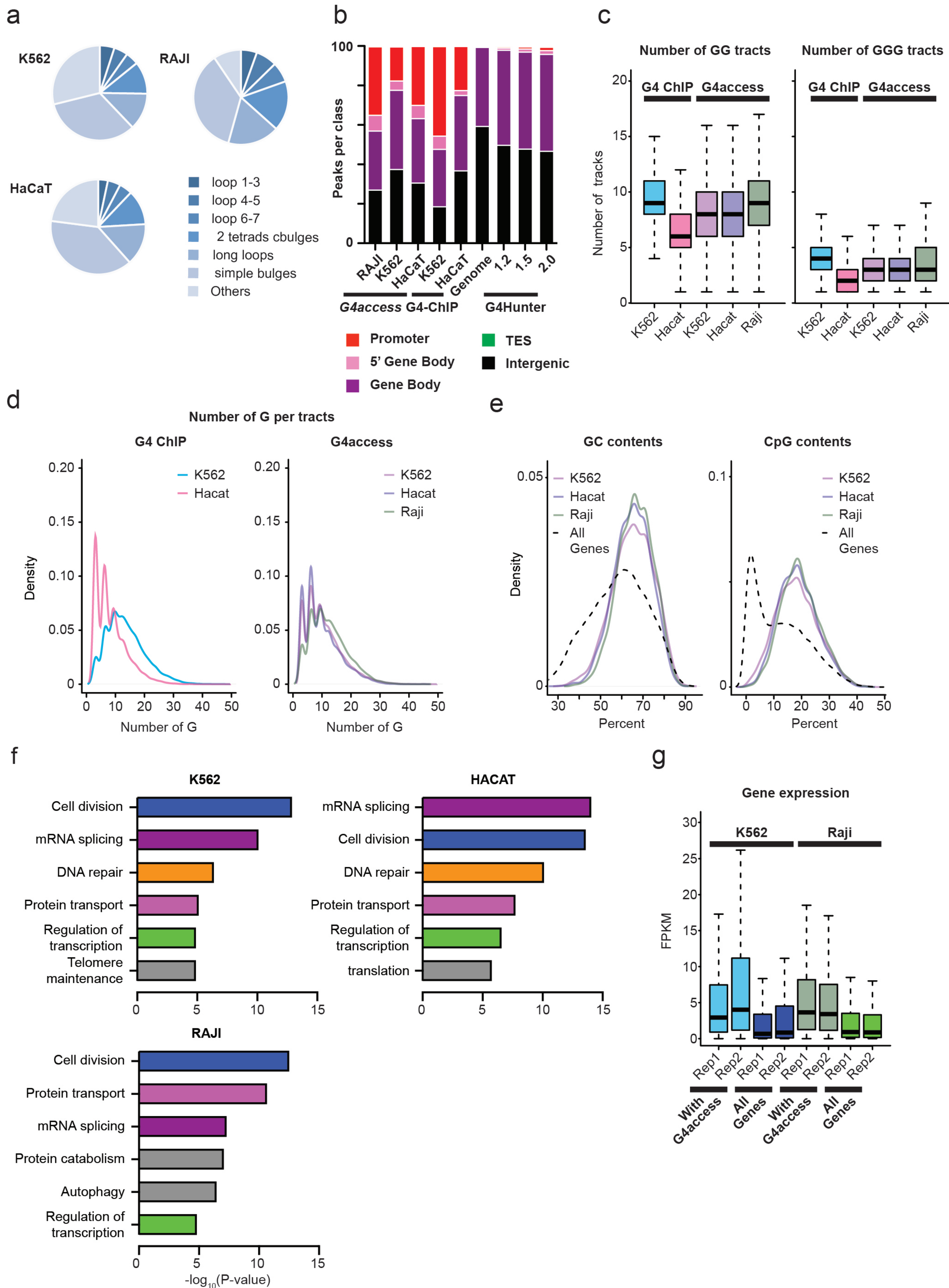
d



e



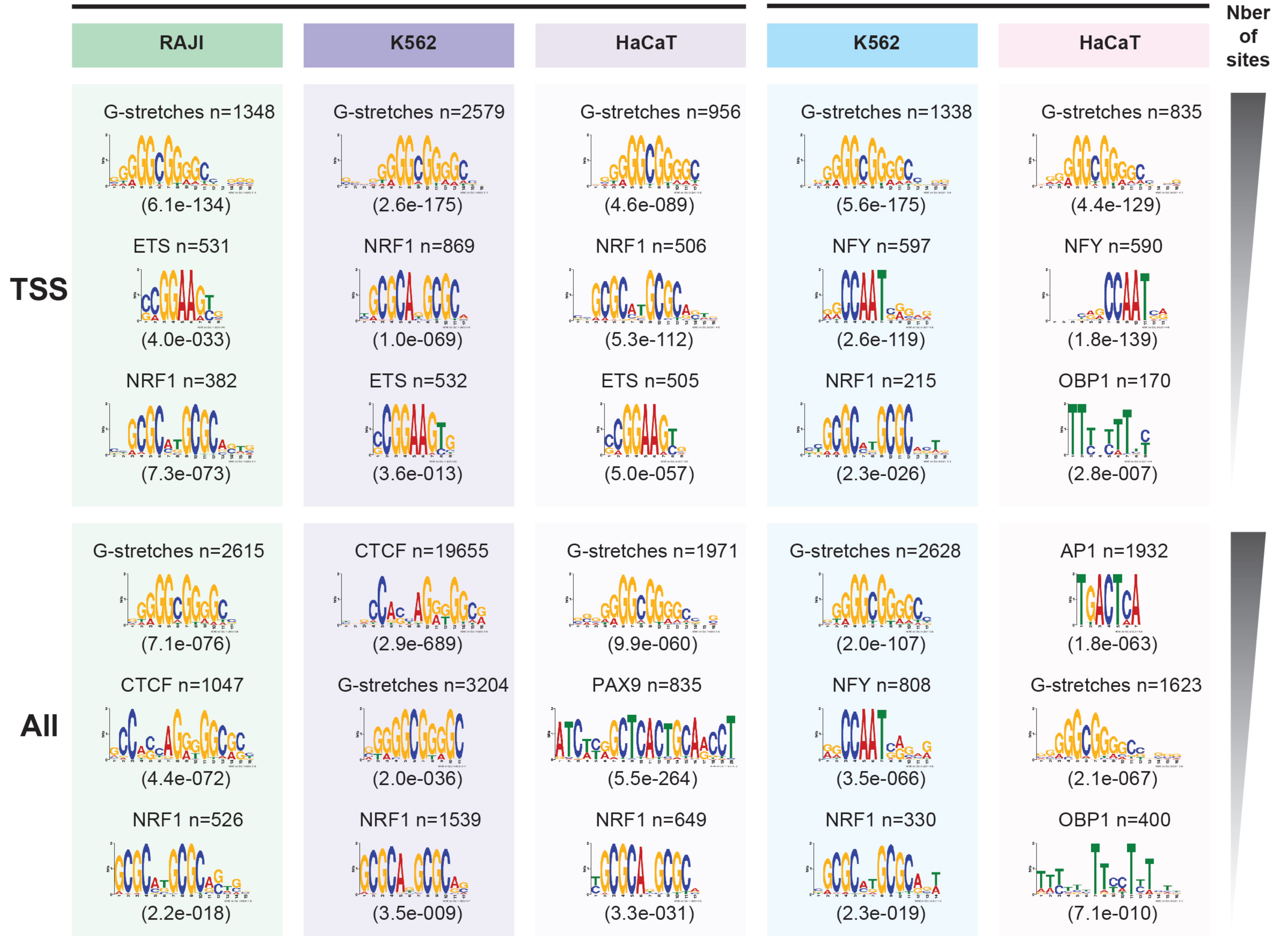
ED Figure 3



ED Figure 4

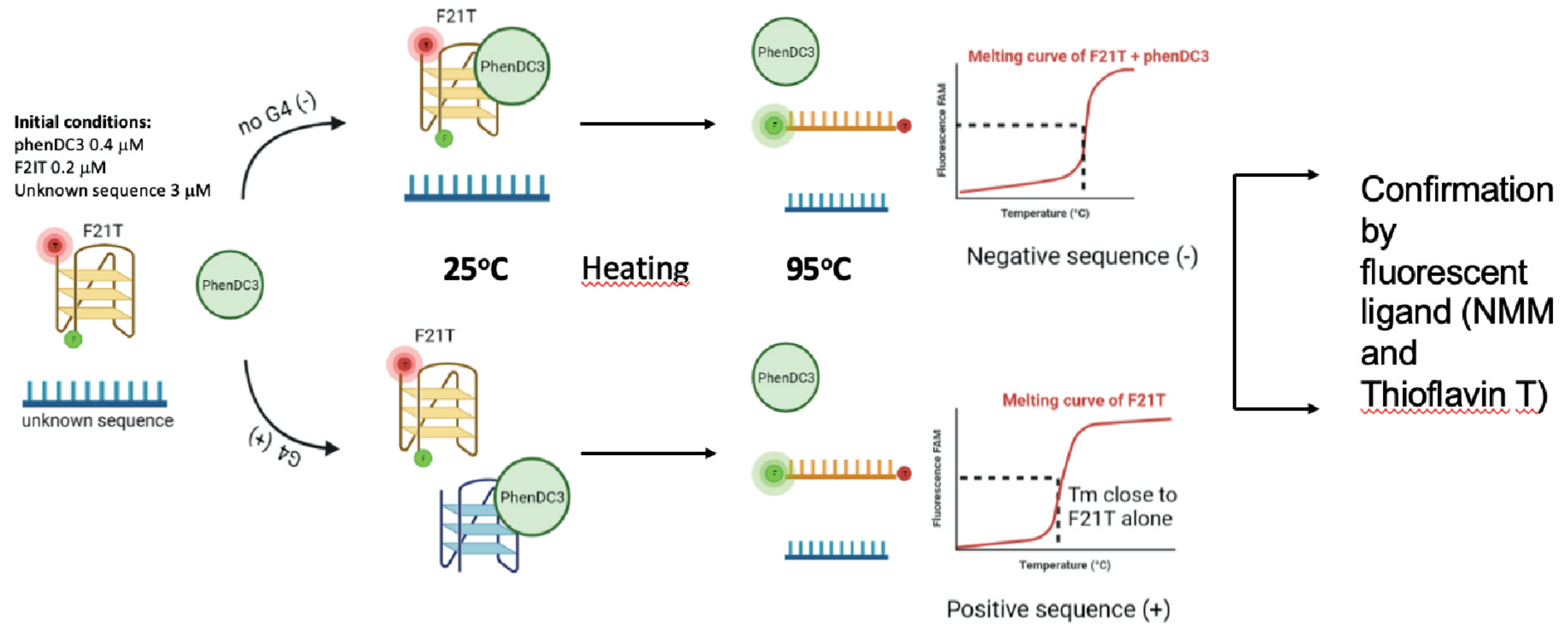
G4access

G4 CHIP

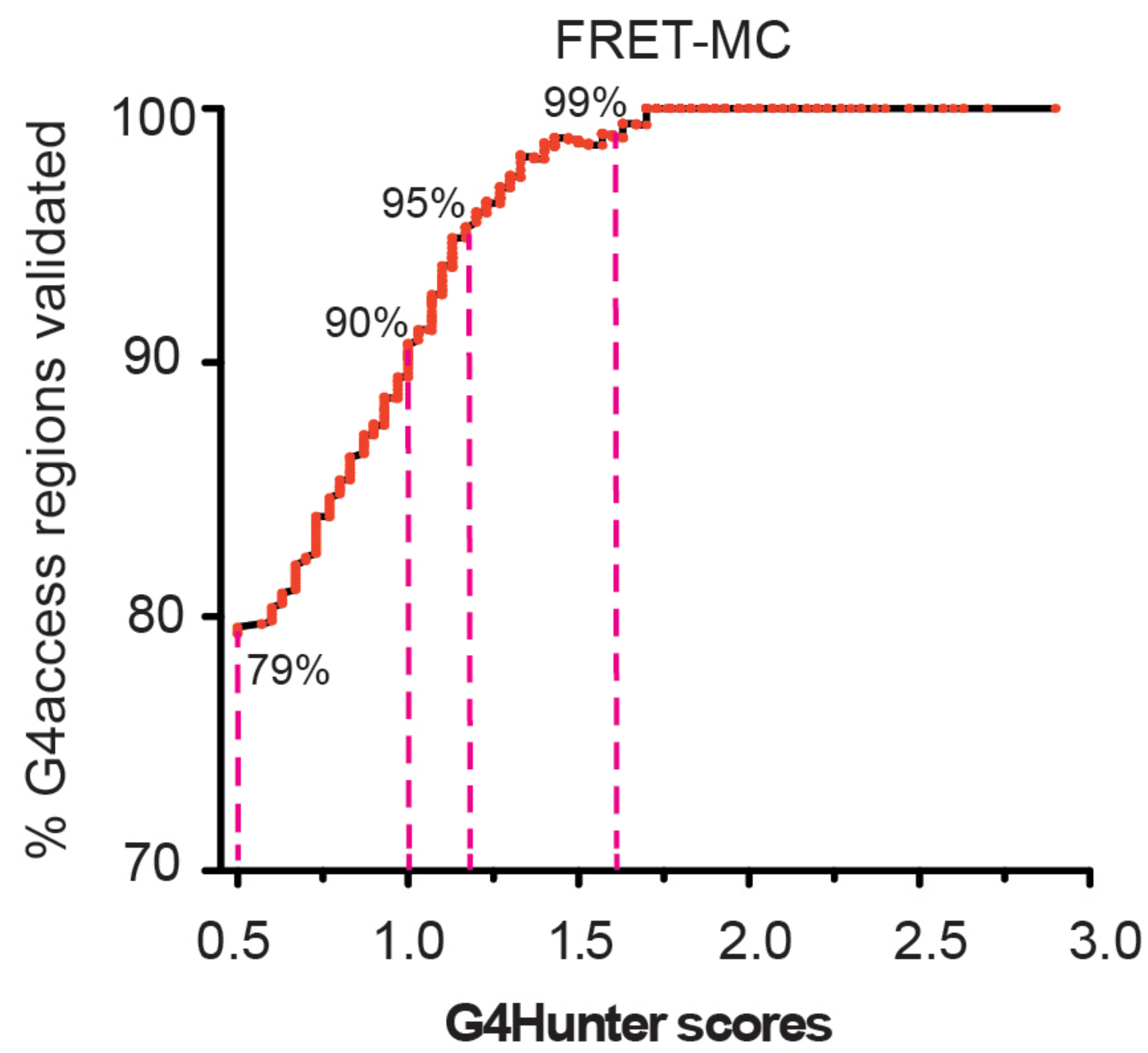


ED Figure 5

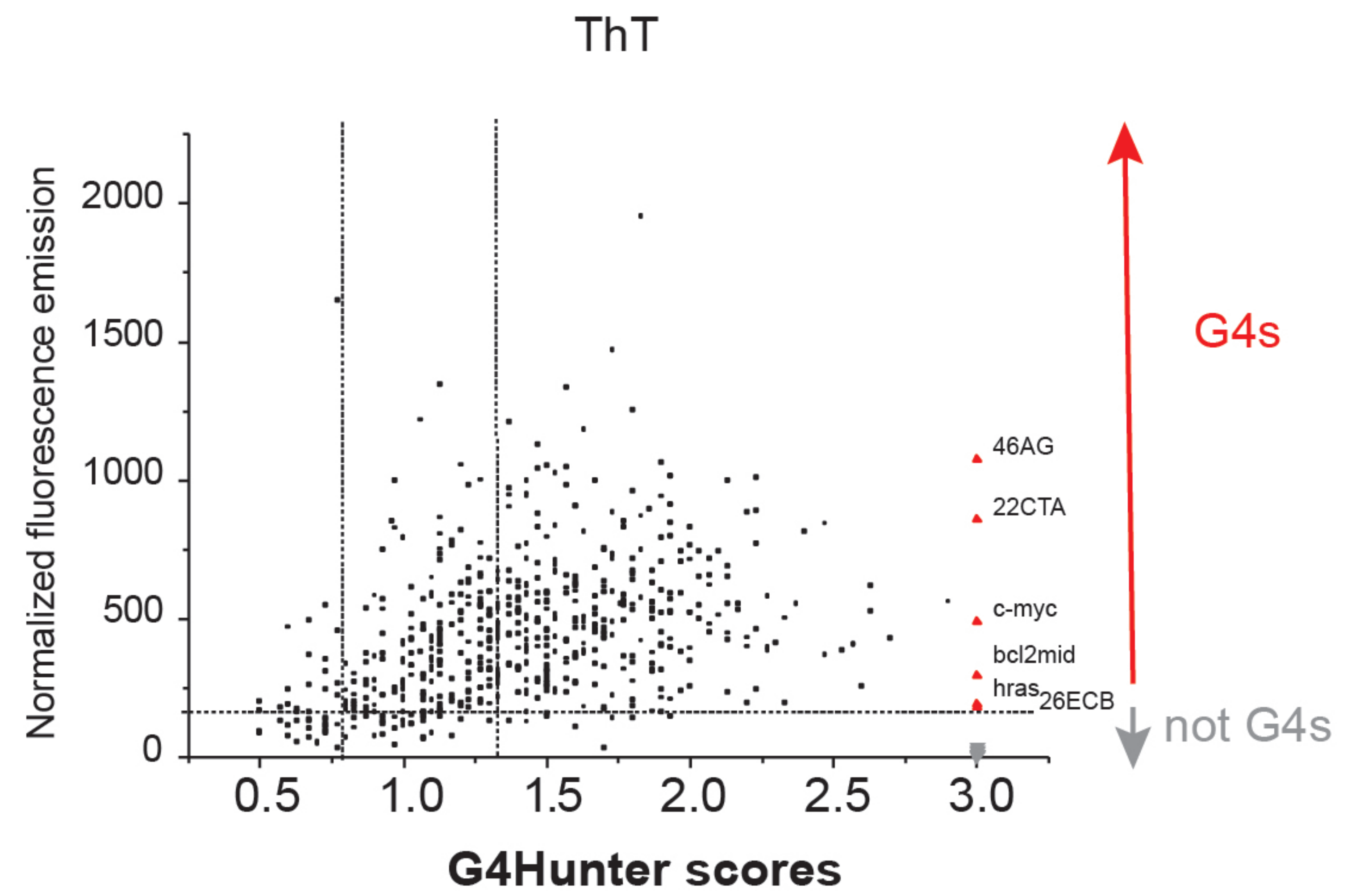
a



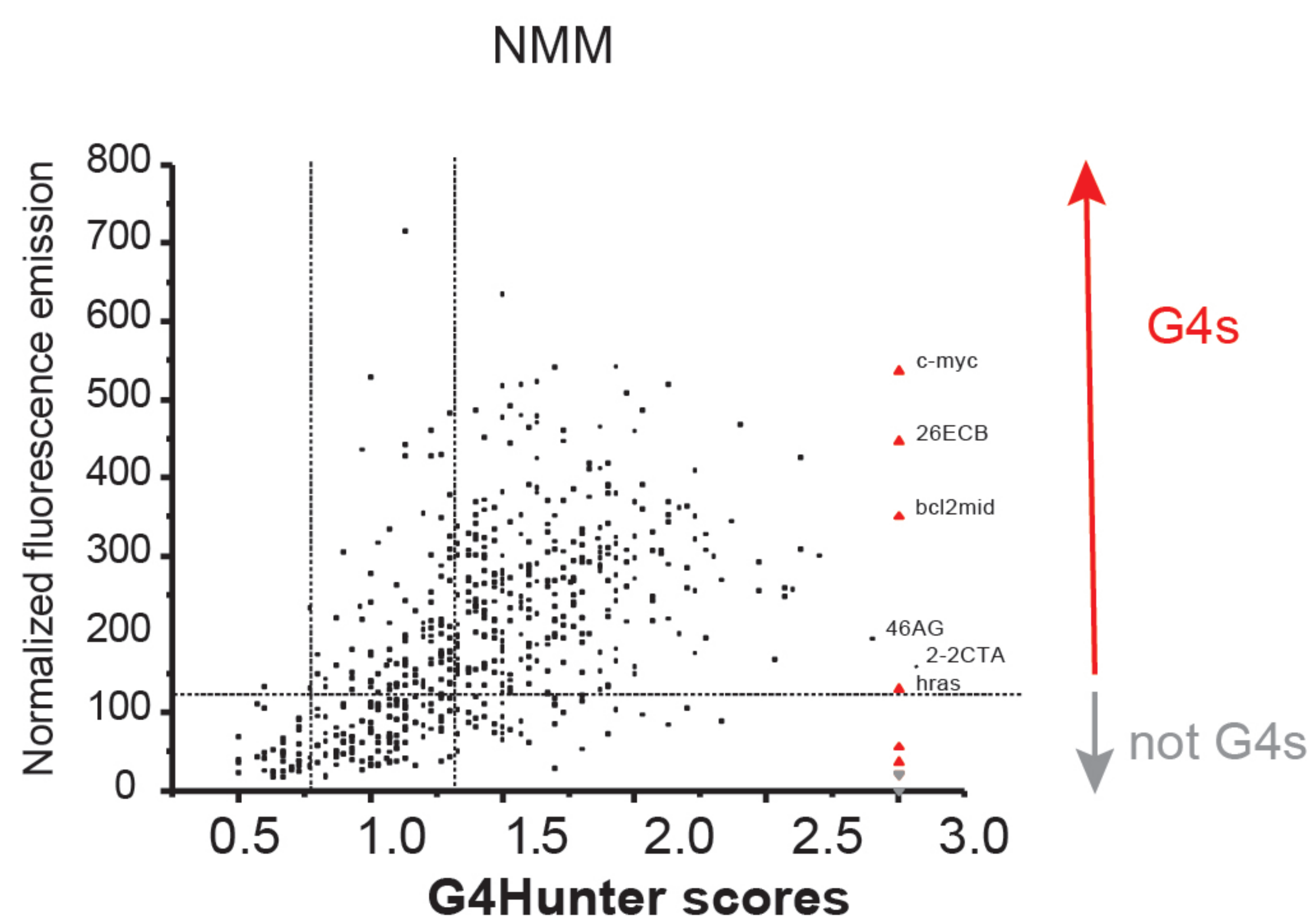
b



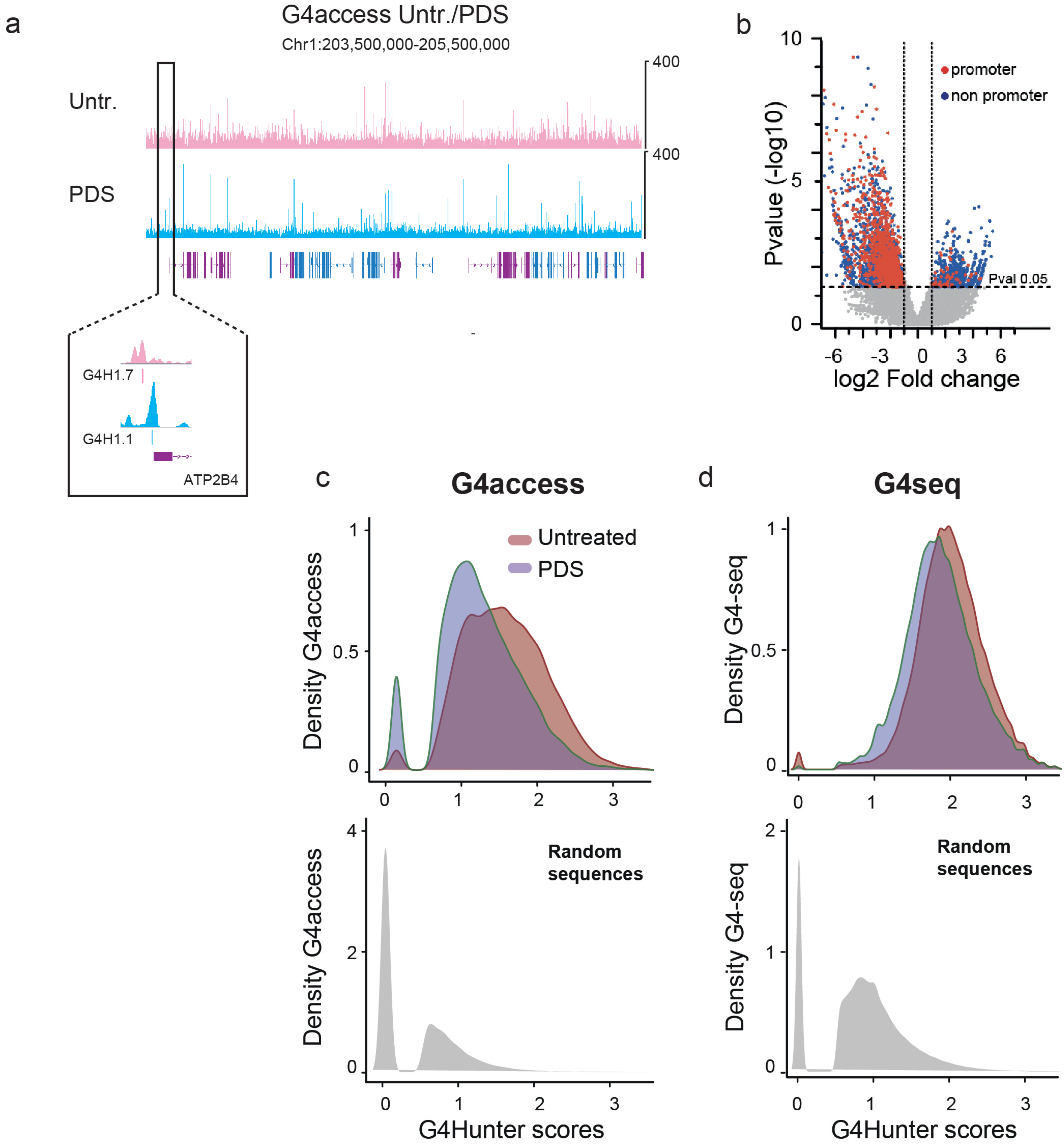
c



d



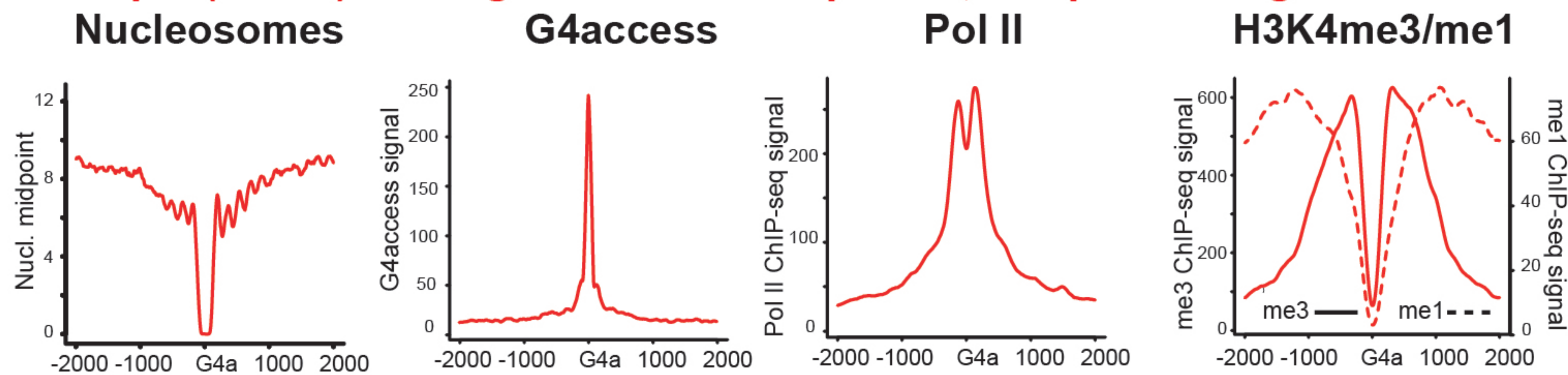
ED Figure 6



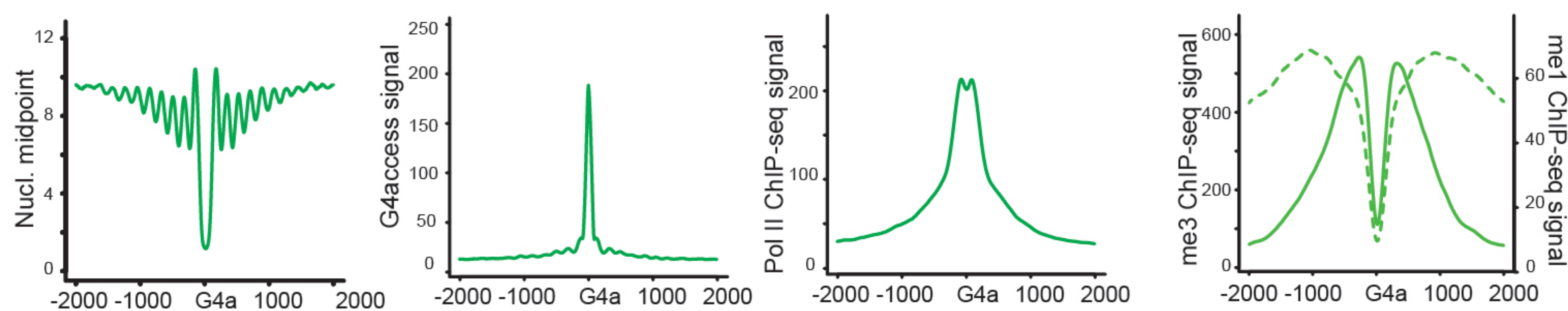
ED Figure 7

a

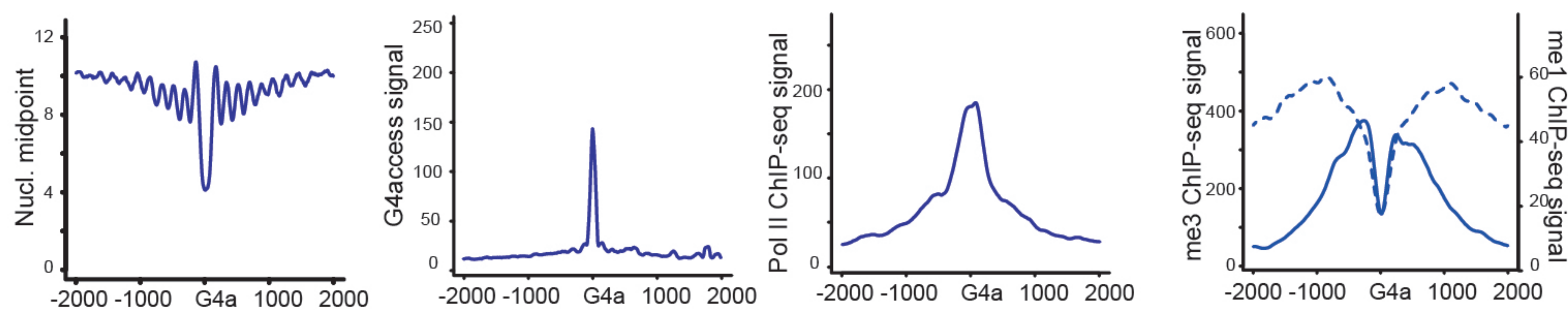
Group 1 (n=844): Strong nucleosome depletion, low positioning



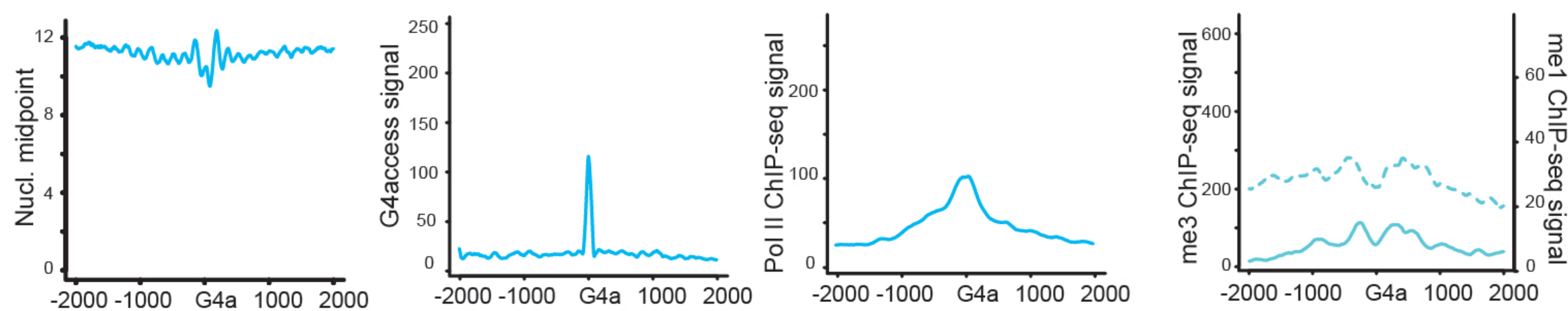
Group 2 (n=6215): Strong nucleosome depletion, strong positioning



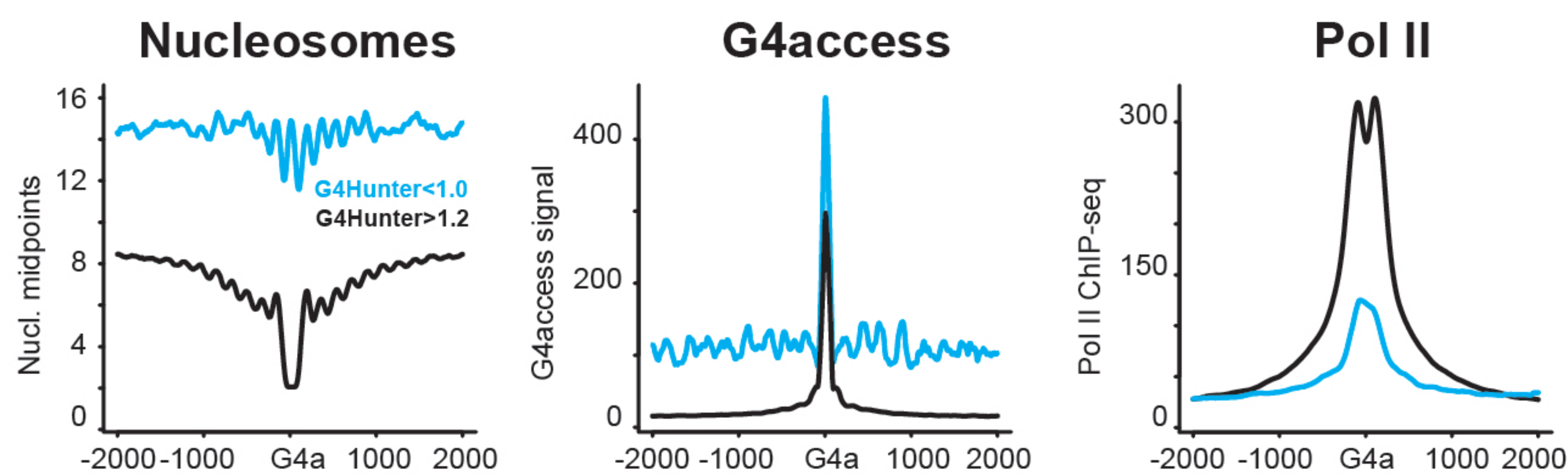
Group 3 (n=2073): Moderate nucleosome depletion, strong positioning



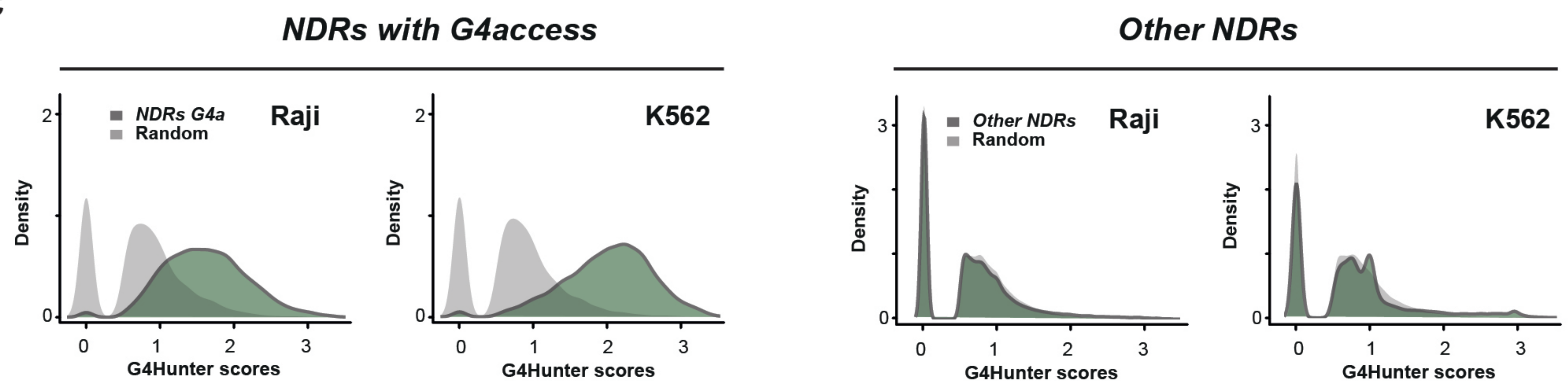
Group 4 (n=2675): No nucleosome depletion, weaker positioning



b



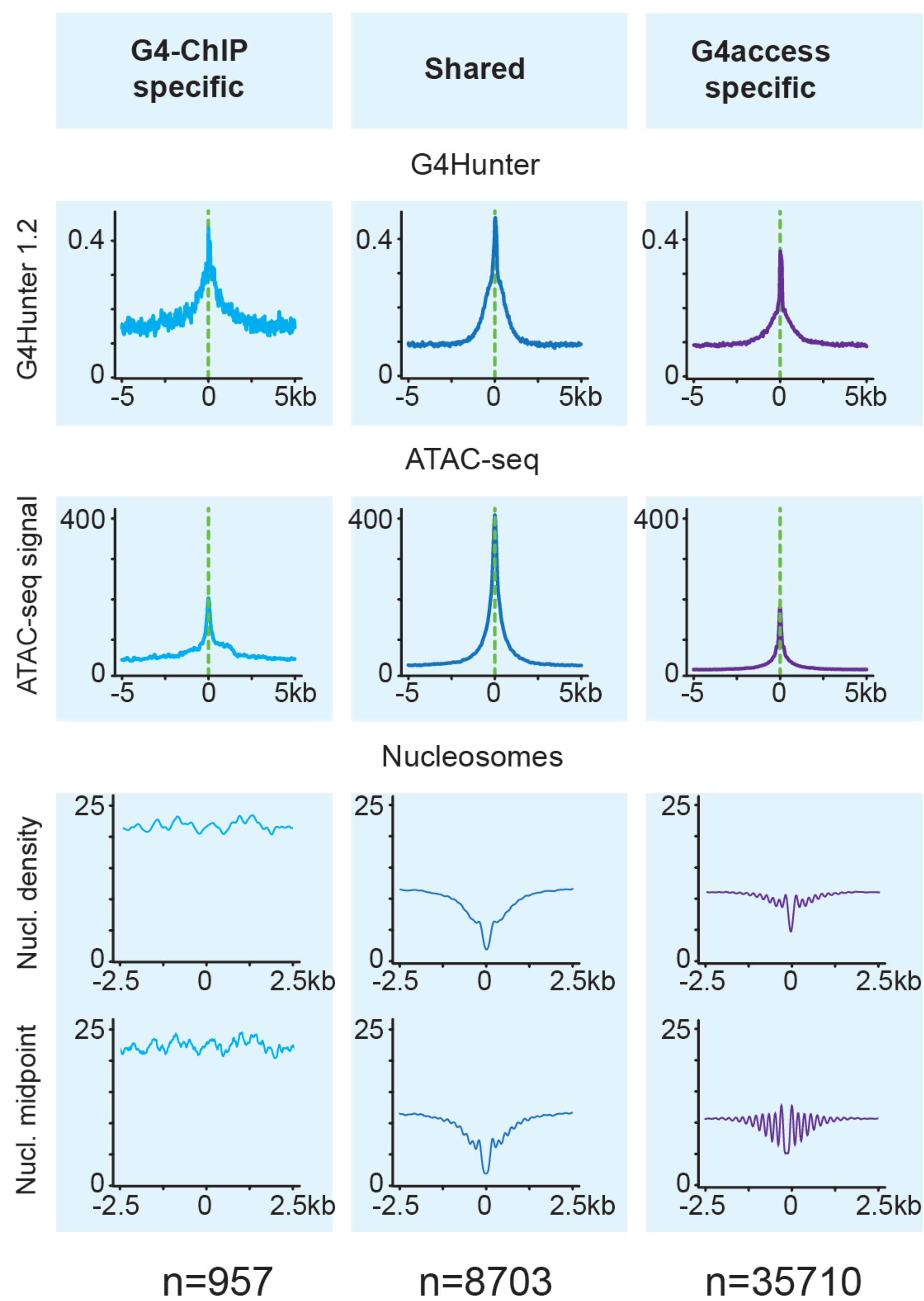
c



ED Figure 8

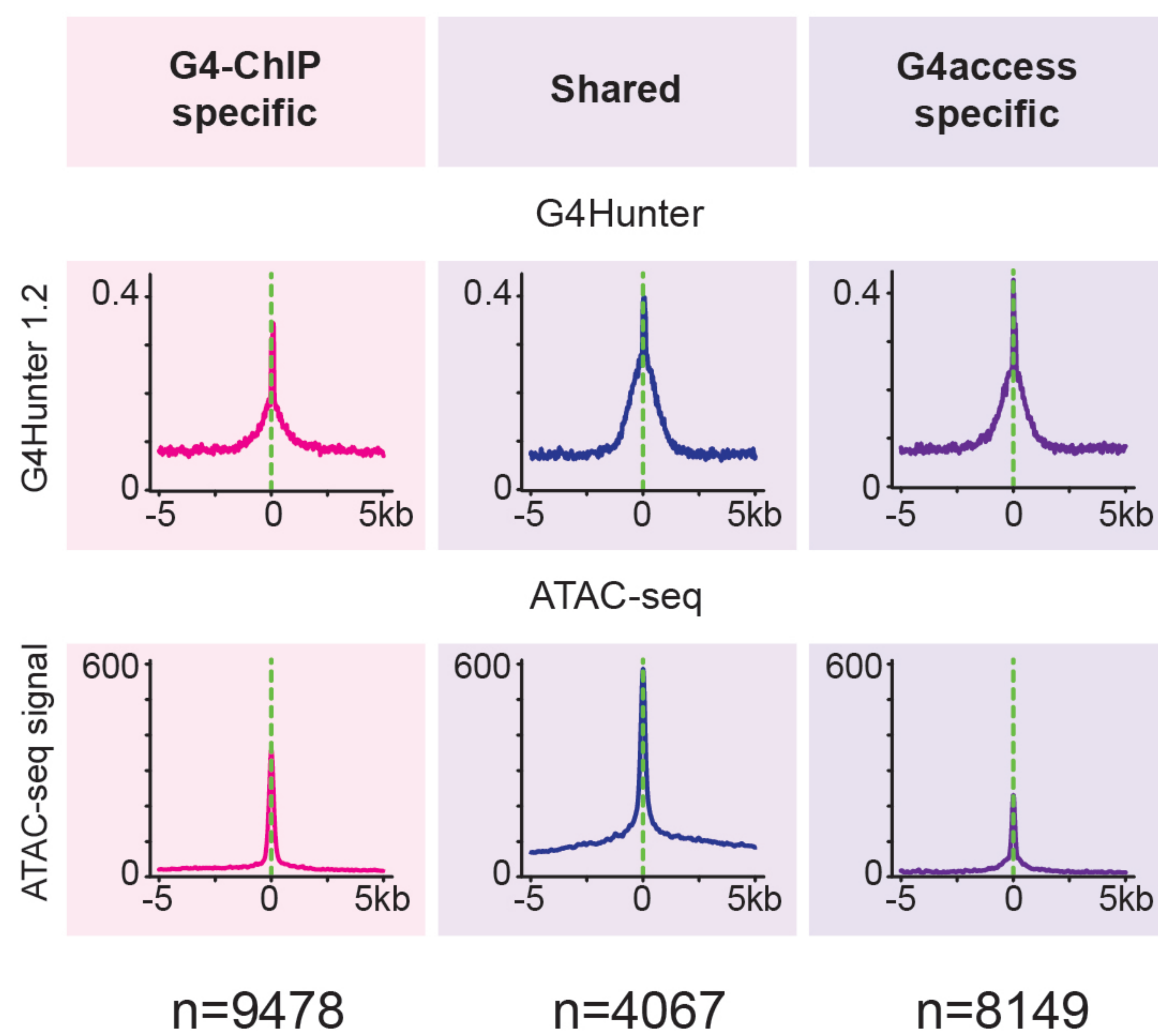
a

K562



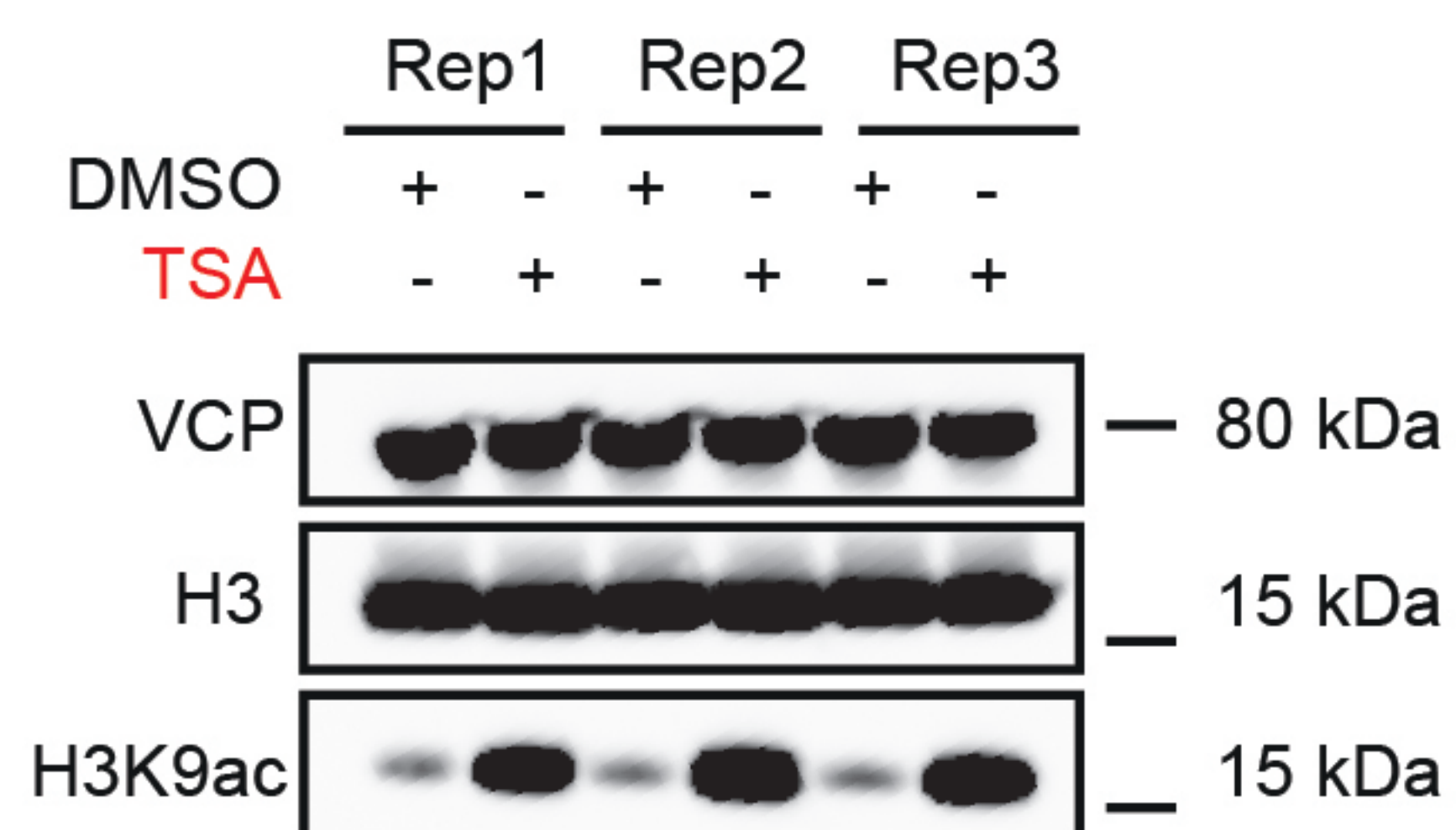
b

HaCaT

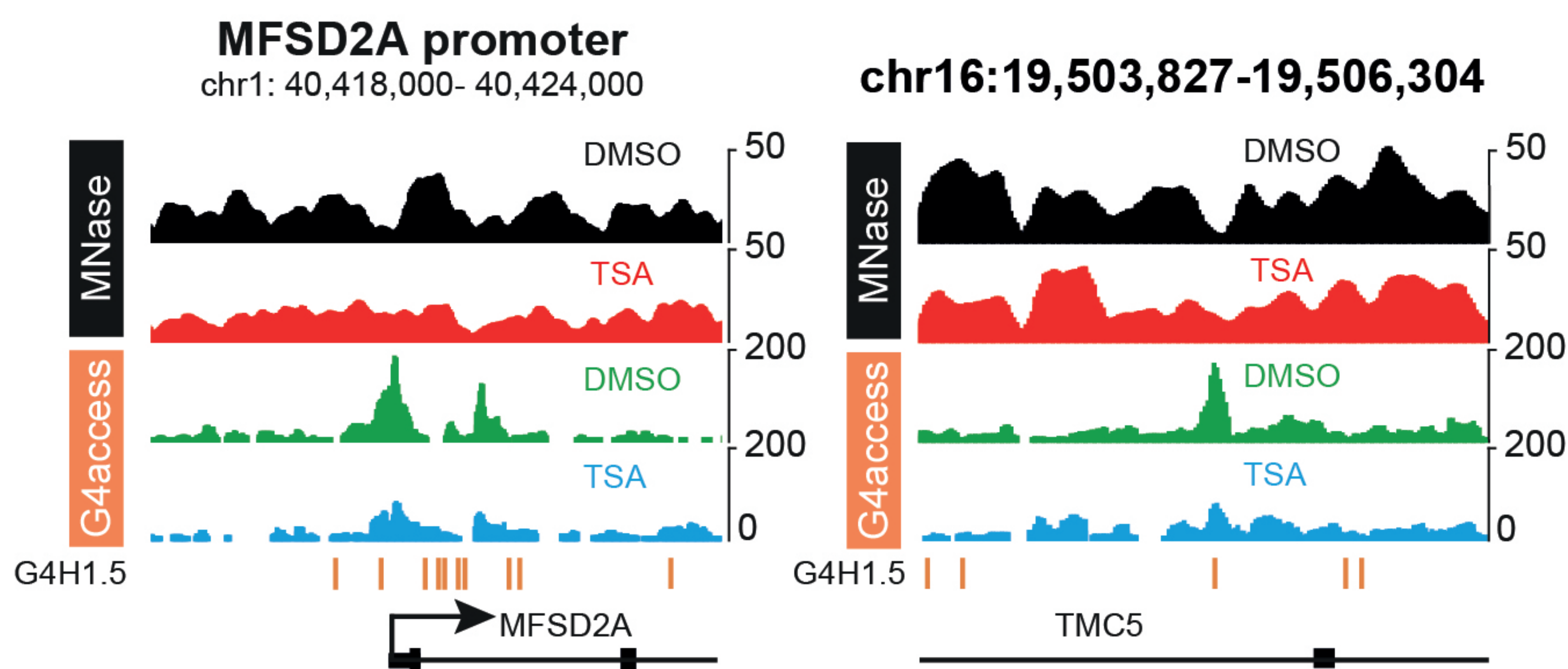


ED Figure 9

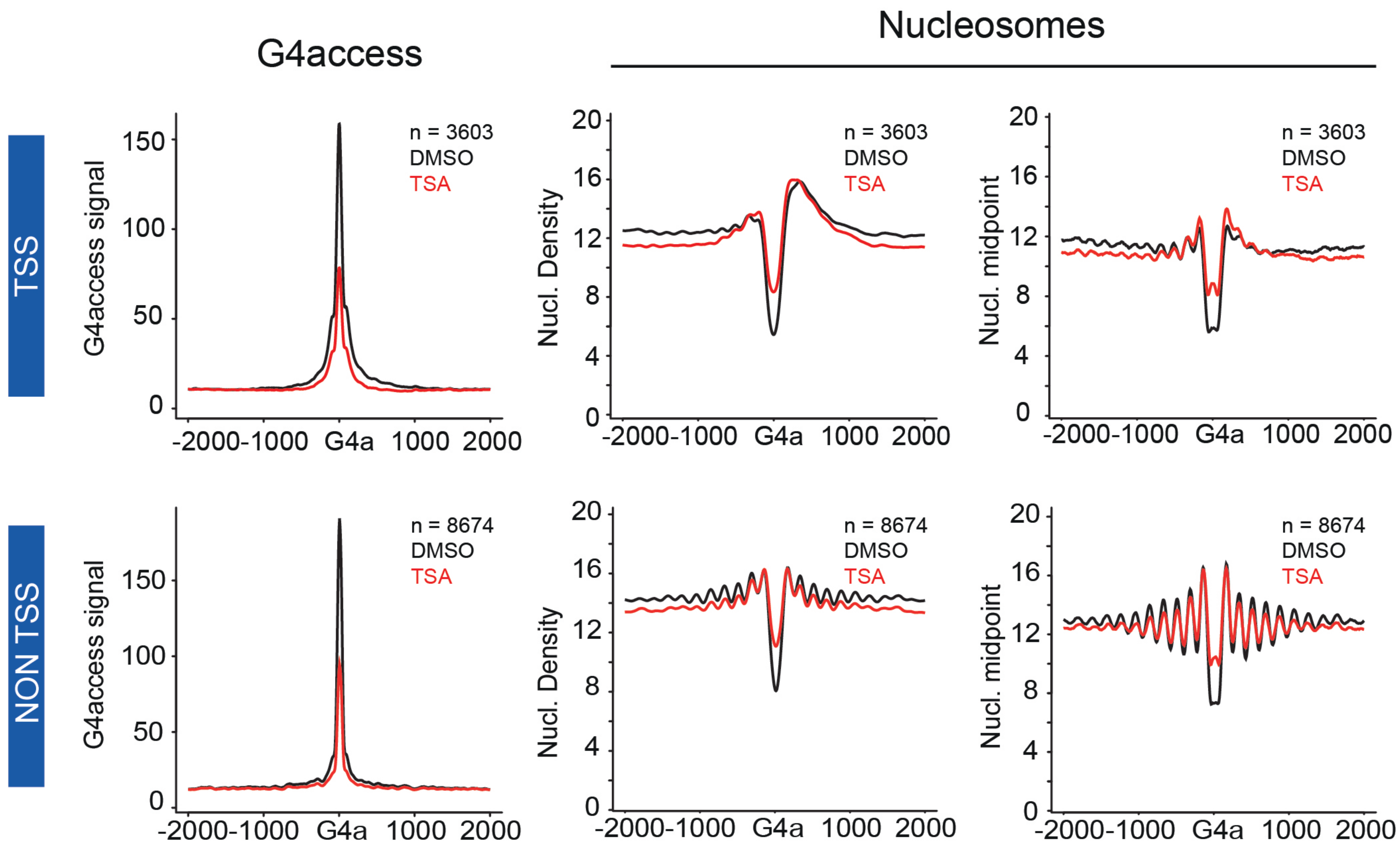
a



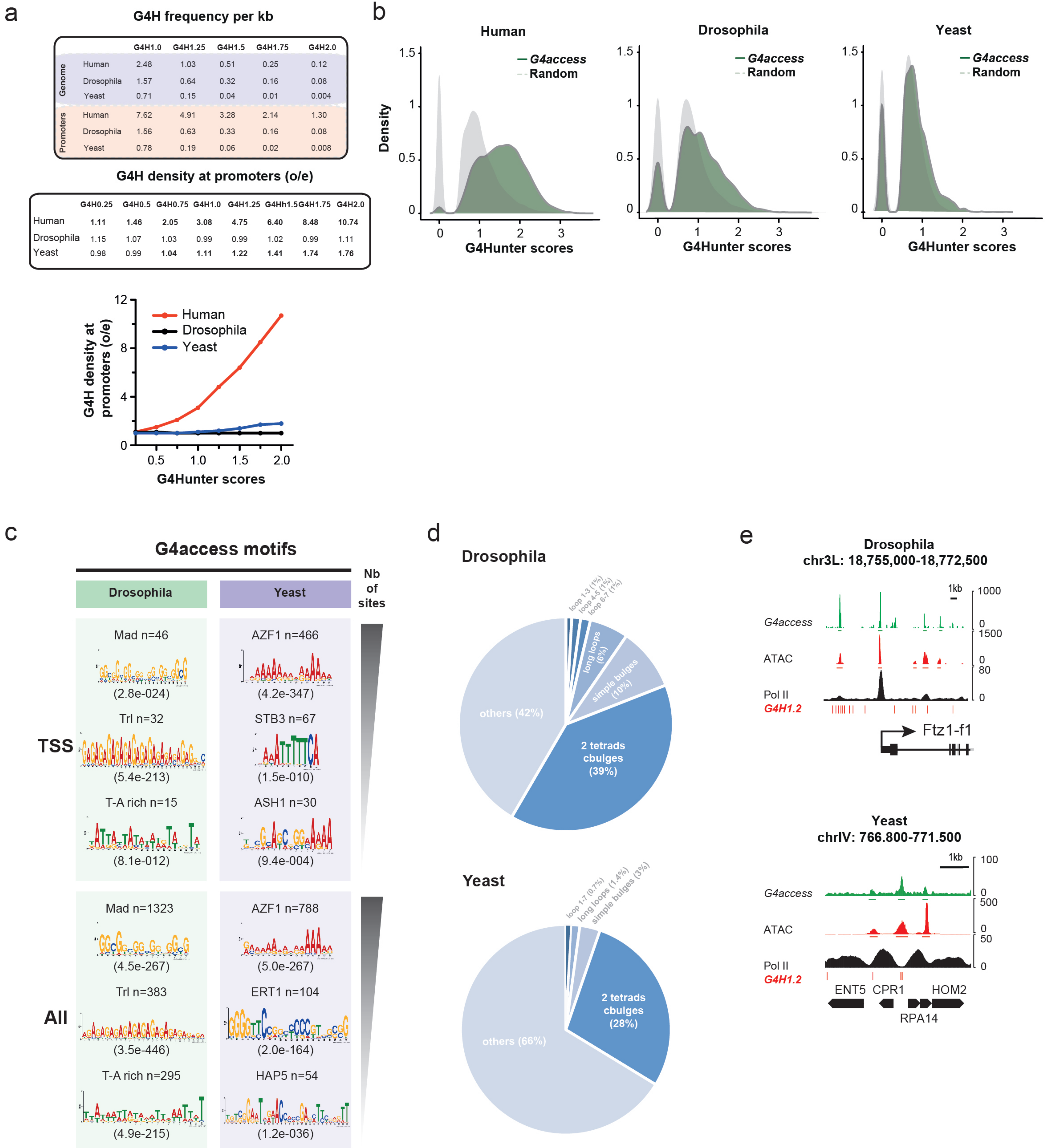
b



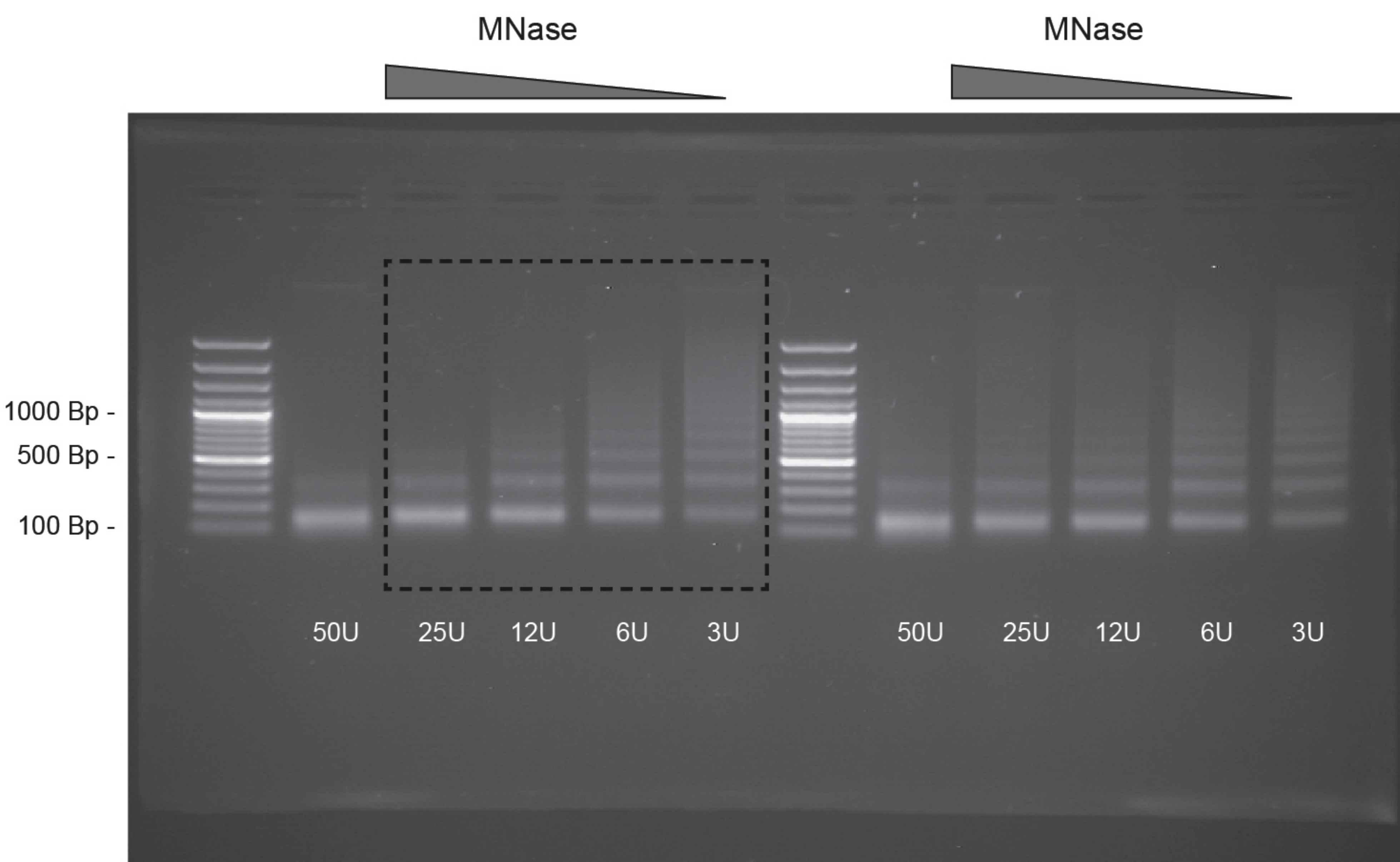
c



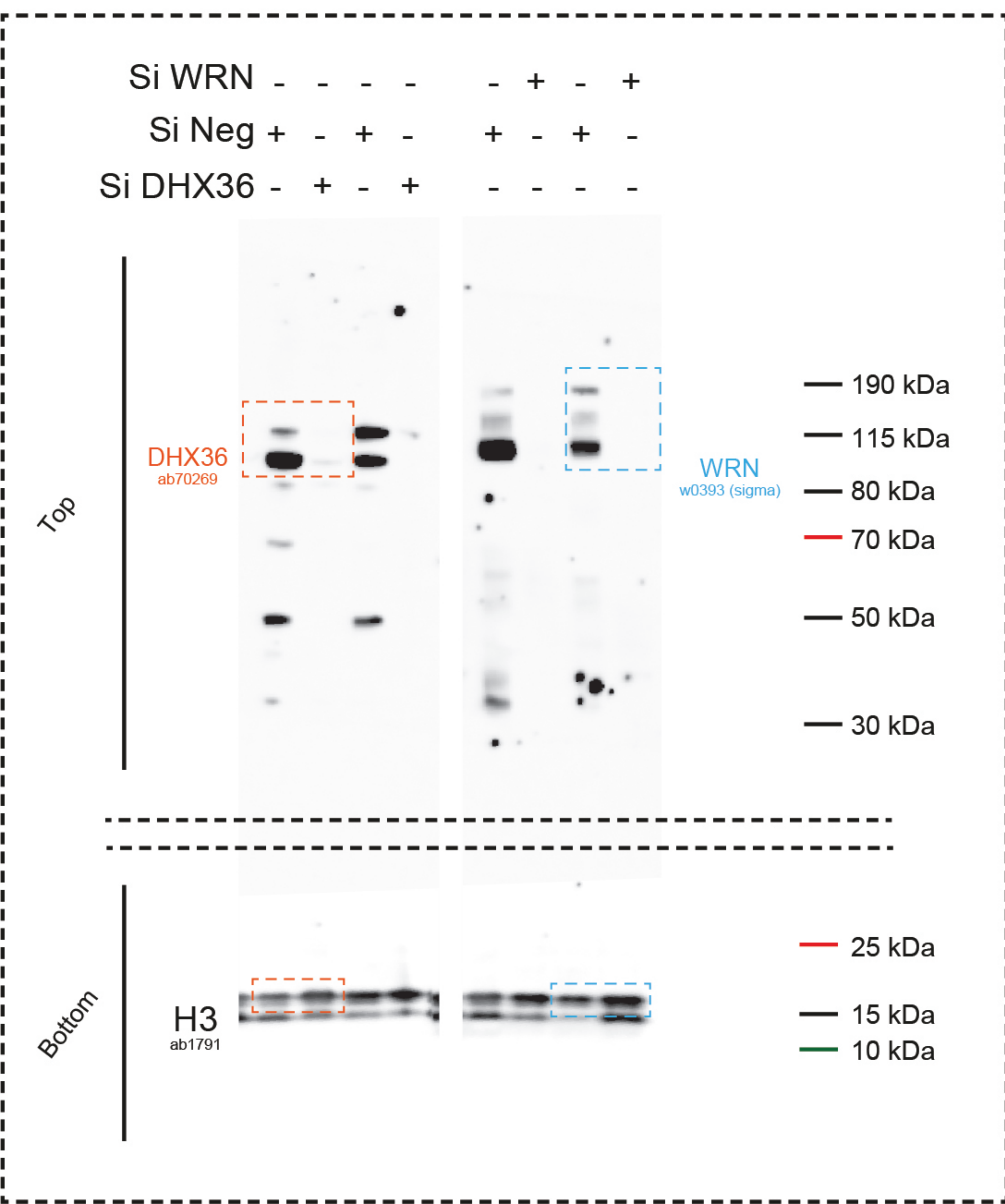
ED Figure 10



source data Fig. 1c



source data Fig. 3a and Fig. 3b



source data Extended Data Fig. 9a

