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-04285615v1

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

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

Cyril Esnault[1]*, Talha Magat[1]*, Amal Zine El Aabidine[1]*, Encar Garcia-Oliver[1]*, Anne Cucchiarini[2], Soumya Bouchouika[1], David Lleres[1], Lutz Goerke[1], Yu Luo[2,3], Daniela Verga[3], Laurent Lacroix[4], Robert Feil[1], Salvatore Spicuglia[5], Jean-Louis Mergny[2], Jean-Christophe Andrau[1#]

[1] Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS-UMR 5535, 1919 Route de Mende, 34293 cedex 5, Montpellier, France.
[2] Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, Inserm, Institut Polytechnique de Paris, Palaiseau, France.
[3] Université Paris-Saclay, Institut Curie, Orsay, France.
[4] Institut de Biologie de l'École Normale Supérieure, ENS, CNRS UMR8197, Inserm U1024, Paris, France.
[5] Aix-Marseille University, INSERM, TAGC, UMR 1090, Marseille, France. Equipe Labellisée Ligue Contre le Cancer.

[#] Corresponding author: jean-christophe.andrau@igmm.cnrs.fr
* These authors contributed equally.

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

Introduction

Eukaryotic promoters encompass a wide range of sequence, but tend to have modest conservation in evolution. For example, while mammalian promoters tend to be GC-rich (>70%), yeast and Drosophila promoters are AT-rich. Despite these pronounced differences, eukaryotic promoters harbor similar properties in their ability to recruit the transcriptional machinery and to exclude and position nucleosomes[1]. We previously showed that mammalian CpG islands (CGIs) intrinsically exclude nucleosomes, independently of transcription[2]. Our more recent work also emphasizes that within CGIs, and more generally in human and mouse promoters, G-quadruplex (G4) forming sequences are likely to play a crucial role in nucleosome exclusion both in cells and *in vitro*[3]. Unimolecular G4s are DNA secondary structures well characterized *in vitro*[4]. Their investigation in living cells is more recent and they are suggested to play essential roles in transcription, replication, genome stability and homeostasis[5]. They can also be predicted by computational algorithms[6,7] such as G4Hunter that calculates a robust likelihood score of genomic sequences forming G4 structures at fixed window sizes.

Various experimental techniques have been developed to characterize G4 formation in cells. The recent development of G4 ChIP has enabled the identification of thousands of G4-forming genomic sequences at promoters and elsewhere[8-10]. However, G4 ChIP is highly dependent on the selectivity of G4 recognition by the BG4 nanobody[11], which may stabilize unstable structures *in vitro*[12,13], introducing a potential bias. Finally, previous observations report that it is difficult to apply G4 ChIP to all cell types, in particular non-cancerous primary cells[8]. Therefore, orthogonal methods are needed to identify G4s formed in the chromatin context.

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

Results

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

We previously showed that CGIs, enriched at mammalian promoters, tend to exclude nucleosomes intrinsically[2]. 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 description[14]. 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 sub-nucleosomal (~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-exonuclease[17]. 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

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

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

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

G4Hunter-predicted G4s were previously validated *in vitro* on a set of mitochondrial sequences[6]. We therefore asked if G4access-identified G4FS could, at large scale, be validated using three independent *in vitro* assays. Firstly, we identified 4743 common G4access peaks in all three cell lines, and then selected 596 representative 30nt regions with G4Hunter scores >0.5 (see methods; Fig. 2a). We next assessed their abilities to form G4 structures by performing FRET-MC[18], Thioflavin T (Th-T) and N-methyl mesoporphyrin IX (NMM) fluorescent ligand assessment (Extended Data Fig. 5a). FRET-MC validated that ~80% of the tested sequences form G4 structures *in vitro*. Moreover, we found G4 formation for 97% of sites with G4Hunter score >1.35 (Fig. 2b and Extended Data Fig. 5b), using ThT and NMM assays (Fig. 2c and Extended Data Fig. 5c-d). Strikingly, FRET-MC, which may be the most robust method, validated 95% of G4access sequences with a score of 1.2 and above (Fig. 2d). This high level of validation exceeds that was previously observed for mitochondrial genome and suggests that G4access further enriches for G4FS.

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

G4access can monitor G4 dynamics in living cells

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

G4s can be targeted by ligands that stabilize them *in vitro*. *In vivo* ligands' mode of action remains relatively enigmatic although it appears that they generate double strand breaks at various genome locations[19]. To get further insights into G4 ligand's action, we treated the cells with Pyridostatin (PDS), a well known G4 ligand[19], for a short time (30 min), to avoid indirect effects, and performed G4access before and after treatment. We found that although many hits are conserved, G4access regions are relatively dynamic after PDS treatment (Extended Data Fig. 6a-b). Many peaks observed, including at promoters, redistributed globally or locally. Surprisingly, we found that the average G4Hunter scores of the PDS-induced G4s were lower, while G4s with high scores tended to decrease (Extended Data Fig. 6c). An example of this

is shown in Extended Data Fig. 6a in which the G4access signals relocate from a strong to a weaker G4FS at the *ATP2B4* promoter. Overall, these results suggest that while strong G4s are only moderately affected by the drug since they are stable, the weaker ones become more enriched in open chromatin areas due to ligand stabilization. However, we could not exclude that the loss of strong G4s might be due to sequencing biases. To address this possibility, we analysed previously published genome-wide *in vitro* data using G4seq experiments with or without PDS[20]. G4seq maps G4s based on error rates incorporation during DNA amplification on purified genomic DNA and thus out of the chromatin context. We found that average scores of G4seq are higher than with G4access and preferential stabilization of the weaker G4s is also observed, although to a lesser extent (Extended Data Fig. 6d), suggesting that weak G4s indeed become preferentially stabilized by brief PDS treatment.

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

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

G4access hallmarks nucleosome exclusion and transcription

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

We next ranked non-promoter G4access regions by increasing nucleosome density (MNase signal) and analyzed nucleosome positioning, Pol II and G4access signal (Fig. 4c and Extended Data Fig. 7a) and defined 4 classes of nucleosome positioning and depletion patterns. We observed that nucleosome positioning is generally conserved in the ¾ of the regions (classes 2 and 3), while a minority of regions in which G4s deplete or position more accurately nucleosomes (class 1 and 4, respectively) do not show such phasing. We speculate that class 1 represents regulatory regions with promoter-like properties while class 4 represents areas in which predicted G4s cannot open chromatin and thus do not affect positioning of their surrounding nucleosomes. In support of this, class 1 regions show stronger H3K4me3 signals as compared to

H3K4me1, consistent with histone mark signature of promoter activity[23] (Extended Data Fig. 7a).

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

Noting the association with transcription, we then asked if G4 formation in open chromatin would be affected by transcriptional perturbation. It has been proposed in the past that transcription induces a torsional stress that may stabilize G4s *in vitro*. Our recent observations suggested the opposite, i.e. that G4s would promote chromatin opening and Pol II recruitment[8,24,25]. To further address this question, we treated the cells with transcription inhibitors specific for initiation or elongation (triptolide and KM05283, respectively) and analysed if this resulted in loss of G4access signal. As previously reported, we found that triptolide treatment stripped Pol II signal across all genic features while KM05283 only removed Pol II from within gene bodies[27] (Fig. 4d, left). Interestingly, at promoter locations elongation blockade did not reduce G4 signals, whereas initiation inhibition reduced G4access levels by half (Fig. 4d, right). At genic locations, distal to the promoter, both inhibitors reduced G4access by half. An example of such variations is shown in Fig. 4e. We conclude that transcriptional inhibition neither abolishes G4 formation, nor its association to open chromatin, but instead reduces it. This suggests a model in which G4 formation precedes Pol II recruitment and becomes further stabilized by the ensuing transcriptional activity. Recent work described similar observations in K562 cells[28] although in this case ChIP-seq did not identify changes in G4 signal in the presence of inhibitors. This difference might originate from the cell types, the techniques used or bioinformatic set-up of the analyses.

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

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

tend to re-position to G4 sites, thus presumably reducing their formation (Extended Data Fig. 9b-c). We note that these results are in contrast with effects observed with another HDAC inhibitor[8,28]. However, when growth condition such as hypoxia result in chromatin compaction at pG4 sites, G-quadruplex formation is also reduced[27].

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

G4s associate with active alleles in hybrid mES cells

Because of their G-rich content, predicted G4s are over-represented at CGIs. We find that the same is true for Differentially Methylated Regions (DMRs) at imprinted gene domains, many of which also correspond to CGIs as well[31]. Because of their parental DNA methylation imprints, DMRs are stably repressed on their methylated allele, and are transcriptionally active on the non-methylated allele[31]. They constitute thus an attractive model to compare repressed and active alleles within the same nuclear context.

We hypothesized that G4 formation in open chromatin might occur specifically at the active, unmethylated alleles of imprinted DMRs. To address this question, we performed G4access in hybrid mouse embryonic stem cells (ESCs) obtained through reciprocal crosses between *M. m. domesticus* C57BL/6J and *M. m. molossinus* JF1[32] strains (BJ and JB; Fig. 5a). These strains are genetically divergent, allowing efficient discrimination of the parental alleles.

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

DNA methylation antagonizes G4 formation and open chromatin

To gain further insight into this question, we analyzed methylation at G4Hunter-predicted regions with or without G4access signal (Fig. 6a-b). These analyses were performed at medium G4 predictions (G4H1.2) that contained at least 2 CpG nucleotides. We found that experimentally-identified G4s are associated with a loss of CpG methylation, consistent with previous observations[10]. Yet, we also found that the

G4s corresponding to weak/medium predictions tend to be more methylated when compared to neighbouring DNA. These results suggest that G4 formation favours open chromatin formation, while pG4 DNA methylation antagonizes this process. To further address this issue at the most basic level, in the context of naked chromatin, we re-analyzed data from *in vitro* nucleosome assembly using human granulocyte genomic DNA[33]. We ranked the *in vitro* nucleosome densities at pG4s by increasing signal (Fig. 6c) and analysed corresponding signals for *in vivo* nucleosomes and DNA methylation. This analysis indicates a clear relationship between methylation levels at pG4s and nucleosome formation *in vitro* and *in vivo* (Fig. 6d). These results propose that high methylation of G4FS results in a sequence context favourable to intrinsic nucleosome formation and G4 structure destabilisation.

G4access enriches for weaker G4s in lower eukaryotes

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

We adapted our G4access protocol in S2 drosophila cells and in a S288C yeast strain, following an adjusted workflow (see methods). We compared the procedure's performance to isolate G4FS in the 3 organisms by plotting the G4access's prediction scores (Extended Data Fig. 10b). As in mammalian cells, we found an enrichment of G4s in the sequenced fragments. However, and as expected from the genomic densities of G4 predictions, we observed lower G4Hunter scores in drosophila and far lower in yeast. Further investigation of the DNA motifs present in G4access data indicated G-rich and GAGA motifs in drosophila and A/Tor G/C-rich sequences in yeast (Extended Data Fig. 10c). Moreover, 60% and 34% of potential G4 subtypes were found in Drosophila and in yeast, respectively (Extended Data Fig. 10d). When browsing the G4access locations in the genomes, our data confirmed that G4FS were more enriched at promoters in yeast as compared to Drosophila. As expected, in both cases, G4access locations correlated with open chromatin ATAC-seq signal (Extended Data Fig. 10e).

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

Discussion

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

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

Our investigations suggest a role for G4s in chromatin opening, nucleosome positioning, Pol II transcription and imprinting control, expanding the possible functions of these secondary DNA structures in the genome. Among these properties, it is remarkable that G4s harbor the apparent ability to position nucleosomes with a phasing comparable to that of insulators[3,35]. We further show that G4s are maintained without active transcription, albeit reduced. This indicates that G4s may be formed prior to transcription and that transcription further stabilizes their structure, reconciliating previous conflicting observations[8,28]. This result is also consistent with the idea that negative supercoiling upstream of the transcription front increases G4 stability[24].

Our study also finds that G4s might contribute to imprinted gene expression. Amongst the DMRs analyzed, all those displaying an allele-specific expression or active unmethylated state were shown to display an allele-specific G4access signal. This suggests that G4s play an activating role in DMRs/CGIs that are not methylated. Our data also indicate that while G4s formed in a given cell antagonize methylation, local CpG methylation also disfavour G4 formation and their related nucleosome exclusion property, providing a possible model for paternal or maternal allele inactivation at DMR. In agreement with this, it was previously shown that CpG methylation results in loss of chromatin opening and phasing at CGIs[36], both features associated with G4FS in the present study. Nevertheless, we note that methylation on G4 oligonucleotides does not clearly modify the properties of G4 formation *in vitro*[37-39]. This could originate from the difference of topology between oligonucleotides and dsDNA embedded in nucleosome *in vivo*. In the latter, topological constraints could result in more sensitivity to DNA methylation on G4 structure, a possibility that requires further investigation.

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

In summary, the G4access technique will enable future investigations to better understand the role of G4s in transcription control but also other genomic processes such as replication[40], DNA repair and the role of helicase in genome stability[41], as well as deepening the mechanistic understanding of various G4-targeting drugs, including some that are believed to display an anti-cancer potential[42].

Acknowledgements:

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

Author contribution statement

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

Competing Interests statement

The authors declare no competing interests.

Figure legends

Figure 1: G4access principle and validation.

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

initial set-up. Bottom left is the fraction of mononucleosomes in the titration curve expressed as percent of all DNA (Mean and SD of n=2 biological replicates are displayed). The 2nd point (30% mononucl.) is considered as the optimal for G4 enrichment in our experimental frame. The right panel indicates mono- (146 bp), di- (320 bp) and multi-nucleosomal Bioanalyzer profiles. **d-** Genome browser view (Chr7: 7.100.000-7.900.000) of G4access signal and corresponding G4-ChIP[8,10] in indicated cell lines. G4H2.0 predictions are indicated below. **e-** Peak size distribution in G4-ChIP and G4access. Vertical dashed lines indicate the maximum (bulk) sizes of fragments in the sequenced samples. **f-** Observed/expected G4Hunter predictions in G4access/ChIP in human cell lines. **g-** G4Hunter prediction scores in G4access and equivalent selection of random genomic DNA fragments. Around 75% of G4access peaks are >G4H1.2, which represents a likelihood >85% of forming a G4 *in vitro*[6]. The difference of distribution is highly significant (*p*-value < $2.2E^{-16}$, two-sided Wilcoxon test). **h-** Repartition of the G4 subtypes in G4access peaks in Raji cells. 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).

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

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

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

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

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

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

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

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

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

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

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

References

1 Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* **10**, 161-172, doi:10.1038/nrg2522 (2009).

2 Fenouil, R. *et al.* CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res* **22**, 2399-2408, doi:10.1101/gr.138776.112 (2012).

3 Esnault, C. *et al.* G-quadruplexes are promoter elements controlling nucleosome exclusion and RNA polymerase II pausing. *bioRxiv*, 2023.2002.2024.529838, doi:10.1101/2023.02.24.529838 (2023).

4 Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet* **13**, 770-780, doi:10.1038/nrg3296 (2012).

5 Hansel-Hertsch, R., Di Antonio, M. & Balasubramanian, S. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat Rev Mol Cell Biol* **18**, 279-284, doi:10.1038/nrm.2017.3 (2017).

6 Bedrat, A., Lacroix, L. & Mergny, J. L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic acids research* **44**, 1746-1759, doi:10.1093/nar/gkw006 (2016).

7 Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic acids research* **33**, 2908-2916, doi:10.1093/nar/gki609 (2005).

8 Hansel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat Genet* **48**, 1267-1272, doi:10.1038/ng.3662 (2016).

9 Li, C. *et al.* Ligand-induced native G-quadruplex stabilization impairs transcription initiation. *Genome Res* **31**, 1546-1560, doi:10.1101/gr.275431.121 (2021).

10 Mao, S. Q. *et al.* DNA G-quadruplex structures mold the DNA methylome. *Nat Struct Mol Biol* **25**, 951-957, doi:10.1038/s41594-018-0131-8 (2018).

11 Ray, S. *et al.* Custom DNA Microarrays Reveal Diverse Binding Preferences of Proteins and Small Molecules to Thousands of G-Quadruplexes. *ACS Chem Biol* **15**, 925-935, doi:10.1021/acschembio.9b00934 (2020).

12 Tran, P. L. T. *et al.* Folding and persistence times of intramolecular G-quadruplexes transiently embedded in a DNA duplex. *Nucleic acids research* **49**, 5189-5201, doi:10.1093/nar/gkab306 (2021).
13 van Holde, K. & Zlatanova, J. Unusual DNA structures, chromatin and transcription. *Bioessays* **16**, 59-68, doi:10.1002/bies.950160110 (1994).
14 Hershman, S. G. *et al.* Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. *Nucleic acids research* **36**, 144-156, doi:10.1093/nar/gkm986 (2008).
15 Dingwall, C., Lomonossoff, G. P. & Laskey, R. A. High sequence specificity of micrococcal nuclease. *Nucleic acids research* **9**, 2659-2673, doi:10.1093/nar/9.12.2659 (1981).
16 Horz, W. & Altenburger, W. Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic acids research* **9**, 2643-2658, doi:10.1093/nar/9.12.2643 (1981).
17 Foulk, M. S., Urban, J. M., Casella, C. & Gerbi, S. A. Characterizing and controlling intrinsic biases of Lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Res*, doi:10.1101/gr.183848.114 (2015).
18 Luo, Y., Granzhan, A., Verga, D. & Mergny, J. L. FRET-MC: A fluorescence melting competition assay for studying G4 structures in vitro. *Biopolymers*, ebip23415, doi:10.1002/bip.23415 (2020).
19 Rodriguez, R. *et al.* Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat Chem Biol* **8**, 301-310, doi:10.1038/nchembio.780 (2012).
20 Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nature biotechnology* **33**, 877-881, doi:10.1038/nbt.3295 (2015).
21 Sun, H., Karow, J. K., Hickson, I. D. & Maizels, N. The Bloom's syndrome helicase unwinds G4 DNA. *J Biol Chem* **273**, 27587-27592, doi:10.1074/jbc.273.42.27587 (1998).
22 Vaughn, J. P. *et al.* The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J Biol Chem* **280**, 38117-38120, doi:10.1074/jbc.C500348200 (2005).
23 Natoli, G. & Andrau, J. C. Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet* **46**, 1-19, doi:10.1146/annurev-genet-110711-155459 (2012).
24 Xia, Y. *et al.* Transmission of dynamic supercoiling in linear and multi-way branched DNAs and its regulation revealed by a fluorescent G-quadruplex torsion sensor. *Nucleic acids research* **46**, 7418-7424, doi:10.1093/nar/gky534 (2018).
25 Esnault, C. *et al.* G-quadruplexes are promoter elements controlling nucleosome exclusion and RNA polymerase II pausing. *Manuscript in review to Cell and Molecular Cell* (2023).
26 Jonkers, I., Kwak, H. & Lis, J. T. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**, e02407, doi:10.7554/eLife.02407 (2014).
27 Medlin, J. *et al.* P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *Embo j* **24**, 4154-4165, doi:10.1038/sj.emboj.7600876 (2005).

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

Methods

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

Cell culture and treatments

K562 (human myelogenous leukemia, gift of E. Soler) and Raji (human Burkitt's lymphoma, gift from D. Eick) cell lines were cultured in RPMI 1640 medium (Thermo Fisher 11875085) supplemented with 10% heat-inactivated fetal bovine serum (Sigma 12103C), 100 units/mL penicillin, and 100 µg/mL streptomycin (Thermo Fisher 15140122) at 37°C in a humidified 5% CO_2 incubator. Human Raji cells, grown at an approximative density of 10^6 cells/ml, were treated with 10 µM pyridostatin (PDS; Sigma SML0678) or H_2O (control) for 30 minutes, with 2 µM TSA (Sigma, T8552) or DMSO (control) for 24 hours for G4access experiment. TSA efficiency was tested by western-blot on H3K9ac histone acetylation (Ab Millipore, CS200583). For transcription inhibition experiments cells were treated for 2 hours with DMSO 0.1% (control), 1 µM triptolide (inhibiting initiation stage; Sigma Merck) or 100 µM KM05283 (inhibiting elongation stage; Thermo Fisher Scientific) prior collection for G4access or Pol II ChIP-seq. HaCaT (spontaneously immortalized human keratinocyte line, gift from the IRCM screening platform in oncology) was cultured in DMEM-high glucose medium (Thermo Fisher 11965084) supplemented with 10% heat-inactivated fetal bovine serum (Sigma 12103C), 100 units/mL penicillin, and 100 µg/mL streptomycin (Thermo Fisher 15140122) at 37°C in a humidified 5% CO_2 incubator.
HeLa cells (gift from E. Bertrand) used for siRNA knockdown assays were maintained in DMEM supplemented with 10% fetal calf serum, penicillin/streptomycin (100 units/L) and glutamin (2.9 mg/L) at 37°C and 5% CO2. Knockdown experiments were performed using Invitrogen siRNA Silencer® Select (siDHX36 and siWRN, 4392420) control siRNA (Dharmacon, ON-TARGET*plus* Non-targeting Pool) and lipofectamine RNAiMAX (Thermo Fisher Scientific) following manufacturer instructions. Knockdown efficiencies were analysed by westernblot using specific antibodies (DHX36: Abcam, ab70269, WRN: Sigma, W0393).

The 2i-medium-derived ESCs hybrid between *M. m. domesticus* strain C57BL/6J and *M. m. molossinus* strain JF1 were derived recently[32]. The two chosen lines, BJ (full laboratory name BJ-WT3) and JB (full laboratory name JB-WT2), are both male, both with a normal karyotype. They were cultured for cell harvesting on 0.1% gelatin-coated dishes (Merck-Millipore SF008) in serum-free ESGRO Complete PLUS medium (Millipore, with LIF and GSK3 inhibitor, SF001) at 37°C in a humidified 5% CO_2 incubator.

Drosophila S2 cells (gift from J. Tazi) were cultured in Schneider's S2 Drosophila medium (Dominique Dutscher, L0207-500) supplemented with 10% heat-inactivated fetal bovine serum (Sigma, 12103C), 100 units/mL penicillin, and 100 µg/mL streptomycin (Thermo Fisher, 15140122) at 27°C and harvested at $2x10^6$ cells/mL. Yeast S288C (BY4741, gift from M. Radman-Livaja) haploid cells were grown up to and collected at mid-log phase (OD_{600}=0.5) in 100mL of YPD (2% glucose).

G4access

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

saline buffer (PBS). HaCaT cells were further subjected to mechanical permeabilization in 5 mL of permeabilization buffer (150 mM of sucrose, 80 mM KCl, 5 mM KH_2PO_4, 5 mM $MgCl_2$, 0.5 mM $CaCl_2$ and 35 mM HEPES pH 7.4) and homogenized using a dounce (40 strokes). HeLa cells tend to clump under such condition. Therefore, nuclei were first purified under sucrose cushion as described[43]. For this, cells were resuspended on ice-cold nuclei buffer#1 (15 mM Tris-HCl pH 7.5, 300 mM sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM ethylene glycol-bis N,N,N',N'-tetra-acetic acid (EGTA), 0.5 mM dithiothreitol (DTT), 0.1 mM phenylmethylsulfonyl fluoride (PMSF), 3.6 µg/ml aprotinin) before addition of ice-cold buffer#2 (15 mM Tris-HCl pH 7.5, 300 mM sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 3.6 µg/ml aprotinin, 0.4% IGEPAL CA-630) and then were carefully layered on ice-cold nuclei buffer #3 (15 mM Tris-HCl pH 7.5, 1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 3.6 µg/ml aprotinin). Cells were then centrifuged at 10.000g for 20 min at 4 ºC. The supernatants were discarded and the nuclei-containing pellets were resuspended in MNase digestion buffer for later steps (see below). For all mammalian cell lines, $5x10^6$ cells per titration points were then re-suspended in 50 µL of prewarmed permeabilization buffer supplemented with 0.2% (v/v) NP40 and incubated for 5 minutes at 37ºC prior digestion. The same procedure was performed for $50x10^6$ cells for S2 Drosophila cells. MNase digestions, were then performed by adding a volume of 500 µL of prewarmed MNase reaction buffer (150 mM sucrose, 50 mM Tris-HCl pH 8, 50 mM NaCl and 2 mM $CaCl_2$) supplemented with either 3, 6, 12, 25 or 50U of MNase (Merck, 10107921001). Digestions were incubated at 37ºC for 10 min and stopped on ice and by adding 11 µL of 500 mM EDTA to each reaction followed by addition of 550 µL of SDS lysis buffer (1% (v/v) SDS, 10 mM EDTA and 50 mM Tris.HCl pH 8). Samples were then incubated 10 minutes on ice. Before DNA purification, 1 mL of water was added to dilute the SDS and the samples were incubated with 5 µL of RNAse A (ThermoFisher, EN0531) at 37 ºC for 2 hours and with 8 µL of proteinase K (Euromedex, 09-0911) at 56 ºC for 2 hours to complete the lysis. To then quality control the MNase digestions: 125 µL of each sample were cleaned-up using QIAquick PCR Purification Kit (QIAGEN, 28106) and assessed by agarose gel and Bioanalyzer 2100, using High-sensitivity DNA Chips (Fig. 1c). At this step, for efficient G4access, samples should present ~30% (+/-5%) of mono-nucleosomes (Fig. 1c). Importantly, this assessment should be performed on purified DNA that does not contain the subnucleosomal fraction, using a bioanalyzer equipment. The remaining of the samples was then purified by phenol-chloroform and ethanol precipitation for subsequent steps. We recommend that, when implementing this method, a wide range of MNase concentrations shall be tested in a first round of preparative experiments to narrow the condition in which the critical fraction of 30% of mononucleosome, as compared to the whole nucleosomal fraction, shall be obtained. Our experiences showed this fraction is on average optimal for best G4 sequence recovery. We also recommend that the bioanalyser and agarose gel QCs shall be used to assess that MNase digestion and lysis are as complete as possible. Whenever chromatin is not properly digested a smear occurs on top of the digestion pattern (high molecular weights from non-permeabilized nuclei) that does not get resolved by increasing digestion time or MNase concentration. Such samples are typically discarded and it is preferable in this case to check that cell lysis is as complete as possible under the microscope so that all nuclei can be digested similarly by the enzyme.

For yeast, a distinct procedure was used. Cells were pelleted and rinsed twice in phosphate-buffered saline buffer (PBS). Pellets from 100mL culture were suspended in 600 µl of cell breaking buffer (20% glycerol, 100 mM Tris pH 7.5), 600 µl zirconia beads (0.5mm), 10 µl of 100x protease inhibitors (Roche, 06538282001). Beads beating was performed in 1.5mL screw cap tubes in a Bullet Blender (Next advance) for 4 x 3min at a strength of 8 at 4°C. Cell suspensions were recovered by centrifugation; a tube was punctured (23 gauge syringes) and the nuclei were collected in collection tube by centrifugation at 170 rcf. The nuclei fraction was then centrifuged for 5min at 20G and the pellets which contained the nuclei were resuspended in 300µl of prewarmed NP buffer (50 mM NaCl, 10 mM Tris pH 7.4, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.2% NP40 (v/v), 0.5 mM spermidine (Sigma, S0266-1G), 0.007% β-Mercaptoethanol (Sigma, M3148-100ML). Digestion took place by adding a volume of 300µL prewarmed NP buffer supplemented with either 60u, 30u, 15u, 7.5u, 3.75u, 1.9u, or 1u of MNase (Merck 10107921001). Digestions were stopped on ice and by adding 150 µL of stop buffer (5% SDS, 50 mM EDTA). Before DNA purification, digestions were incubated with 5 µL of RNAse A (ThermoFisher, EN0531) at 37 ºC for 2 hours and with 10 µL of proteinase K (Euromedex, 09-0911) at 56ºC for 2 hours. Purification was performed with two consecutive phenol and one chloroform steps followed by ethanol and linear acrylamide precipitation. Importantly, purified DNAs were once again incubated with 5µl RNAse A (ThermoFisher, EN0531) to get rid of persistent RNA contaminations. As for other cells described above, a criterion of 30% mononucleosome was used for the choice of the subnucleosomal fraction to be sequenced. We note that the absence of K+ ions in this yeast-specific procedure might impact an optimal G4 recovery.

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

G4access library preparation

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

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

ChIP-seq

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

For Pol II ChIP, Protein-G coated Dynabeads were incubated at 4°C in blocking solution (0.5% BSA in PBS) carrying Pol II F12 (Santa Cruz 10 sc-55492, lot H2019) specific antibodies. Sonicated chromatins ($10x10^6$ Raji cells equivalent and $5x10^6$ drosophila S2 cells equivalent (2:1 spike-in ratio)) was added to pre-coated beads (250 µL) and the mix was incubated overnight at 4°C on a rotating wheel. After incubation with chromatin, beads were washed 7 times with Wash buffer (50 mM HEPES pH 7.6, 500 mM LiCl, 1 mM EDTA pH 8, 1% NP-40, 0.7% Na-Deoxycholate, 1X protease inhibitor cocktail) followed by one wash with TE-NaCl buffer (10 mM Tris pH 8 and 1 mM EDTA pH 8, 50 mM NaCl) and a final wash with TE buffer (10 mM Tris pH 8 and

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

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

Nascent Chr-RNA-seq

Chromatin-associated RNAs (ChrRNAs) presented in ED Figure 3g were either analyzed from GSE90238 or isolated from $2x10^7$ Raji cells as follows. Nuclear fractionation was performed by incubating cells in 4mL of Buffer I (10mM Tris-HCl pH 7.5, 10 mM NaCl, 2.5mM MgCl2, 0.5% IGEPAL CA-630) on ice for 5 minutes. Next, we carefully underlaid 1 mL of Buffer II (10mM Tris-HCl pH 7.5, 10 mM NaCl, 2.5mM MgCl2, 0.5% IGEPAL CA-630, 10% sucrose) before harvesting the nuclear fraction at 1400 rcf for 5 minutes at 4°C. Nuclei were resuspended with 125µL of NUN1 buffer (20mM Tris-HCl pH 8, 75mM NaCl, 0.5mM EDTA, 50% Glycerol) followed by 1.2mL of NUN2 buffer (20mM HEPES-KOH pH 7.6, 300mM NaCl, 0.2mM EDTA, 7.5mM MgCl2, 1% IGEPAL CA-630, 1M Urea). After 15 minutes of vigorous vortexing, the chromatin was centrifuged at 15000 rcf for 10 minutes at 4°C and washed with 500µL of NUN2 buffer. After discarding the supernatant, the chromatin was resuspended in 500µL of Trizol. At this stage, the chromatin pellet is very tight and needs to be dissolved in Trizol by repeated pipetting with decreasing volume tips (1mL-200µL-10µL) then pushing through very small syringe needles. Then RNA was extracted from chromatin according to the Trizol manufacturer's protocol and resuspended in 50µL of nuclease-free water (Invitrogen, ref#AM9906) followed by TurboDNase (Invitrogen, ref#AM2238) treatment. Trizol RNA extraction and TurboDNase were performed two more times and RNA was resuspended in 20µL of nuclease-free water. Purified RNAs were quantified by Qubit and quality was assessed using the RNA Assay kit (Agilent RNA 6000 Pico reagents, ref#1567-1513) with Bioanalyzer 2100 (Agilent Technologies, USA). ChrRNAs were then subjected to library preparation using the True-seq stranded total RNA library prep gold kit (Illumina, ref#220599) using 1 µg of ChrRNA, with a total of15 cycles of amplification and following the manufacturer's instructions (including ribo-depletion).

Quantitative PCR analysis of targeted G4s

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

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

FRET melting profiles

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

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

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

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

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

Bioinformatic procedures

Genomic data set processing, peak calling and differential analysis

For both native and re-analyzed published data sets considered in this study, raw sequencing reads were aligned using Bowtie2[44] (version 2.1.0) to the human (hg19), mouse (mm9), drosophila (dm6) and yeast (sc3) genomes. Aligned reads were elongated *in silico* using the DNA fragment size inferred using an in-house developed PASHA (version 0.99.21) R (version 3.3.1) pipeline[45] or using MACS2[46] (version 2.1.2) which also allows peak calling for G4-ChIP and G4access (peaks were considered below a *p*-value of 10^{-10} from the narrow peaks table). PASHA was used for ChIP-seq and MNase-seq datasets, using drosophila spike-in for ChIP normalization[47] (Fig. 4d), and MACS2 was used for G4access and G4 ChIP for the sake of consistency in comparison with previously published G4 ChIP analyses. MACS2 was run using input DNA as control and with recommended settings[46]. Bedgraph files generated by MACS2 were then converted to wig files (bin10) and scaled using the sequencing depth with PASHA. Wiggle files representing average enrichment score every 10bp were generated. For nucleosome positioning analyses (midpoints), to determine the average nucleosome positions, wiggle files representing the central nucleotides of DNA fragments were also generated (Fig. 1a, 4a and Extended Data Fig. 6a, 7a, 9c). Finally, for nucleosome densities representation and analyses (MNase-seq), we smoothed the signal by replacing each 10bp bin by the average of the 5 surrounding bins on each side and using an in-house script.

To assess differences in G4access peak intensities, DESeq[48] (version 1.26.0) was used with the MACS2 peak definition as genomic references. Differences between the conditions (control versus treated cells) were called at a *p*-value below 0.05.

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

Motif search analysis and NDRs determination

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

exclude nucleosomes[2]. NDRs and deepest NDR points were defined using an in-house script. First by creating an inversed track of the MNase-seq signal (y=1/x, 0 values were replaced by the minimal value found in the region -300/+100 of annotated TSSs in the sequencing signals). Then, peaks and peak summits corresponding to NDRs and their deepest points were determined. The peak-calling was performed using wigpeakcaller[49] (https://doi.org/10.7554/eLife.02105.022) fixing a threshold based on the peak height and the gap between two adjacent signals (Fig. 1a and Extended Data Fig. 7c). *De novo* motif discovery analyses were thus performed at the lowest nucleosome density (deepest point of MNase-seq signal) at promoters overlapping CGIs between -500 to TSS (Fig. 1a) and on G4access or G4-ChIP (Extended Data Fig. 4) datasets using MEME-ChIP[50] (default settings, version 5.0.2). Fragments from -30 to +30 bp centered on the deepest point of MNase-seq signal or peak summits for the other datasets were used and tested using the jaspar 2020 core non redundant database. Fragments overlapping the annotated TSS region (-200 bp to the TSS) were defined using Bedtools (version 2.21.0). For yeast datasets (Extended Data Fig. 10a), TSS determined by CAGE were used (http://www.yeastss.org/download/). The First 3 motifs are displayed ranked by site numbers (Fig. 1a, Extended Data Fig. 4 and 10c).

Allelic data analysis of mESC

Raw sequencing reads were aligned strictly to the mm9 or JF1[51] genomes using Bowtie2. Reads overlapping distinguishable SNPs (https://molossinus.brc.riken.jp/mogplus/#JF1) between the two genomes were considered to assess allelic signals of G4 access and RNA-seq datasets. At these loci, reads overlapping SNP from each allele were determined and counted using HTSeq-count (version 0.6.1p1) at G4access peaks (for differential G4access analysis) and within gene annotations (for differential gene expression analysis). Assessment of statistical analyses of differential signals were performed using DESeq[48] (version 1.26.0) (Fig. 5b-d; Supplementary Table 1).

The list of 31 murine imprinted DMRs, all corresponding to CGIs, was in-house curated by the Feil laboratory and is largely based on previous reports by others[52-54] together with Feil laboratory's own compilation of imprinted DMRs. Mm9 was used as a reference genome. The list of these annotations with genomic coordinates is provided in Supplementary Table 1.

G4 motif and enrichment analysis.

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

G4Hunter scores of all experiments were tested against fragments of the same size distribution randomly selected in the genomes (10 permutations), thus allowing the evaluation of the observed versus expected from random selection scores (Fig. 1f).
Finally, G4 motifs were analyzed using a published code to assign G4 classes[8]: Loop size 1–3, 4–5 and 6–7; sequences with at least one loop of the respective length; long

loop: sequences with a G4 with any loop of length >7 (up to 12 for any loop and 21 for the middle loop); 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 (Fig. 1h, Extended Data Fig. 3a, Extended Data Fig. 10d). In this analysis random sequences with the same size distribution were used as control.

Computation of G4Hunter scores

G4Hunter scores were computed using a previously published principle[55] with specific functions (see provided R scripts). First, G4Hunter hits were extracted from the reference genomes (hg19, mm9, dm6 and sc3; with a window of 25 and a minimal score of 0.5). Each hit is characterized by its genomic coordinates and a "max_score" reflecting the score of the highest G4Hunter within this window. These scores reflect the relative G4 propensity of the peaks. Peaks with no overlapping G4Hunter hit have a score of 0. Random genomic regions of same size distribution were used as control.

To assess GC richness effects in G4Hunter score distributions, all peaks were resized at 90bp around their summits. Shuffled DNA sequences using the same pool of nucleotides and the same size distribution were generated and compared to G4access or G4-ChIP datasets using G4Hunter analysis at 25 bp window settings and selecting the best scores in the fragments[6] (Extended Data Fig. 2e).

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

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

Gene expression and gene ontology analyses

Genes exhibiting G4access peaks in their promoters (within -200bp and their TSS) were determined using Bedtools[56]. Fragments Per Kilobase per Million (FPKM) of all genes were determined using Cufflinks[57] (version 2.2.1) (Extended Data Fig. 3g). Charts were drawn from all expressed genes defined as genes with FPKM>0. (RNA-seq dataset used are indicated in Supplementary Table 4). Gene ontology was performed using DAVID webtool[58] (https://david.ncifcrf.gov/) on the top 3000 promoters ranked by G4access signals (Extended Data Fig.3f).

Methylation analysis

WGBS datasets were analyzed using Bismarck (version 0.22.3)[59] for Fig. 6a-b (K562 cells) and the computed beta files provided in ref[60] (GSE186458) for Fig. 6c-d (human granulocytes) . A Selection on G4H1.2 exhibiting at least 2CpG within the annotation and overlapping with G4access was performed (Fig. 6). A subset of 100000 randomly selected G4H1.2 which do not overlap with G4access peaks was used as control (Fig. 6b).

Average binding profiles and heatmaps

To generate average binding or G4Hunter profiles (Fig. 1a, 4a, 4d, Extended Data Fig. 1a, 7a-b, 8, 9c), R scripts were developed and used for retrieving bin scores in defined regions from 10 bp bin sized wiggle files[45]. Heatmaps were generated, viewed and color-scaled according to sample read depth using Java TreeView[61] (version 1.2.0-osx) (Fig. 4c, 6a,c). Regions were defined as centered on the G4access peak summits (from the narrow peak table of MACS2 at *p*-value < 10^{-10}). For Fig.4a and 4c, and because we applied a filter of G4access peaks with weak/moderate G4 predictions, G4access peaks (from the narrow peak table of MACS2, *p*-value < 10^{-8}) overlapping G4 predictions G4H1.2 or QP longloops were considered (n=10018). The longloop predictions were generated using the Quad-Parser consensus[14] QP3-7, allowing one loop of a maximum length of 30 nt. To generate average binding profiles of Pol II (Fig. 4d left), hg19 Ensembl gene annotations were used to extract values from wiggle files associated with the selected genes. Bin scores inside these annotations and in a region of 5kb before the TSSs and after 5kb of annotated termination sites were determined. Based on the gene list selections, bin scores from wiggle files were used to re-scale values between TSSs and transcription termination sites (gene body) of all genes using linear interpolation. In total, 1000 points were interpolated for the gene body of each selected gene in all average profiles presented.

Metaprofiles of G4access at gene bodies were performed as above at sites that do not overlap with H3K27ac ChIP-seq signals to avoid enhancers and unannotated promoters (3d right). Finally, Deeptools[62] (version 3.3.0) was used to generate metaprofiles in Fig. 6.

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

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

Analysis of genomic locations of predicted G4s or G4access peaks

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

Analysis of signal variation between replicates

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

Statistics and reproducibility

Sample size

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

Data exclusion

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

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

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

Replication

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

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

Randomization

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

Blinding

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

Data availability

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

Code availability

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

Methods-only references

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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. Tm and ∆Tm 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 37ºC (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, -log10 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 µM 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 significatively 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-blots 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.

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

Figure 1

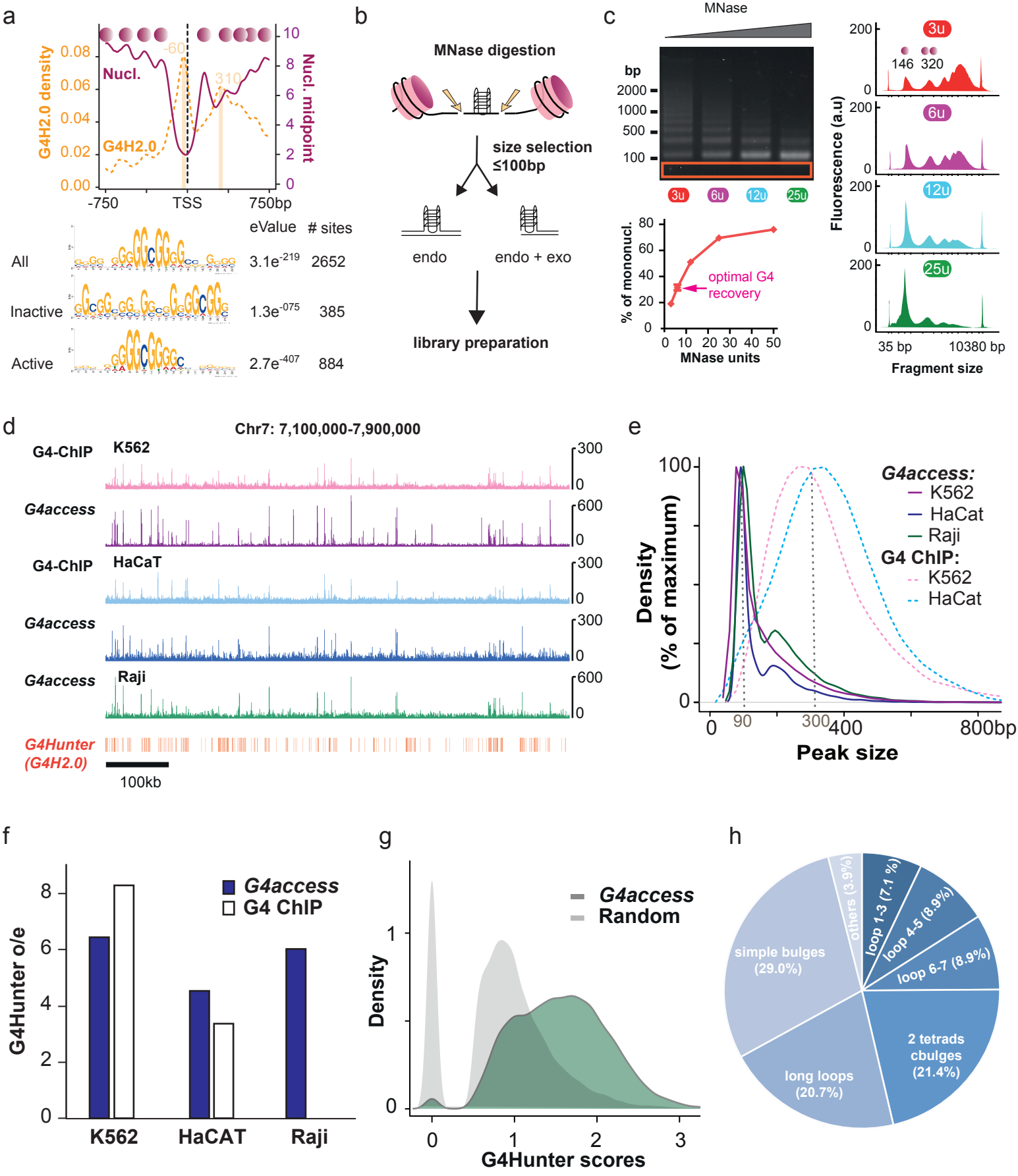

a
G4H2.0 density
Nucl.
G4H2.0
-60
310
Nucl. midpoint
-750
TSS
750bp
eValue
sites
All
3.1e-219
2652
Inactive
1.3e-075
385
Active
2.7e-407
884
b
MNase digestion
size selection
≤100bp
endo
endo + exo
library preparation
c
MNase
bp
2000
1000
500
100
3u
6u
12u
25u
% of mononucl.
optimal G4
recovery
MNase units
Fluorescence (a.u)
146
320
35 bp
10380 bp
Fragment size
d
Chr7: 7,100,000-7,900,000
G4-ChIP
K562
G4access
HaCaT
Raji
G4Hunter
(G4H2.0)
100kb
e
Density
(% of maximum)
G4access:
K562
HaCat
Raji
G4 ChIP:
K562
HaCat
Peak size
f
G4Hunter o/e
G4access
G4 ChIP
K562
HaCAT
Raji
g
G4access
Random
Density
G4Hunter scores
h
simple bulges
(29.0%)
others (3.9%)
loop 1-3 (7.1 %)
loop 4-5 (8.9%)
loop 6-7 (8.9%)
2 tetrads
cbulges
(21.4%)
long loops
(20.7%)

Figure 2

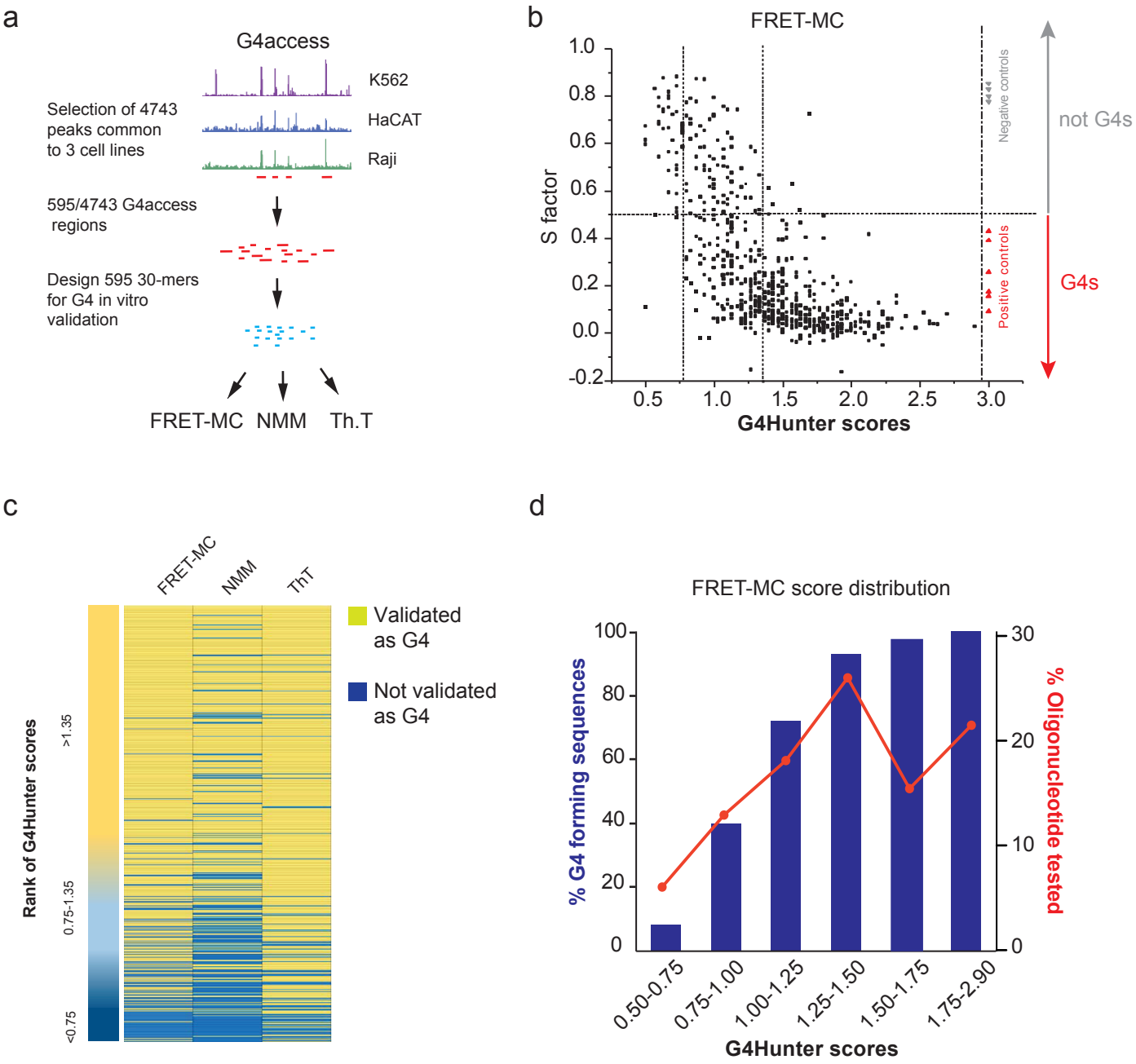

a
G4access
K562
HaCAT
Raji
Selection of 4743 peaks common to 3 cell lines
595/4743 G4access regions
Design 595 30-mers for G4 in vitro validation
FRET-MC
NMM
Th.T
b
FRET-MC
S factor
G4Hunter scores
Negative controls
Positive controls
not G4s
G4s
c
FRET-MC
NMM
ThT
Rank of G4Hunter scores
>1.35
0.75-1.35
<0.75
Validated as G4
Not validated as G4
d
FRET-MC score distribution
% G4 forming sequences
% Oligonucleotide tested
0.50-0.75
0.75-1.00
1.00-1.25
1.25-1.50
1.50-1.75
1.75-2.90
G4Hunter scores

Figure 3

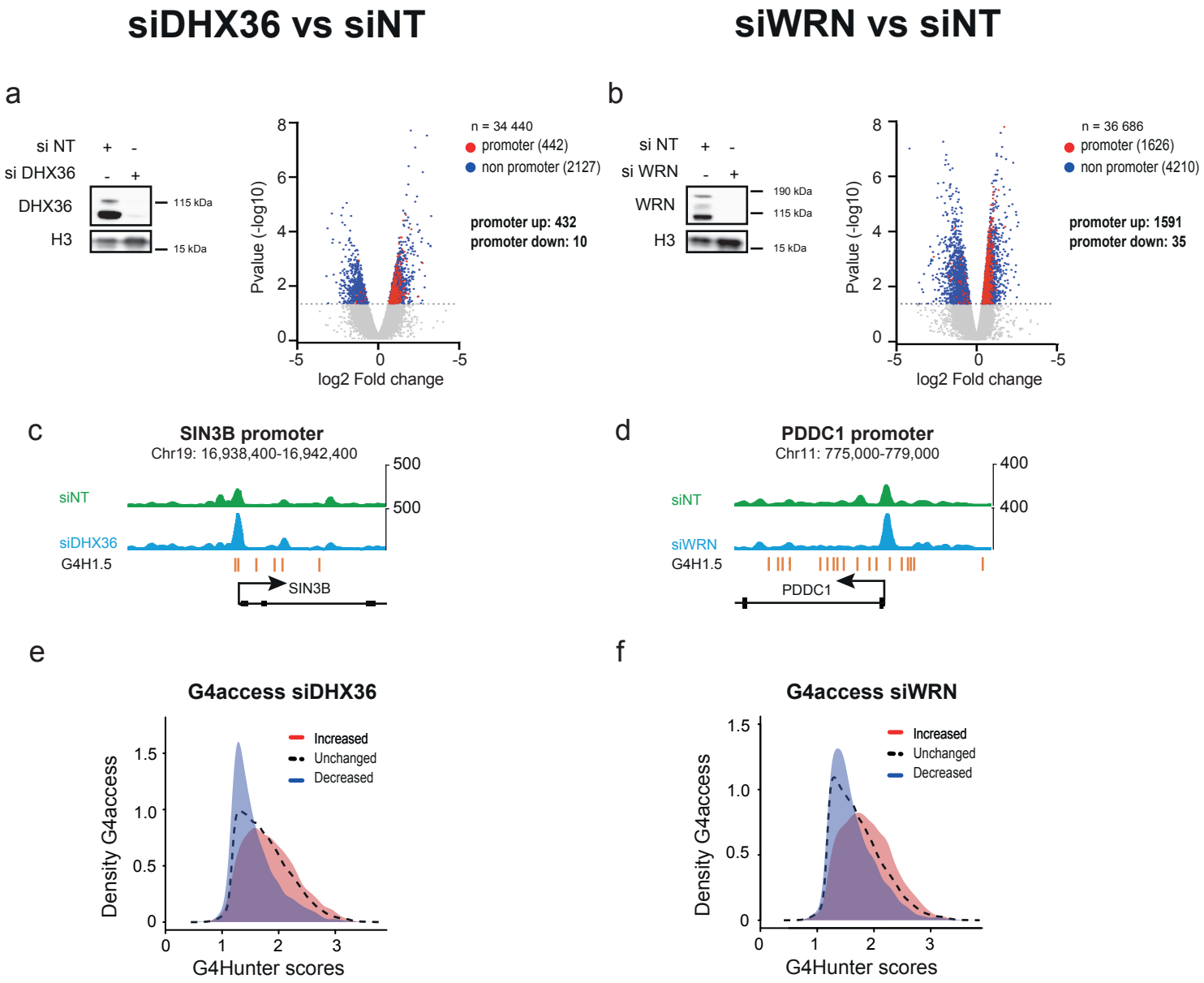

siDHX36 vs siNT
siWRN vs siNT
a
si NT + -
si DHX36 - +
DHX36
H3
115 kDa
15 kDa
Pvalue (-log10)
log2 Fold change
n = 34 440
promoter (442)
non promoter (2127)
promoter up: 432
promoter down: 10
b
si NT + -
si WRN - +
WRN
H3
190 kDa
115 kDa
15 kDa
n = 36 686
promoter (1626)
non promoter (4210)
promoter up: 1591
promoter down: 35
c
SIN3B promoter
Chr19: 16,938,400-16,942,400
siNT
siDHX36
G4H1.5
SIN3B
d
PDDC1 promoter
Chr11: 775,000-779,000
siNT
siWRN
G4H1.5
PDDC1
e
G4access siDHX36
Increased
Unchanged
Decreased
Density G4access
G4Hunter scores
f
G4access siWRN
Increased
Unchanged
Decreased
Density G4access
G4Hunter scores

Figure 4

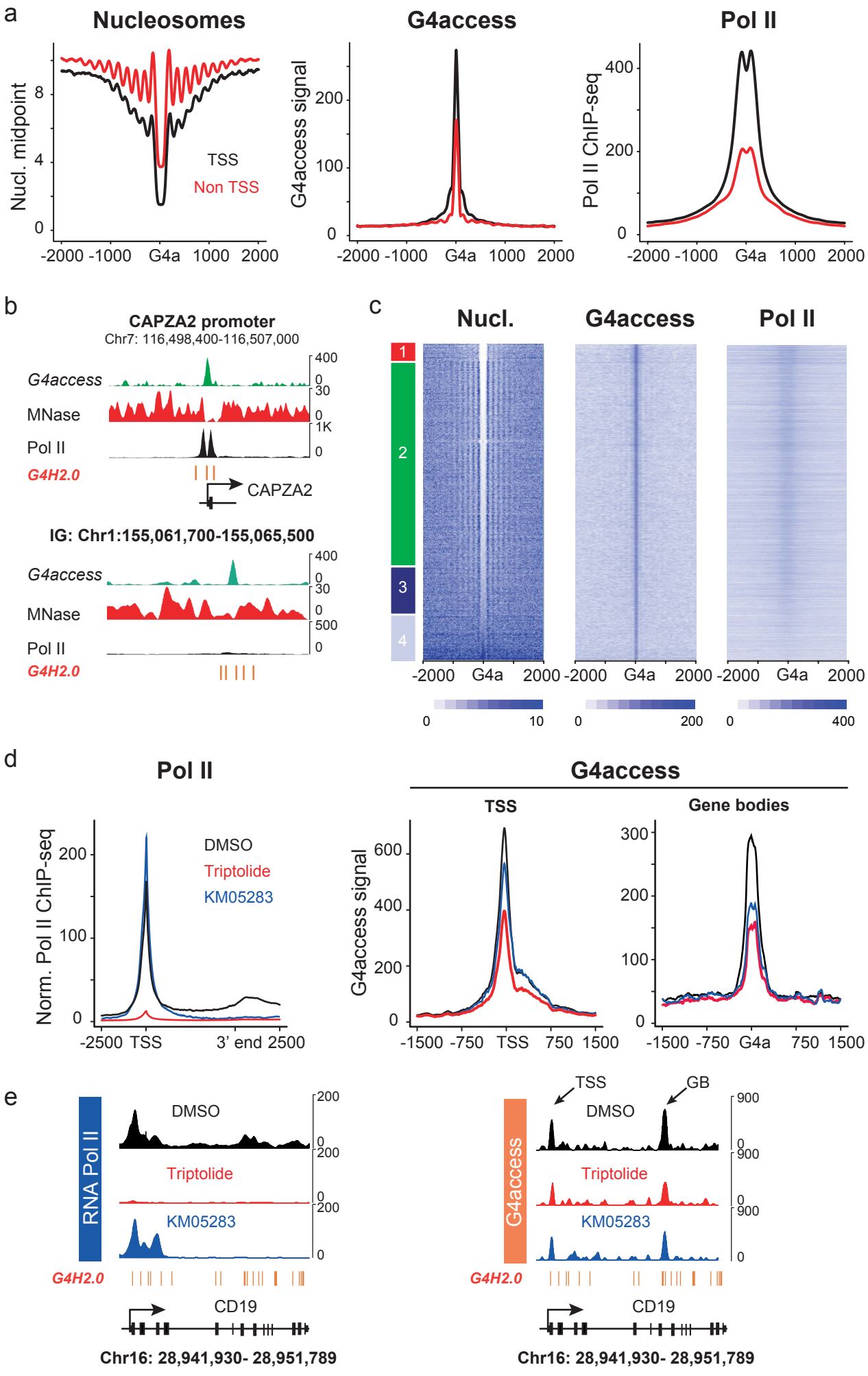

a
Nucleosomes
G4access
Pol II
Nucl. midpoint
G4access signal
Pol II ChIP-seq
TSS
Non TSS
-2000 -1000 G4a 1000 2000
b
CAPZA2 promoter
Chr7: 116,498,400-116,507,000
G4access
MNase
Pol II
G4H2.0
CAPZA2
IG: Chr1:155,061,700-155,065,500
c
Nucl.
G4access
Pol II
d
Pol II
G4access
TSS
Gene bodies
Norm. Pol II ChIP-seq
G4access signal
DMSO
Triptolide
KM05283
-2500 TSS 3' end 2500
-1500 -750 TSS 750 1500
-1500 -750 G4a 750 1500
e
RNA Pol II
G4access
DMSO
Triptolide
KM05283
G4H2.0
CD19
Chr16: 28,941,930- 28,951,789

Figure 5

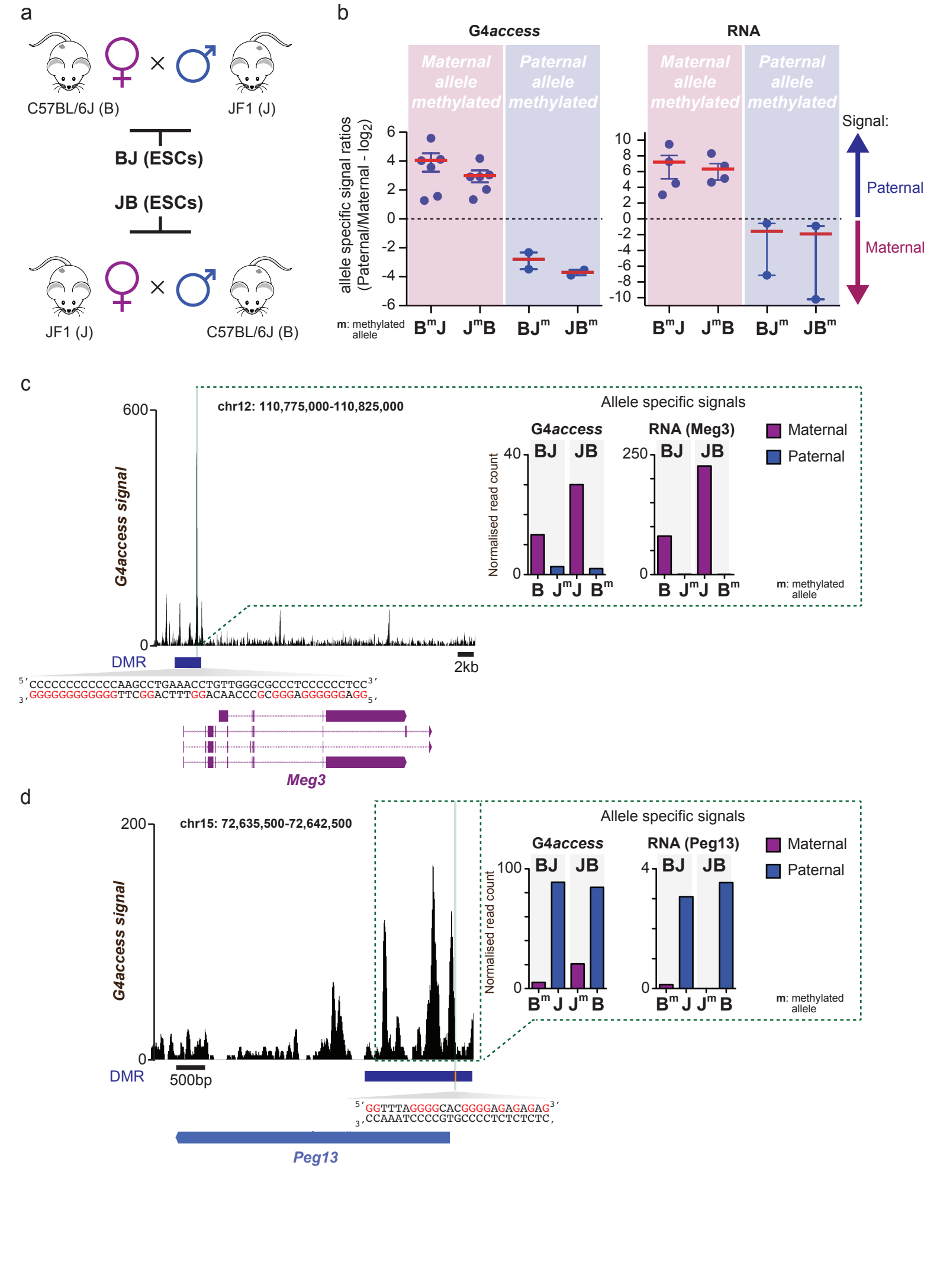
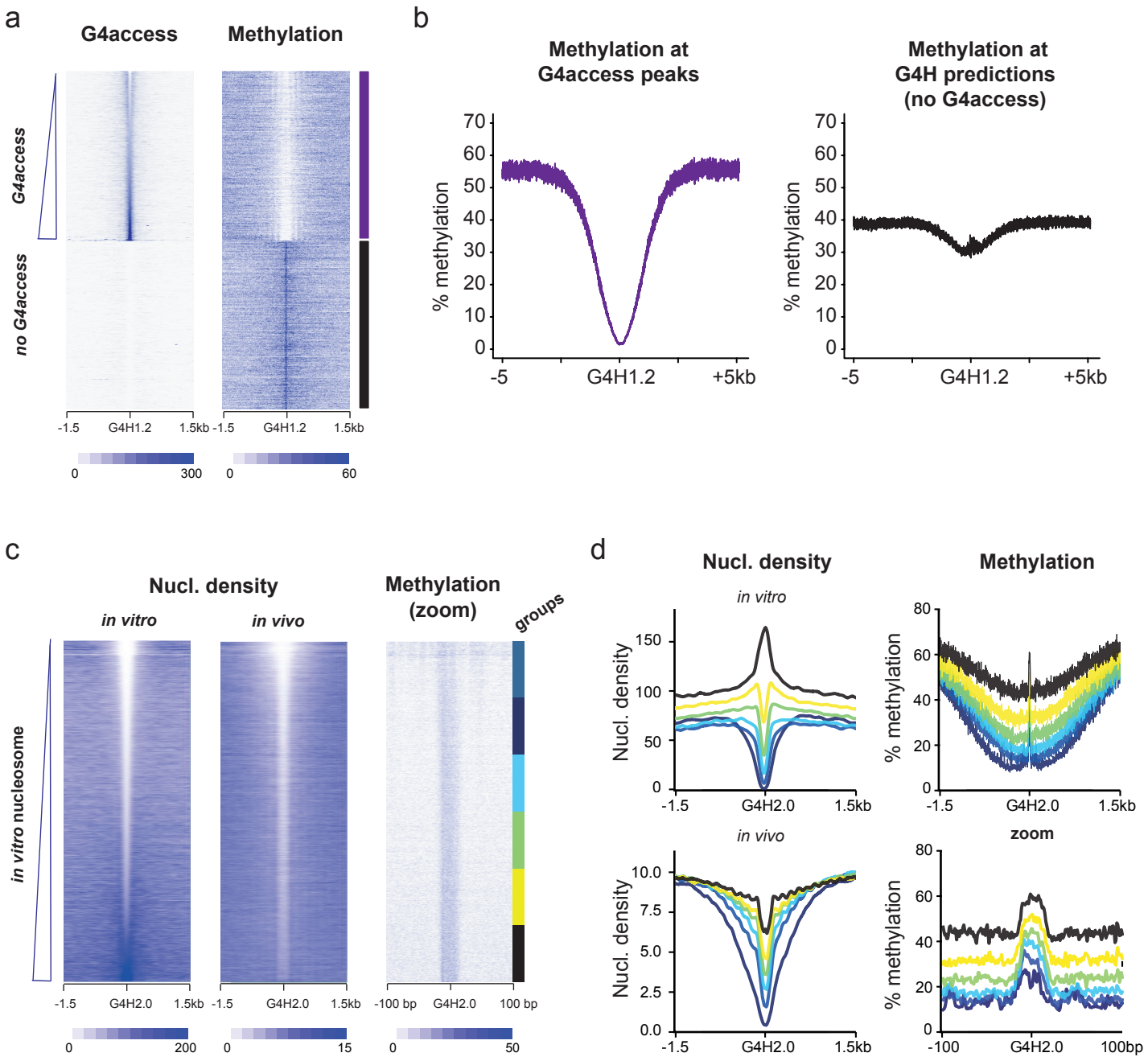
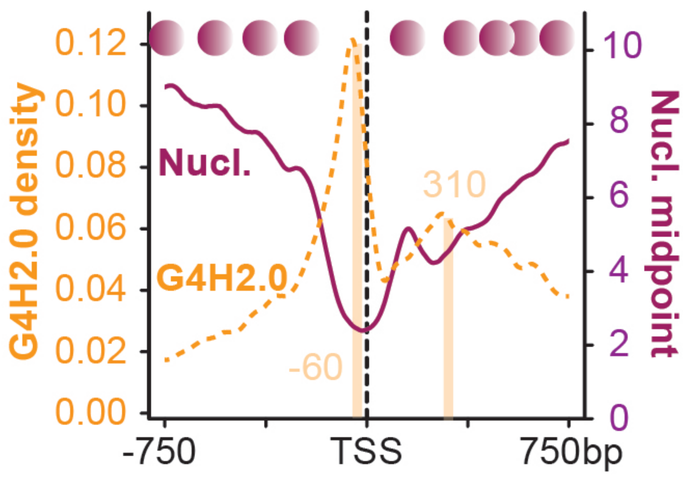

a
C57BL/6J (B)
JF1 (J)
BJ (ESCs)
JB (ESCs)
JF1 (J)
C57BL/6J (B)
b
G4access
RNA
Maternal allele methylated
Paternal allele methylated
allele specific signal ratios (Paternal/Maternal - log2)
m: methylated allele
BmJ
JmB
BJm
JBm
Signal:
Paternal
Maternal
c
chr12: 110,775,000-110,825,000
G4access signal
DMR
2kb
Allele specific signals
G4access
RNA (Meg3)
Normalised read count
Maternal
Paternal
m: methylated allele
Meg3
d
chr15: 72,635,500-72,642,500
G4access signal
DMR
500bp
Allele specific signals
G4access
RNA (Peg13)
Normalised read count
Maternal
Paternal
m: methylated allele
Peg13

Figure 6

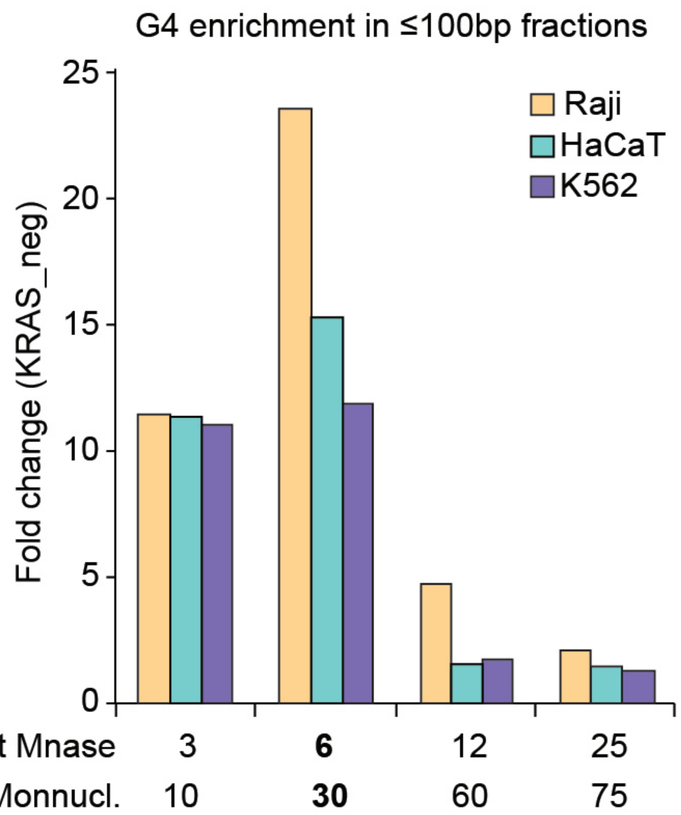

a
G4access
Methylation
G4access
no G4access
-1.5
G4H1.2
1.5kb
0
300
60
b
Methylation at G4access peaks
Methylation at G4H predictions (no G4access)
% methylation
-5
G4H1.2
+5kb
c
Nucl. density
Methylation (zoom)
groups
in vitro
in vivo
in vitro nucleosome
G4H2.0
200
15
50
-100 bp
100 bp
d
Nucl. density
Methylation
in vitro
in vivo
zoom
-100
100bp

ED Figure 1

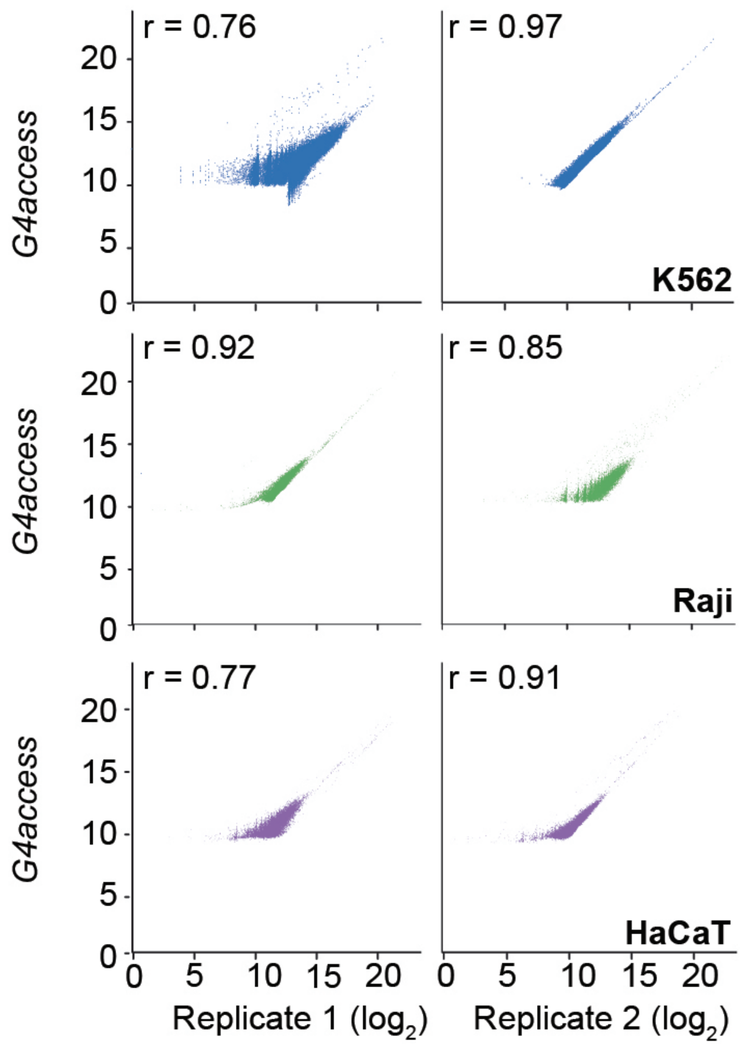
a



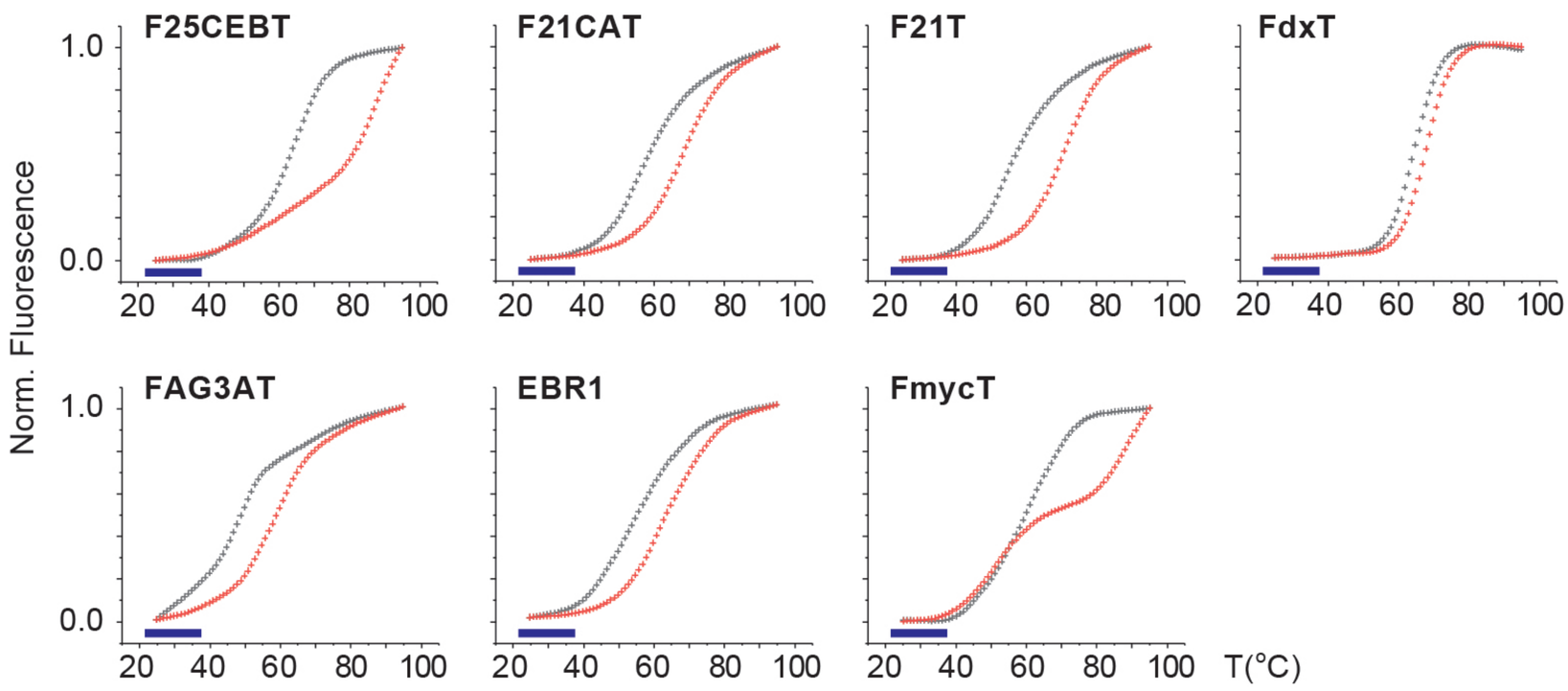
b



e



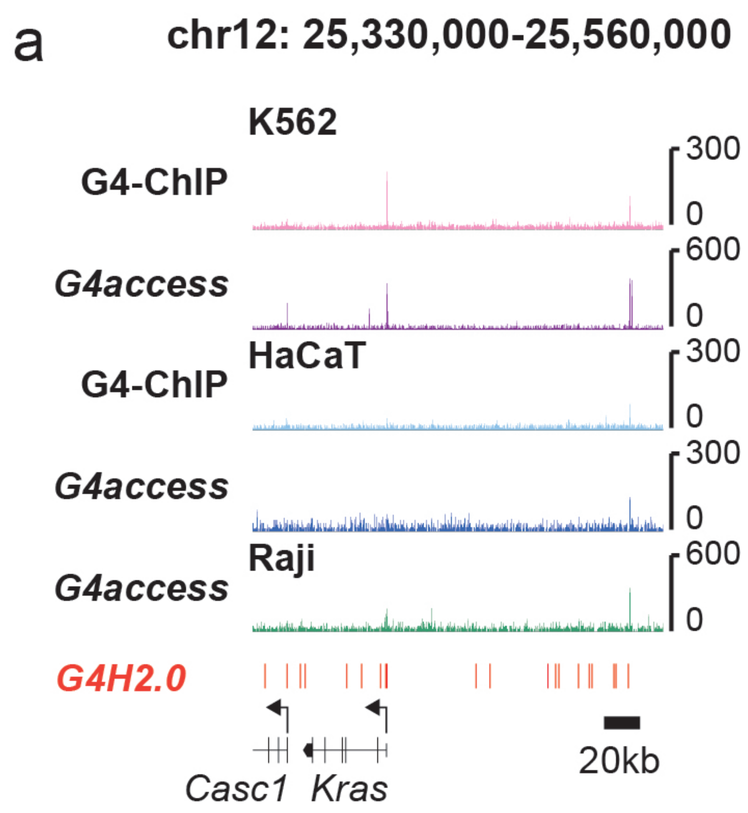
C

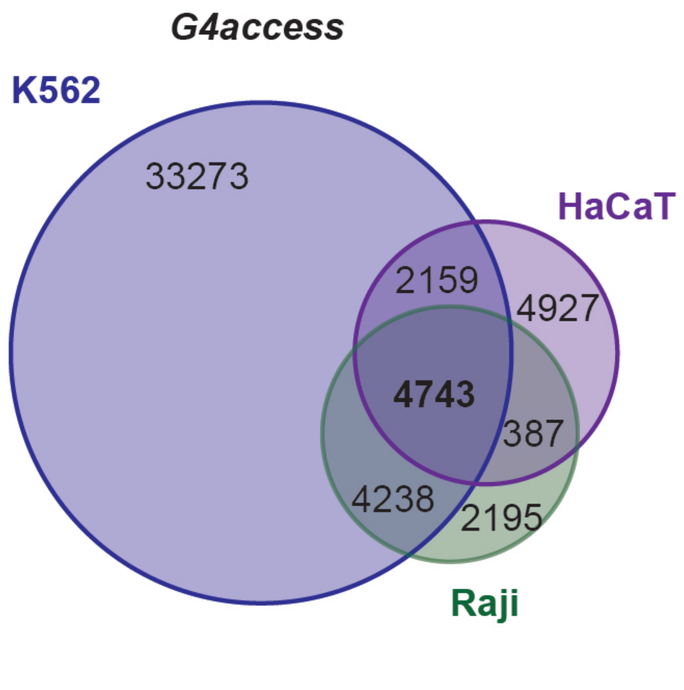


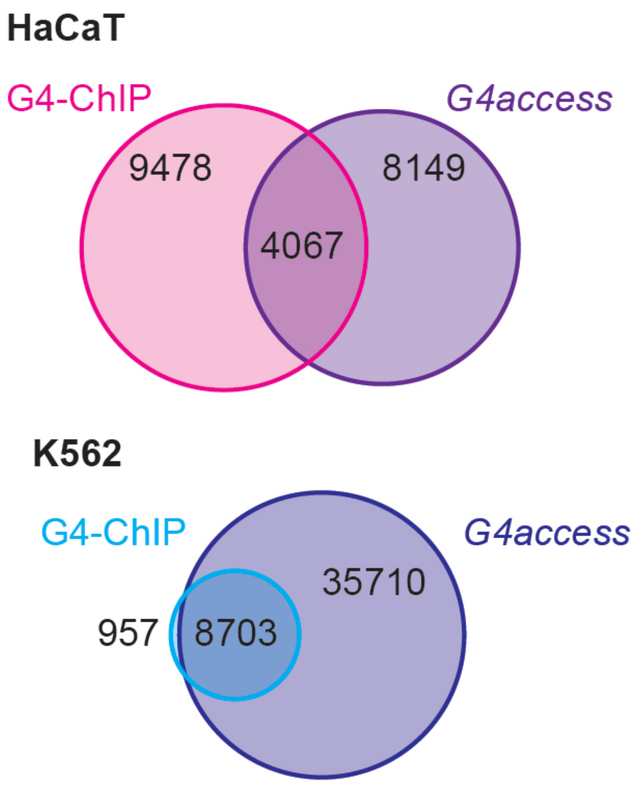
d

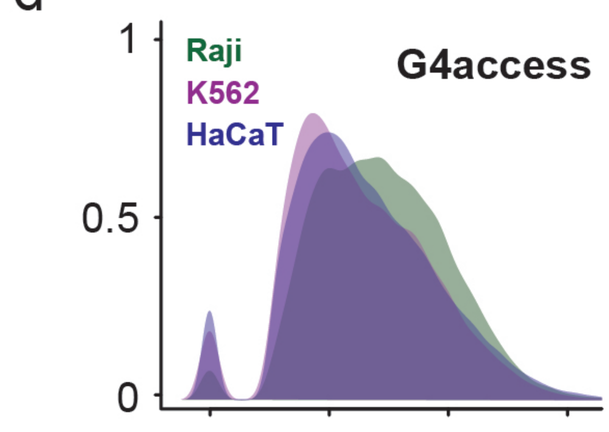
Name	Sequence	G4H score	Tm1 140mM KCl	Tm2 45mM NaCl 10mM KCl	Δ Tm (°C)
F25CEBT	AGGGTGGGTGTAAGTGTGGGTGGGT	1.56	79.6	63.4	16.2
F21CTAT	GGGCTAGGGCTAGGGCTAGGG	1.57	67.8	59.6	8.2
F21T	GGGTTAGGGTTAGGGTTAGGG	1.71	69.5	57	12.5
FAG3AT	AGGGAAGGGAAGGGAAGGGA	1.8	57.6	51.2	6.4
EBR1	GGGCAGGGGGTGATGGGGAGGAGCCAGGG	1.9	62.8	57	5.8
FmycT	TTGAGGGTGGGTAGGGTGGGTAA	1.61	/	59.5	/
FdxT	TATAGCTAT-PEG-TATAGCTATA	n. a.	68	65	3

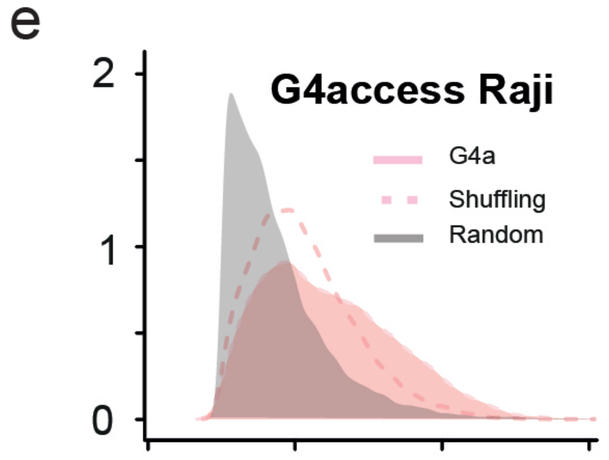
ED Figure 2

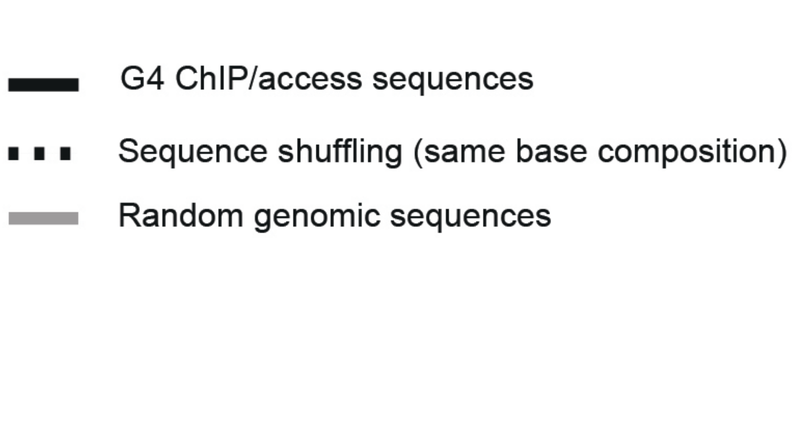

a
chr12: 25,330,000-25,560,000
K562
G4-ChIP
300
0
G4access
600
0
HaCaT
G4-ChIP
300
0
G4access
300
0
Raji
G4access
600
0
G4H2.0
Casc1
Kras
20kb


b
G4access
K562
33273
HaCaT
2159
4927
4743
387
4238
2195
Raji


c
HaCaT
G4-ChIP
G4access
9478
8149
4067
K562
G4-ChIP
G4access
957
8703
35710


d
Raji
K562
HaCaT
G4access
1
0.5
0

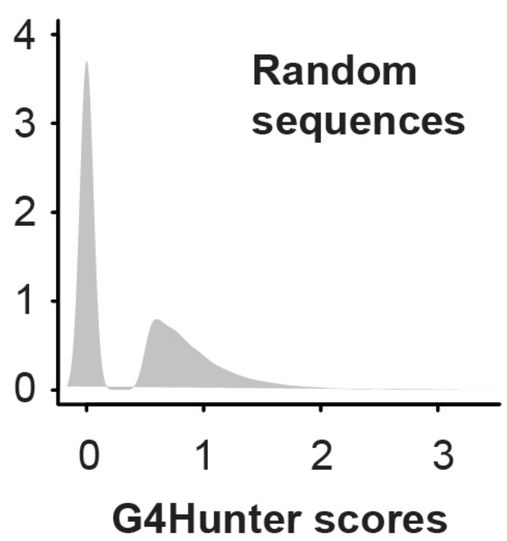

e
G4access Raji
G4a
Shuffling
Random
2
1
0

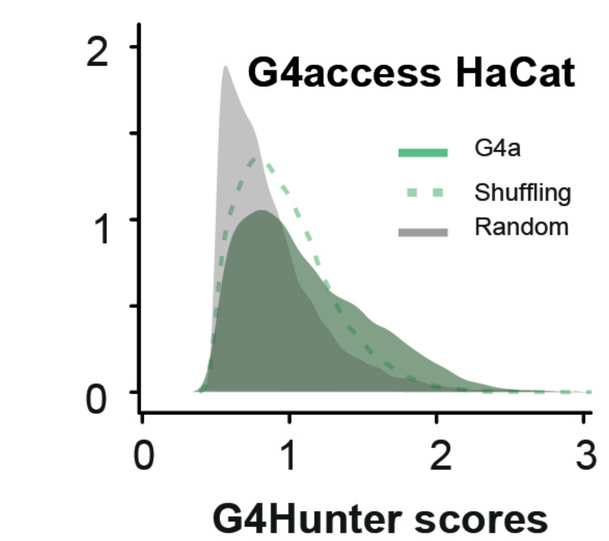

G4 ChIP/access sequences
Sequence shuffling (same base composition)
Random genomic sequences

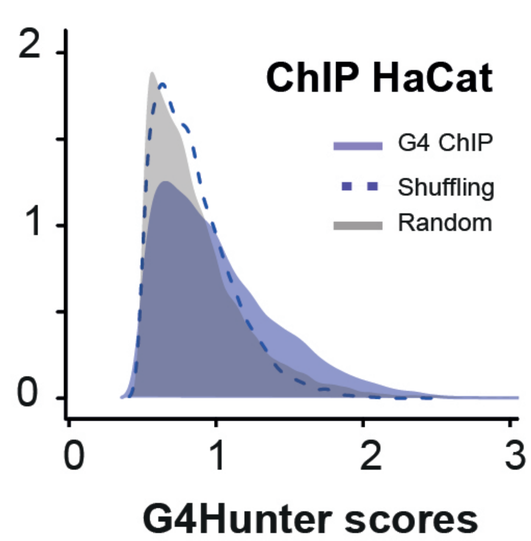
K562
HaCaT
G4 ChIP
Density
1
0.5
0

G4access K562
G4a
Shuffling
Random
Density
2
1
0

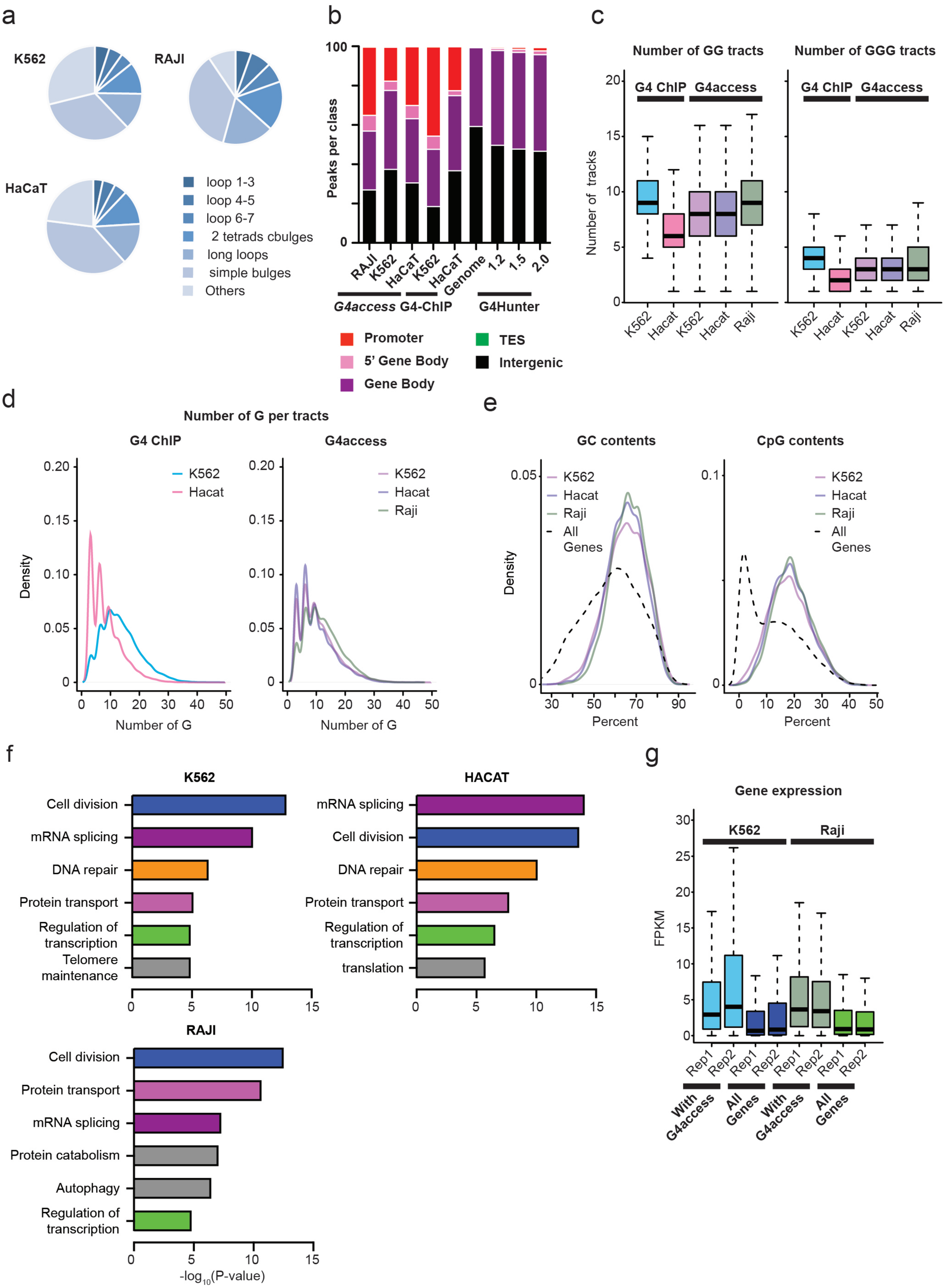
ChIP K562
G4 ChIP
Shuffling
Random
2
1
0


Random sequences
4
3
2
1
0
0
1
2
3
G4Hunter scores

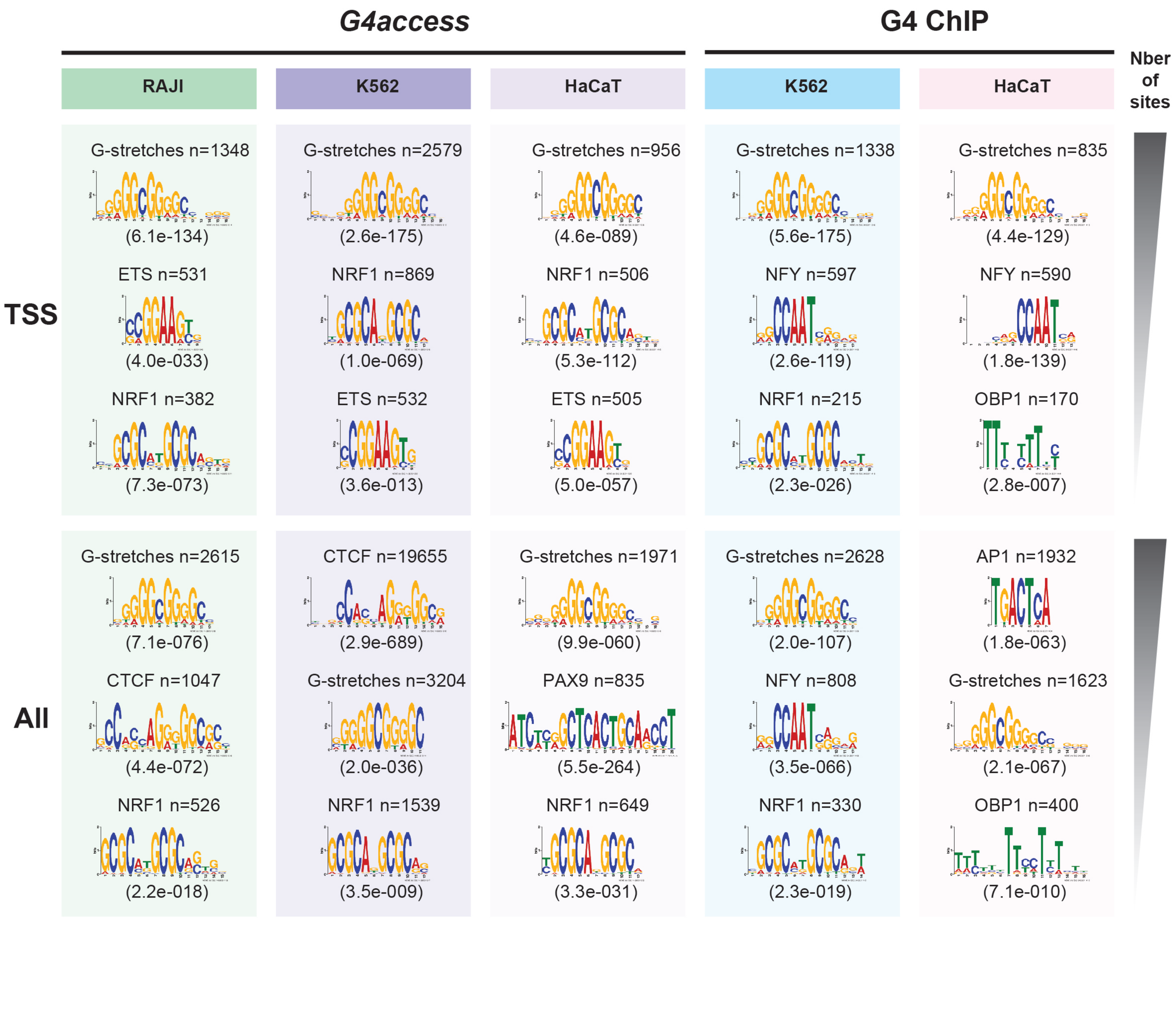

G4access HaCat
G4a
Shuffling
Random
2
1
0
0
1
2
3
G4Hunter scores


ChIP HaCat
G4 ChIP
Shuffling
Random
2
1
0
0
1
2
3
G4Hunter scores

ED Figure 3

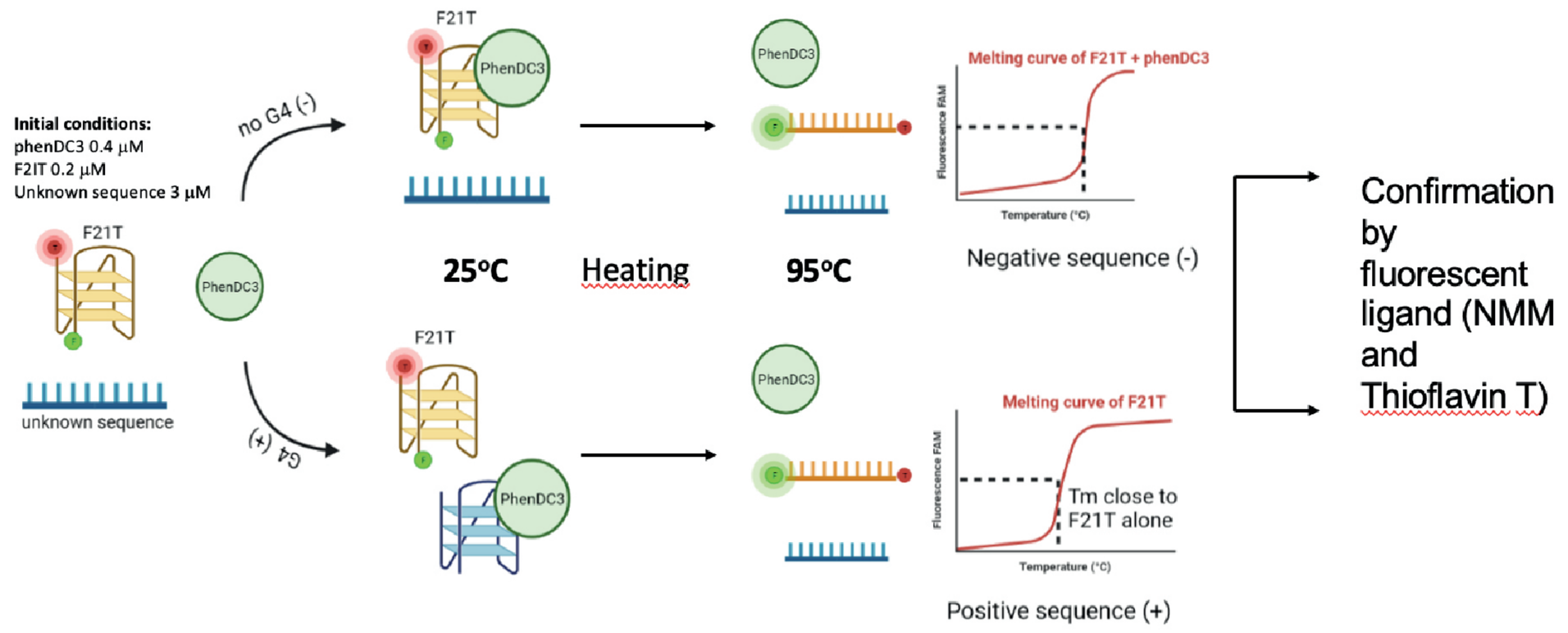

a
K562
RAJI
HaCaT
loop 1-3
loop 4-5
loop 6-7
2 tetrads cbulges
long loops
simple bulges
Others
b
Peaks per class
100
0
RAJI
K562
HaCaT
K562
HaCaT
Genome
1.2
1.5
2.0
G4access
G4-ChIP
G4Hunter
Promoter
5' Gene Body
Gene Body
TES
Intergenic
c
Number of GG tracts
Number of GGG tracts
G4 ChIP
G4access
Number of tracks
20
15
10
5
0
K562
Hacat
K562
Hacat
Raji
d
Number of G per tracts
G4 ChIP
G4access
K562
Hacat
Raji
Density
0.20
0.15
0.10
0.05
0
0 10 20 30 40 50
Number of G
e
GC contents
CpG contents
K562
Hacat
Raji
All Genes
Density
0.05
0.1
30 50 70 90
Percent
f
K562
Cell division
mRNA splicing
DNA repair
Protein transport
Regulation of transcription
Telomere maintenance
HACAT
mRNA splicing
Cell division
DNA repair
Protein transport
Regulation of transcription
translation
RAJI
Cell division
Protein transport
mRNA splicing
Protein catabolism
Autophagy
Regulation of transcription
0 5 10 15
-log10(P-value)
g
Gene expression
K562
Raji
FPKM
30
25
20
15
10
5
0
Rep1
Rep2
With G4access
All Genes
With G4access
All Genes

ED Figure 4

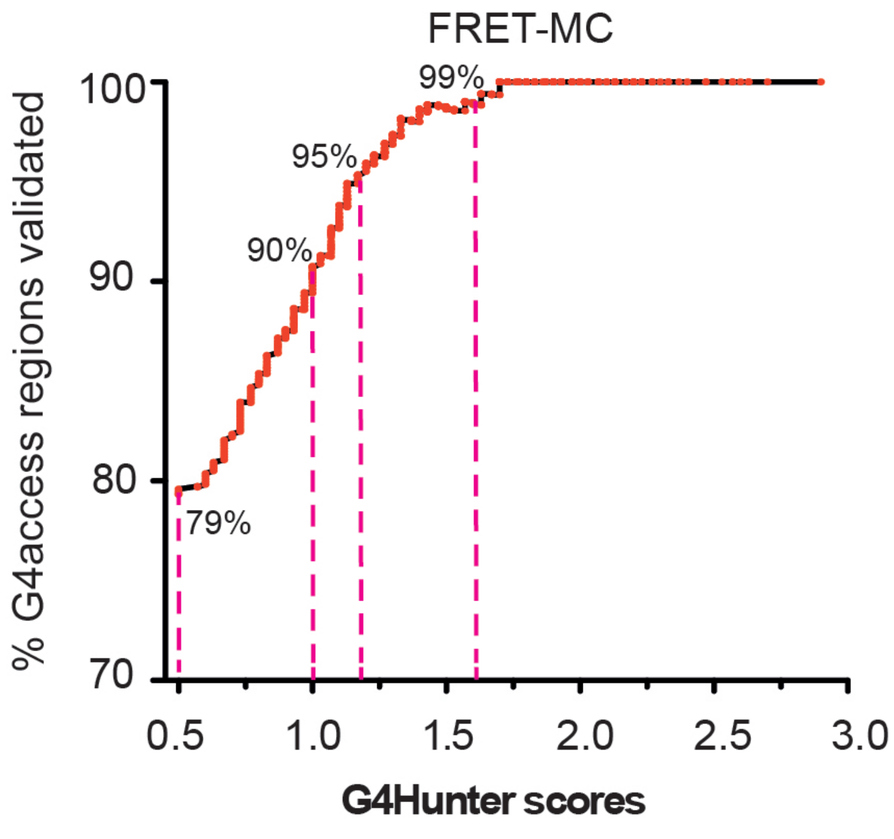

G4access
G4 ChIP
RAJI
K562
HaCaT
K562
HaCaT
Nber of sites
TSS
All
G-stretches n=1348
(6.1e-134)
ETS n=531
(4.0e-033)
NRF1 n=382
(7.3e-073)
G-stretches n=2579
(2.6e-175)
NRF1 n=869
(1.0e-069)
ETS n=532
(3.6e-013)
G-stretches n=956
(4.6e-089)
NRF1 n=506
(5.3e-112)
ETS n=505
(5.0e-057)
G-stretches n=1338
(5.6e-175)
NFY n=597
(2.6e-119)
NRF1 n=215
(2.3e-026)
G-stretches n=835
(4.4e-129)
NFY n=590
(1.8e-139)
OBP1 n=170
(2.8e-007)
G-stretches n=2615
(7.1e-076)
CTCF n=1047
(4.4e-072)
NRF1 n=526
(2.2e-018)
CTCF n=19655
(2.9e-689)
G-stretches n=3204
(2.0e-036)
NRF1 n=1539
(3.5e-009)
G-stretches n=1971
(9.9e-060)
PAX9 n=835
(5.5e-264)
NRF1 n=649
(3.3e-031)
G-stretches n=2628
(2.0e-107)
NFY n=808
(3.5e-066)
NRF1 n=330
(2.3e-019)
AP1 n=1932
(1.8e-063)
G-stretches n=1623
(2.1e-067)
OBP1 n=400
(7.1e-010)

ED Figure 5

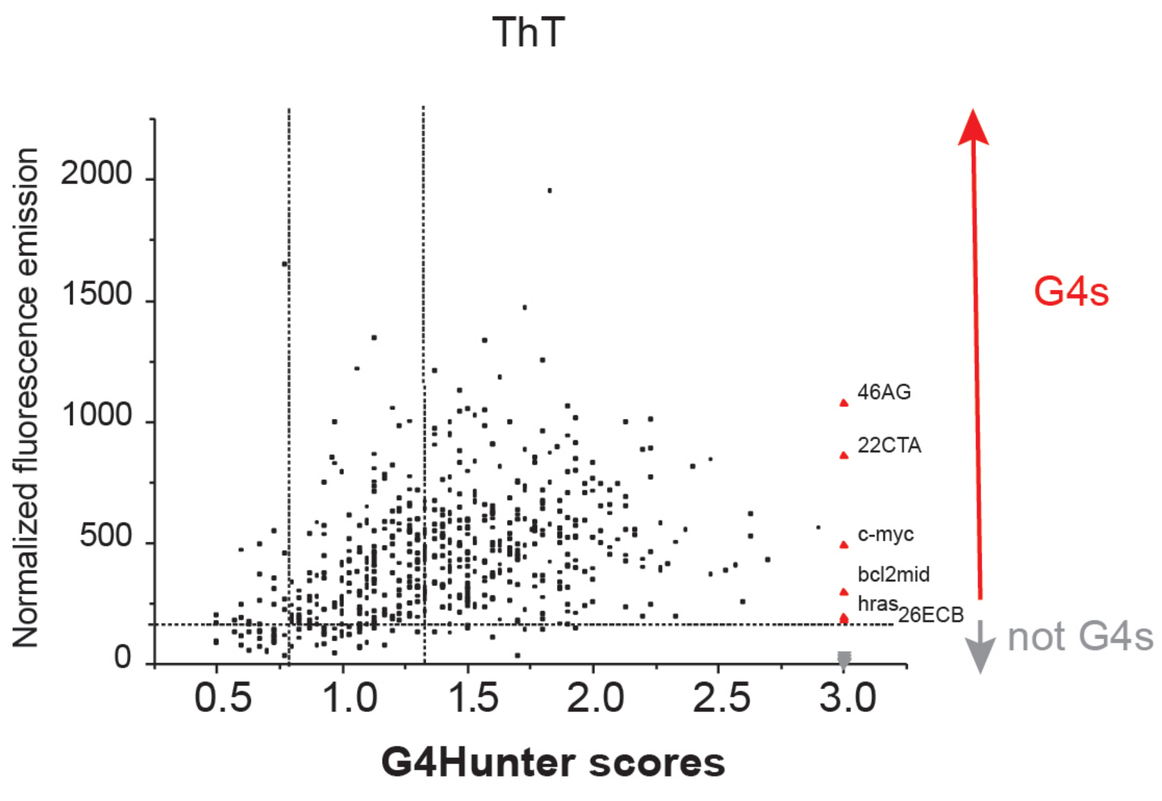
a



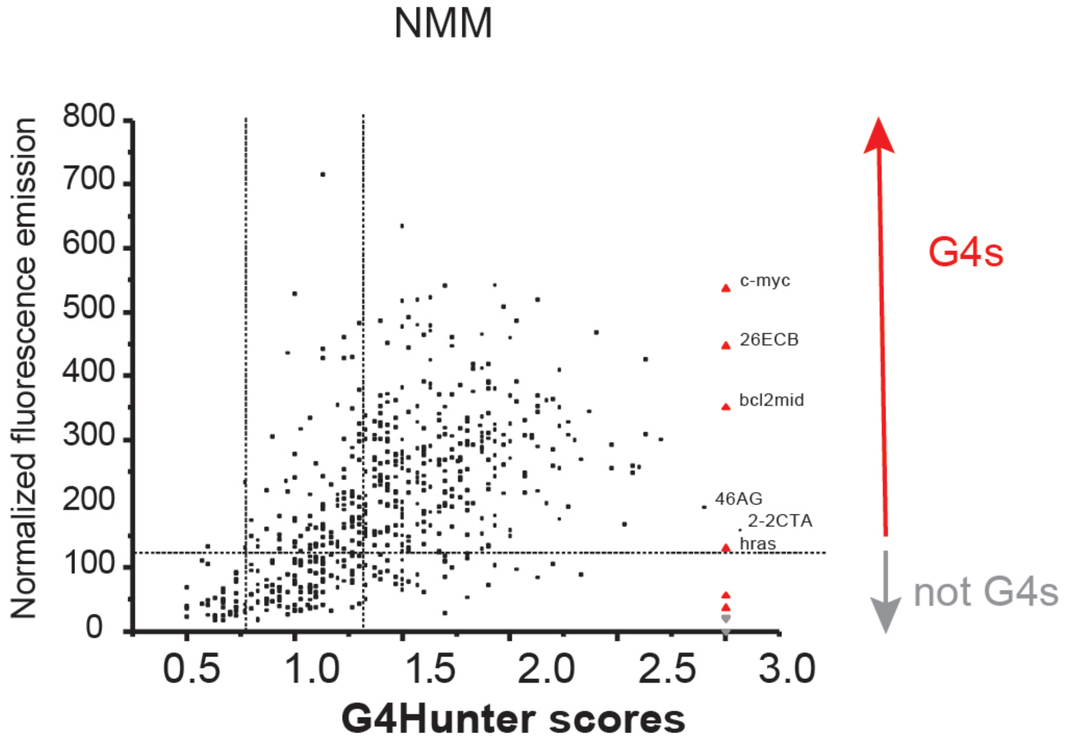
b



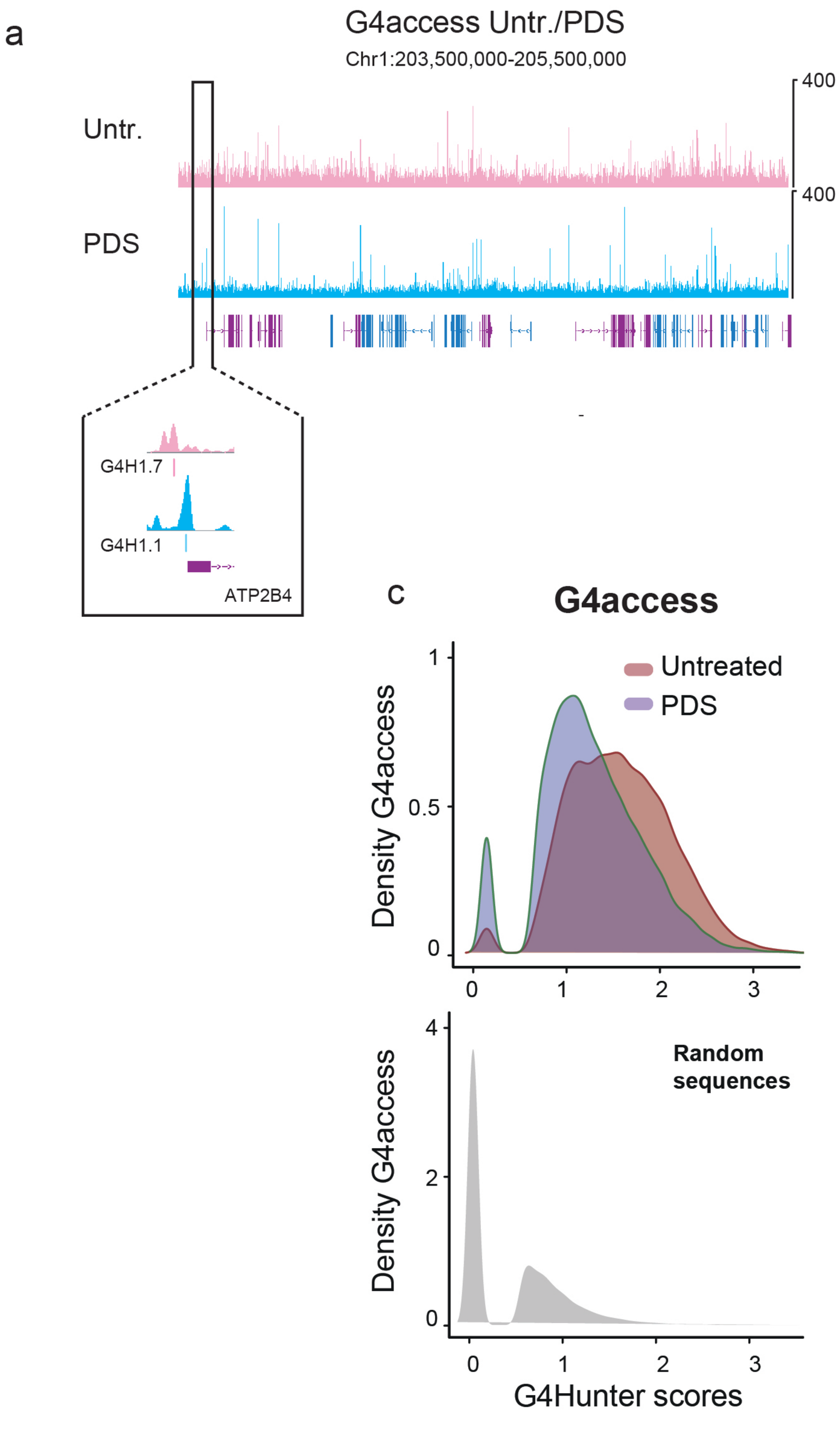
c



d

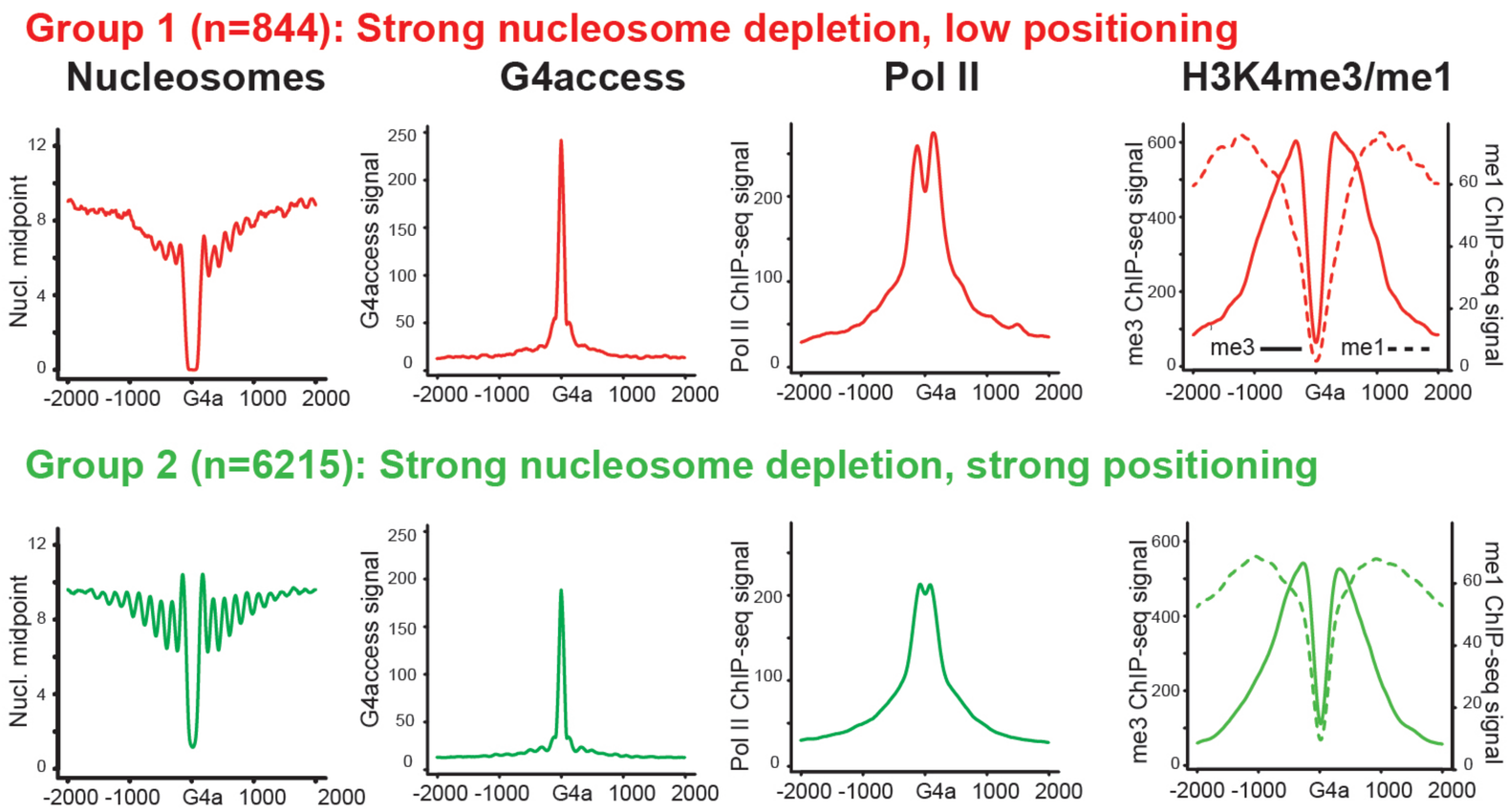


ED Figure 6

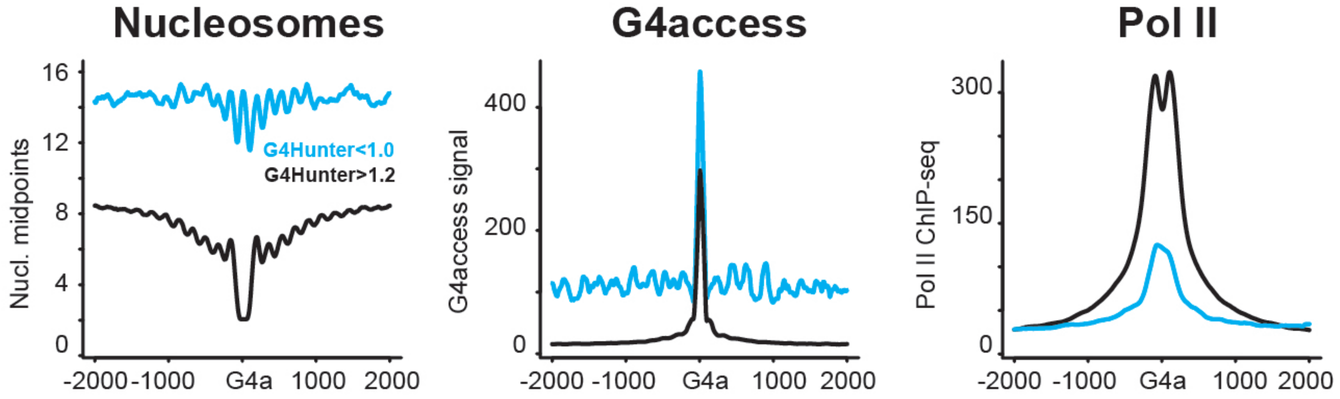

a
G4access Untr./PDS
Chr1:203,500,000-205,500,000
Untr.
PDS
400
400
G4H1.7
G4H1.1
ATP2B4
b
Pvalue (-log10)
promoter
non promoter
Pval 0.05
log2 Fold change
c
G4access
Untreated
PDS
Density G4access
Random sequences
G4Hunter scores
d
G4seq
Density G4-seq
Random sequences
G4Hunter scores

ED Figure 7

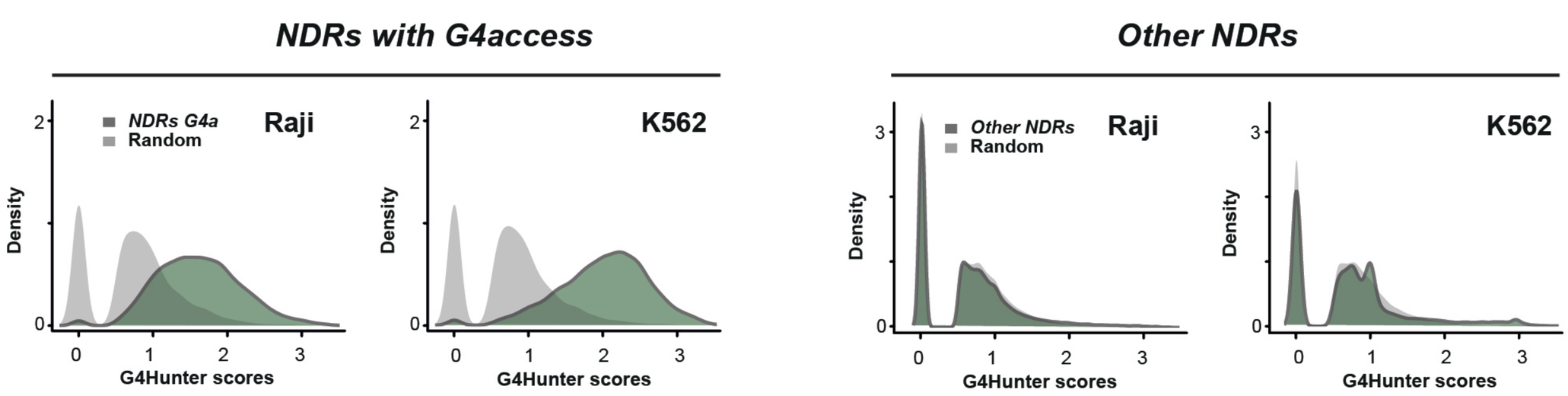
a



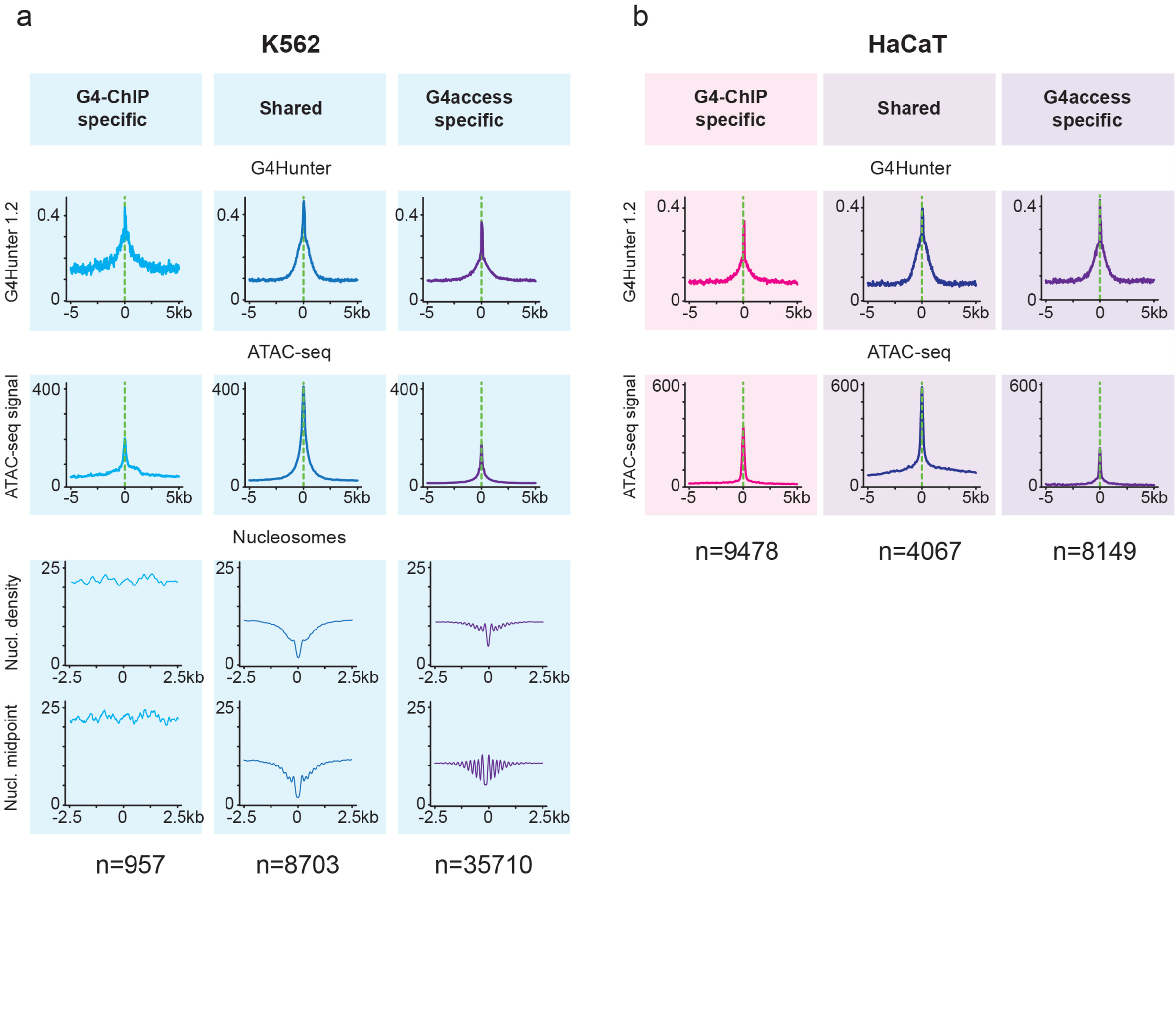
b



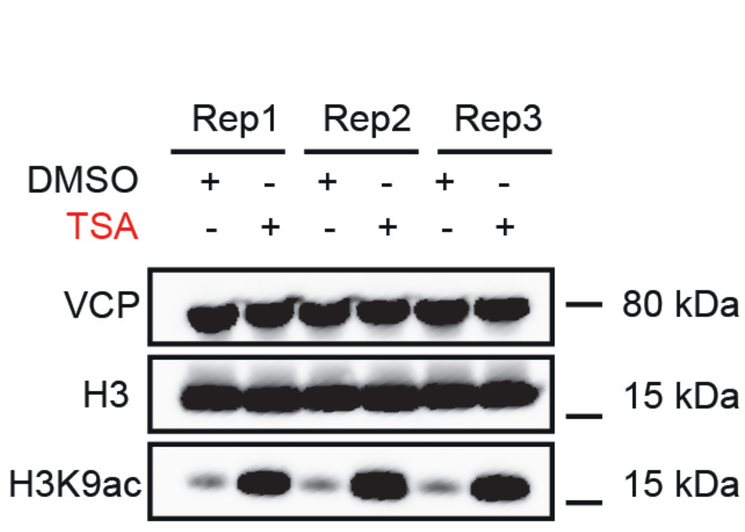
C

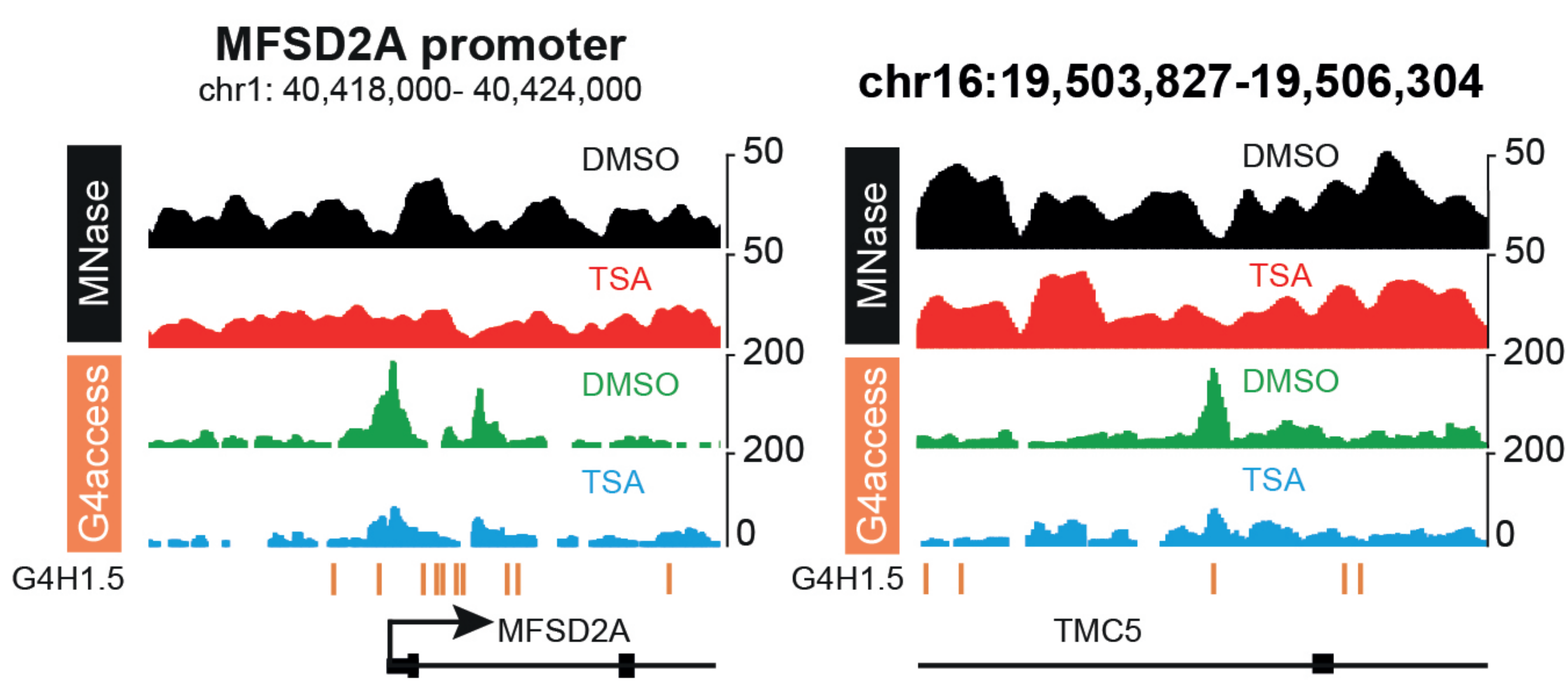


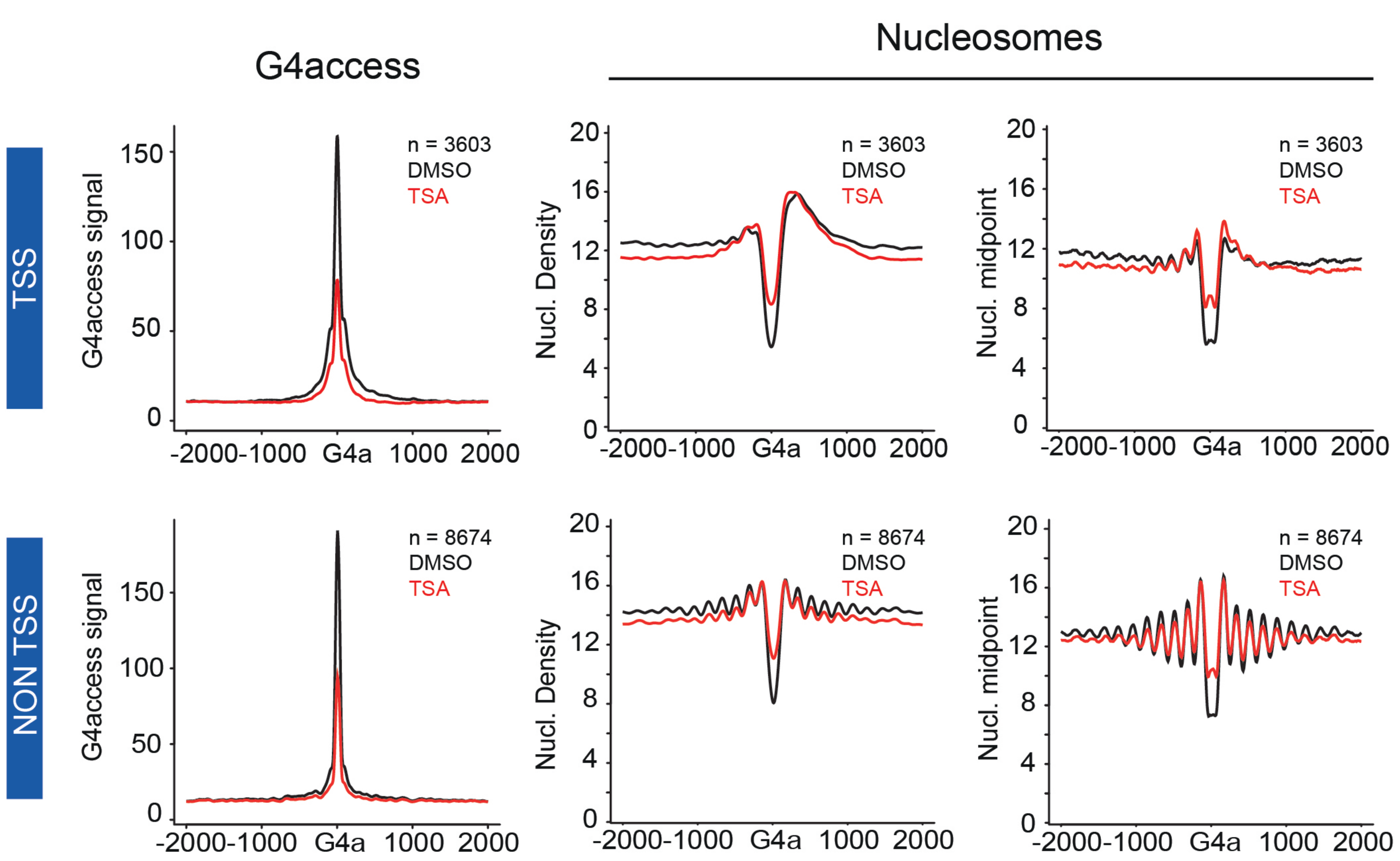
ED Figure 8


a
K562
G4-ChIP specific
Shared
G4access specific
G4Hunter
G4Hunter 1.2
0.4
0
-5
0
5kb
ATAC-seq
ATAC-seq signal
400
0
Nucleosomes
Nucl. density
Nucl. midpoint
25
-2.5
2.5kb
n=957
n=8703
n=35710
b
HaCaT
G4-ChIP specific
Shared
G4access specific
G4Hunter
G4Hunter 1.2
0.4
0
-5
0
5kb
ATAC-seq
ATAC-seq signal
600
n=9478
n=4067
n=8149

ED Figure 9


a
Rep1
Rep2
Rep3
DMSO + - + - + -
TSA - + - + - +
VCP
H3
H3K9ac
80 kDa
15 kDa
15 kDa


b
MFSD2A promoter
chr1: 40,418,000- 40,424,000
chr16:19,503,827-19,506,304
MNase
G4access
DMSO
TSA
50
200
0
G4H1.5
MFSD2A
TMC5


c
G4access
Nucleosomes
TSS
NON TSS
G4access signal
Nucl. Density
Nucl. midpoint
n = 3603
n = 8674
DMSO
TSA
-2000-1000 G4a 1000 2000

ED Figure 10

a

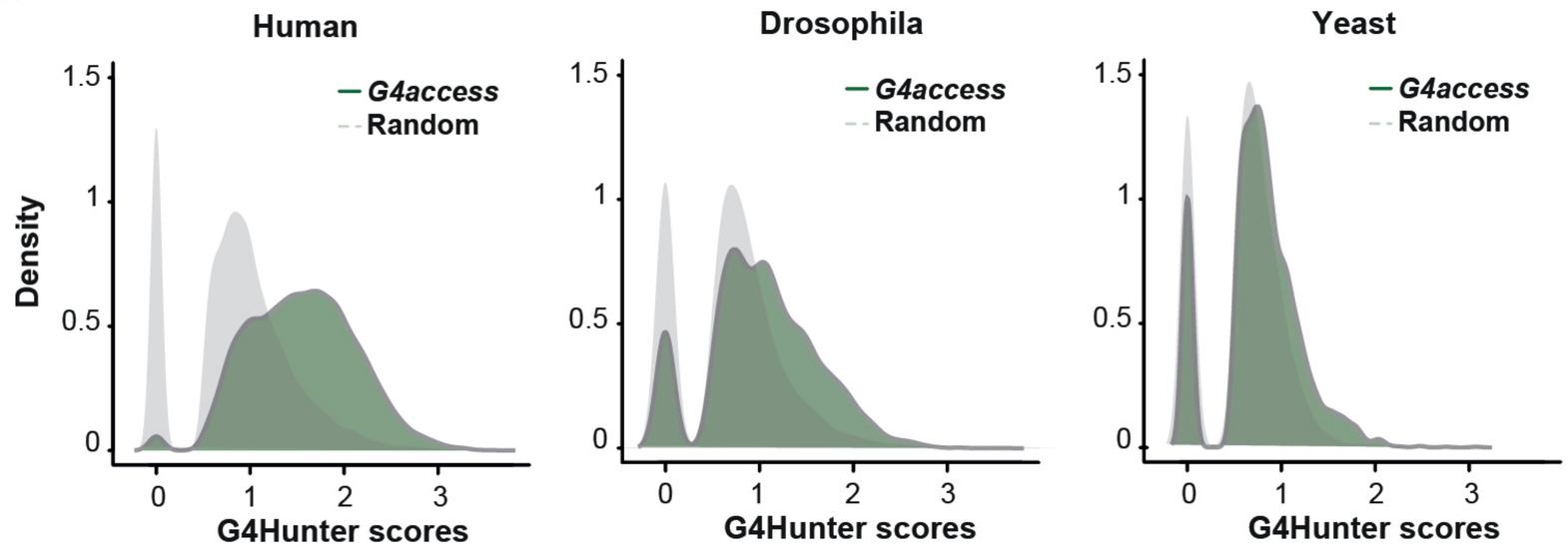
G4H frequency per kb

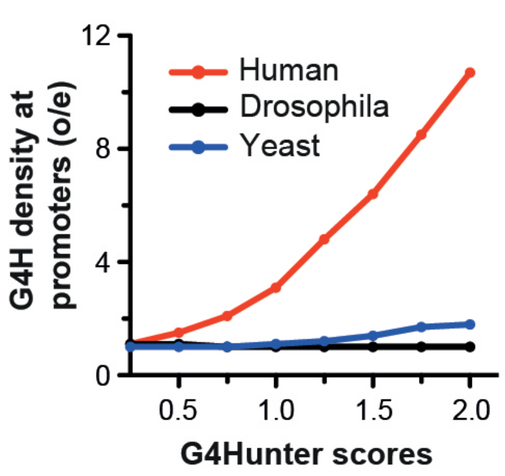
		G4H1.0	G4H1.25	G4H1.5	G4H1.75	G4H2.0
Genome	Human	2.48	1.03	0.51	0.25	0.12
	Drosophila	1.57	0.64	0.32	0.16	0.08
	Yeast	0.71	0.15	0.04	0.01	0.004
Promoters	Human	7.62	4.91	3.28	2.14	1.30
	Drosophila	1.56	0.63	0.33	0.16	0.08
	Yeast	0.78	0.19	0.06	0.02	0.008

G4H density at promoters (o/e)

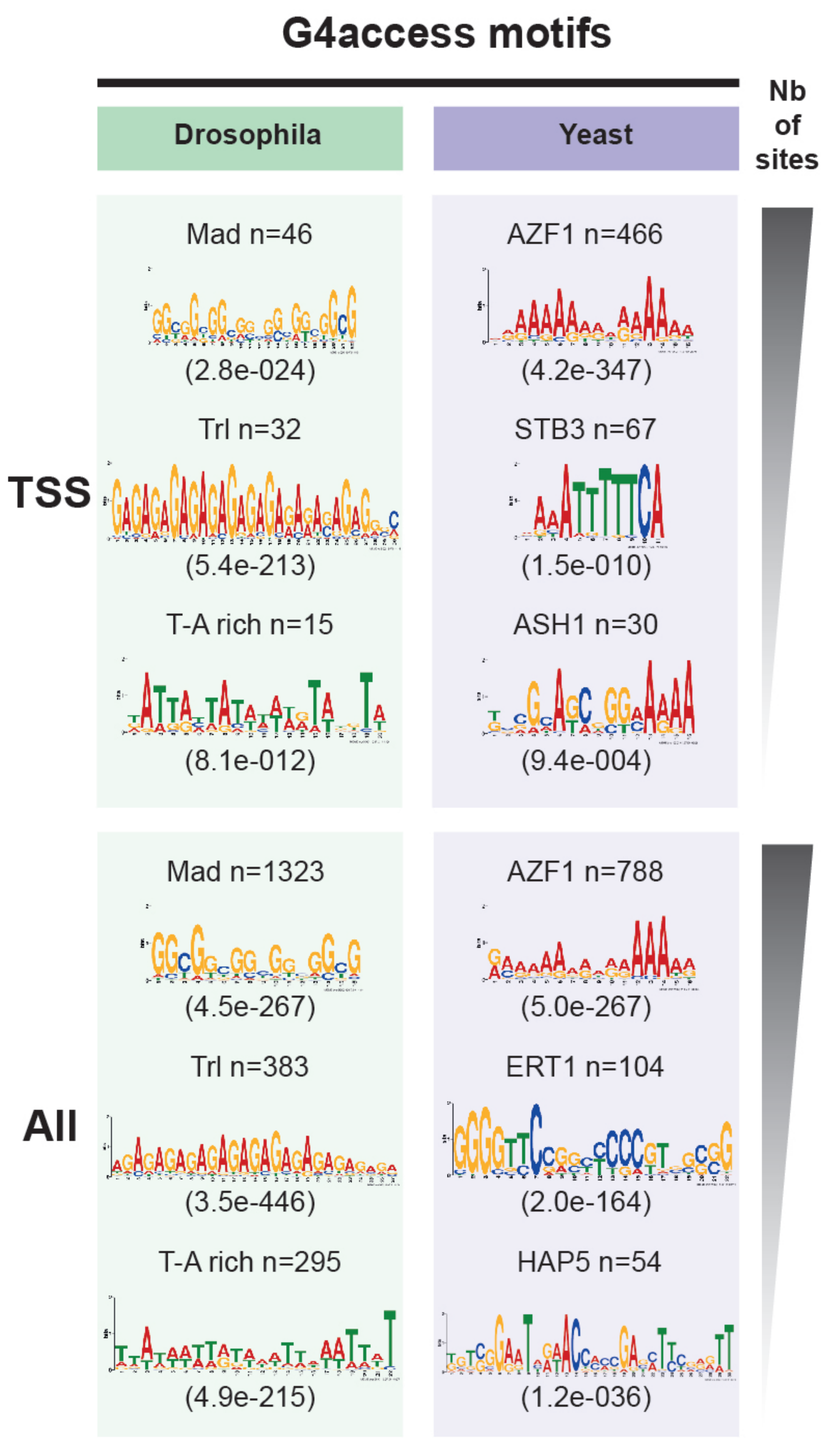
	G4H0.25	G4H0.5	G4H0.75	G4H1.0	G4H1.25	G4Hh1.5	G4H1.75	G4H2.0
Human	**1.11**	**1.46**	**2.05**	**3.08**	**4.75**	**6.40**	**8.48**	**10.74**
Drosophila	1.15	1.07	1.03	0.99	0.99	1.02	0.99	1.11
Yeast	0.98	0.99	**1.04**	**1.11**	**1.22**	**1.41**	**1.74**	**1.76**

b

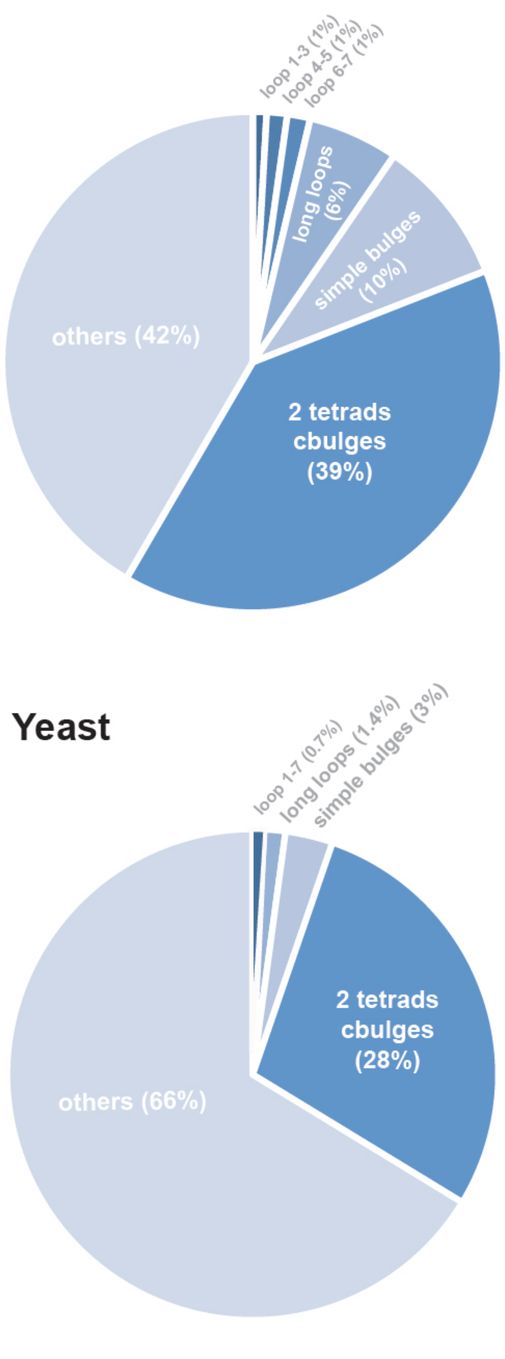




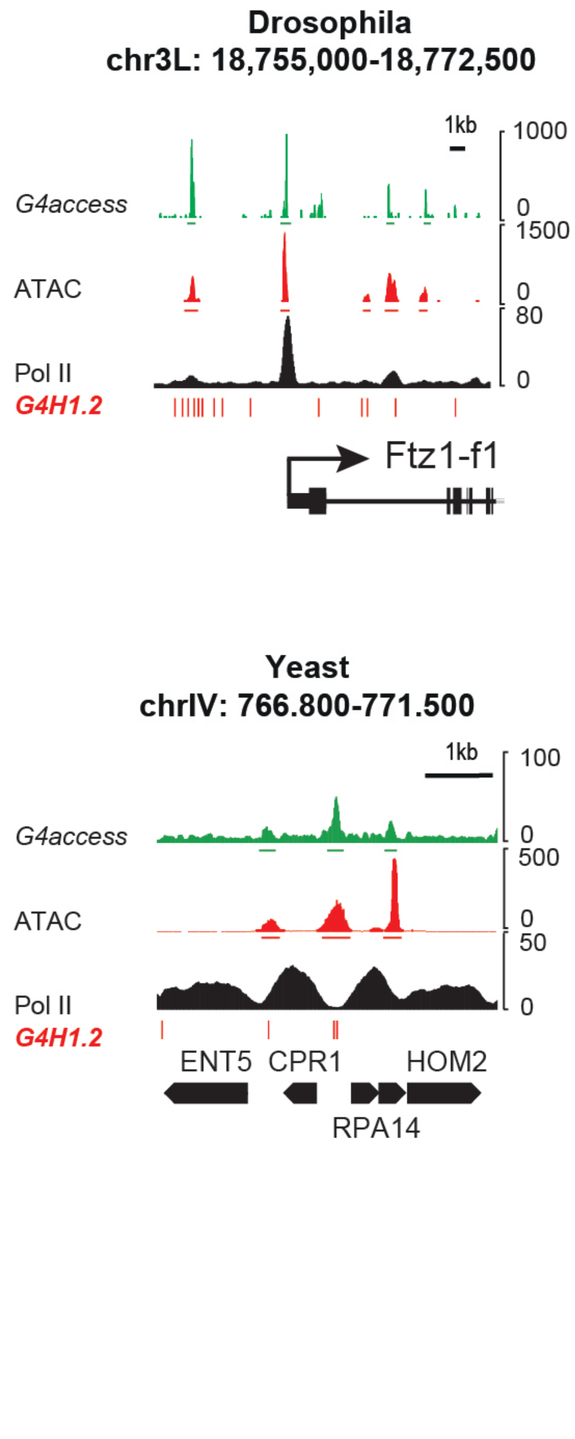
c



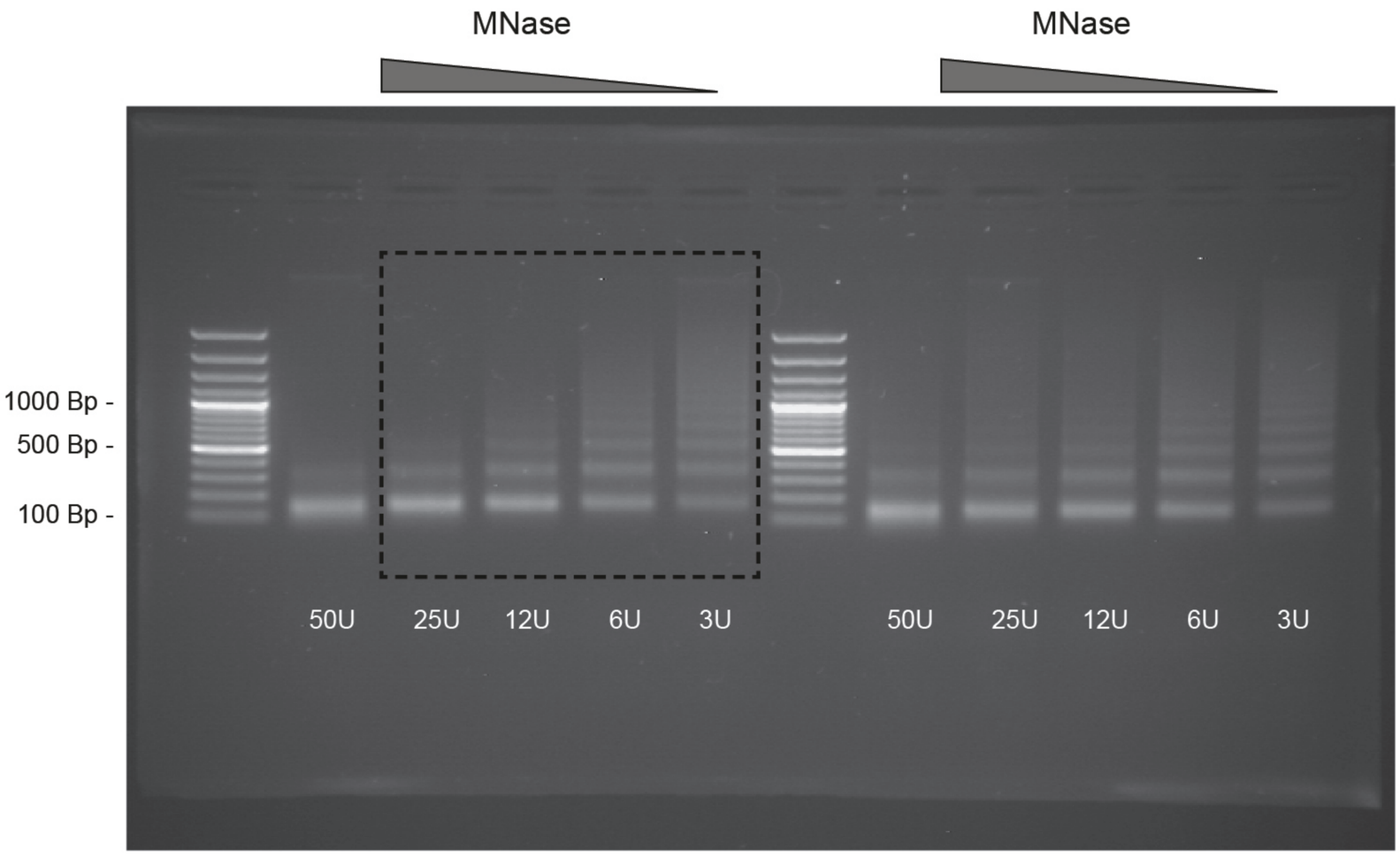
d



e



source data Fig. 1c



source data Fig. 3a and Fig. 3b

source data Extended Data Fig. 9a