



**HAL**  
open science

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

Cyril Esnault, Talha Magat, Amal Zine El Aabidine, Encar Garcia-Oliver, Anne Cucchiarini, Soumya Bouchouika, David Llères, 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-04793061>

**HAL Id: hal-04793061**

**<https://hal.science/hal-04793061v1>**

Submitted on 25 Nov 2024

**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.



HAL Authorization

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

Received: 22 October 2021

Accepted: 31 May 2023

Published online: 03 July 2023

 Check for updates

Cyril Esnault<sup>1,7</sup>, Talha Magat<sup>1,7</sup>, Amal Zine El Aabidine<sup>1,7</sup>, Encar Garcia-Oliver<sup>1,7</sup>, Anne Cucchiari<sup>2</sup>, Soumya Bouchouika<sup>1</sup>, David Lleres<sup>1</sup>, Lutz Goerke<sup>1</sup>, Yu Luo<sup>2,3</sup>, Daniela Verga<sup>3</sup>, Laurent Lacroix<sup>4</sup>, Robert Feil<sup>1</sup>, Salvatore Spicuglia<sup>5,6</sup>, Jean-Louis Mergny<sup>2</sup> & Jean-Christophe Andrau<sup>1</sup>✉

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 new tool for studying G4s in cellular dynamics and highlights their association with open chromatin, transcription and their antagonism to DNA methylation.

Eukaryotic promoters encompass a wide range of sequences 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<sup>1</sup>. We previously showed that mammalian CpG islands (CGIs) intrinsically exclude nucleosomes, independently of transcription<sup>2</sup>. 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 have a crucial role in nucleosome exclusion both in cells and *in vitro*<sup>3</sup>. Unimolecular G4s are DNA secondary structures well-characterized *in vitro*<sup>4</sup>. Their investigation in living cells is more recent, and they are suggested to have essential roles in transcription, replication, genome stability and homeostasis<sup>5</sup>. They can also be predicted by computational algorithms<sup>6,7</sup> such as G4Hunter which calculates a robust likelihood score of genomic sequences forming G4 structures at fixed window sizes.

<sup>1</sup>Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS-UMR 5535, Montpellier, France. <sup>2</sup>Laboratoire d’Optique et Biosciences, Ecole Polytechnique, CNRS, Inserm, Institut Polytechnique de Paris, Palaiseau, France. <sup>3</sup>Université Paris-Saclay, Institut Curie, Orsay, France. <sup>4</sup>Institut de Biologie de l’École Normale Supérieure, ENS, CNRS UMR8197, Inserm U1024, Paris, France. <sup>5</sup>Aix-Marseille University, INSERM, TAGC, UMR 1090, Marseille, France. <sup>6</sup>Equipe Labellisée Ligue Contre le Cancer, Marseille, France. <sup>7</sup>These authors contributed equally: Cyril Esnault, Talha Magat, Amal Zine El Aabidine, Encar Garcia-Oliver. ✉e-mail: [jean-christophe.andrau@igmm.cnrs.fr](mailto:jean-christophe.andrau@igmm.cnrs.fr)

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<sup>8–10</sup>. However, G4–ChIP is highly dependent on the selectivity of G4 recognition by the BG4 nanobody<sup>11</sup>, which may stabilize unstable structures *in vitro*<sup>12,13</sup>, introducing a potential bias. Finally, previous observations report that it is difficult to apply G4–ChIP to all cell types, in particular noncancerous primary cells<sup>8</sup>. Therefore, orthogonal methods are needed to identify G4s formed in the chromatin context.

Here we describe ‘G4access’, an antibody- 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 (MNase), 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 II (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 the 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 mouse embryonic stem cells (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 (pG4)-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<sup>2</sup>. 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, pG4s densities at promoters confirmed an association between pG4 and the lowest nucleosome density (Fig. 1a and Extended Data Fig. 1a), consistent with the previous description<sup>14</sup>. Based on this observation, we hypothesized that we could

map G4s in living cells based on their nucleosome exclusion potential. For this purpose, we developed a technique to isolate subnucleosomal (–147 bp) DNA fragments that are protected by DNA secondary structures using 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 endonuclease and exonuclease activity and has been reported to have a cutting preference before G-stretches<sup>15,16</sup>, while G4s are also resistant to  $\lambda$ -exonuclease<sup>17</sup>. 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). Finally, we 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 section ‘G4access hallmarks nucleosome exclusion and transcription’) and is consistent with G4–ChIP profiles<sup>8,10</sup> (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–pG4s were comparable across the two techniques (Fig. 1f,g and Extended Data Fig. 2d). Shuffling nucleotides while keeping base composition constant confirmed the methods’ specificity (Extended Data Fig. 2e). Notably, 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*<sup>6</sup>. We confirmed this observation by scoring G4 subtype structures from motif predictions in individual cell lines. We found that 75–90% of the sequences fit one or another pG4 category (Extended Data Fig. 3a). When considering peaks common to the three 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 datasets. 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

**Fig. 1 | G4access principle and validation.** **a**, G4FS at promoters and motifs associated with 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 before library preparations. **c**, Initial setup, 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 the initial setup. Bottom left is the fraction of mononucleosomes in the titration curve expressed as a percent of all DNA (mean and s.d. of  $n = 2$  biological replicates are displayed). The second point (30% mononucleosome) is considered as 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

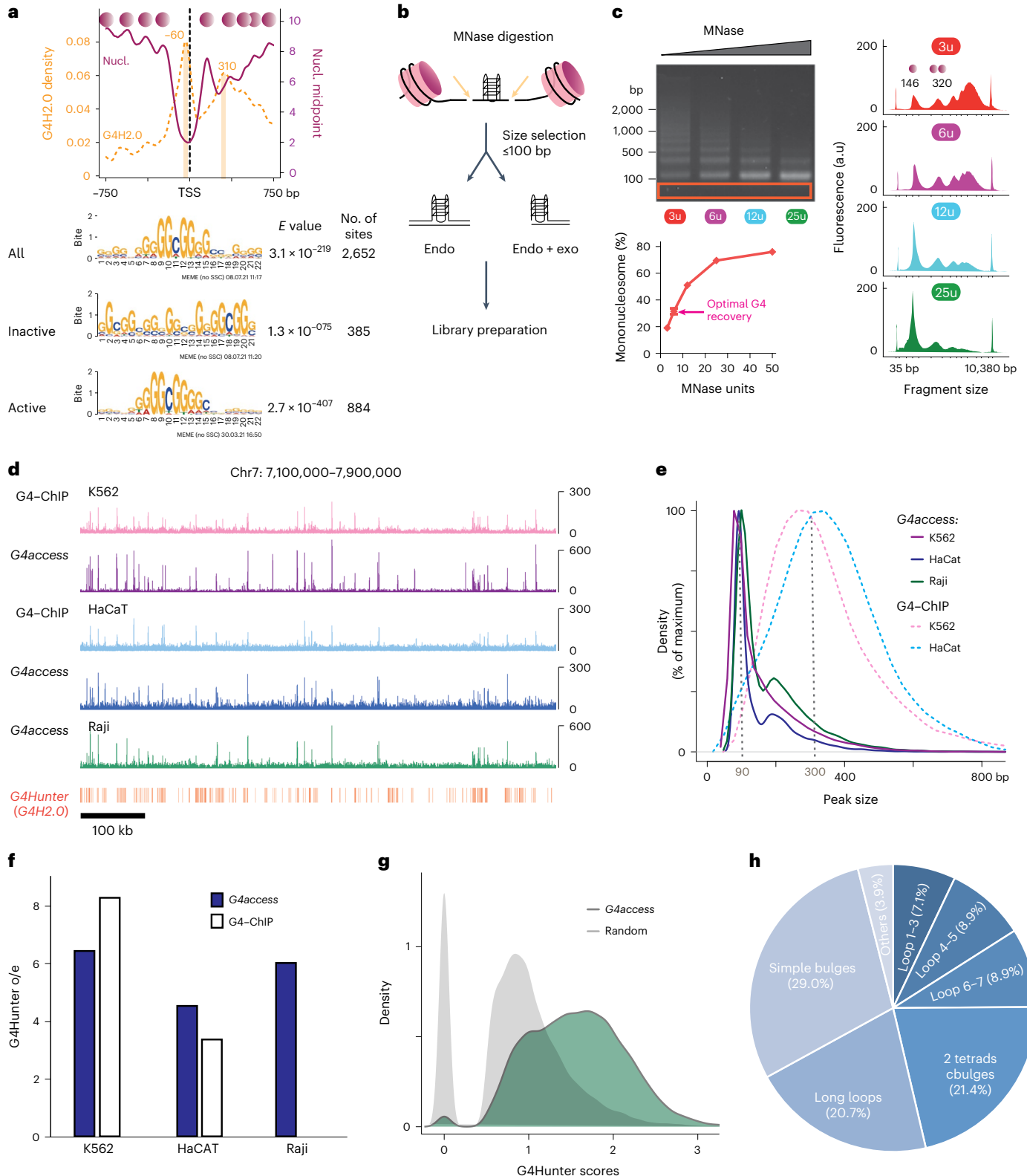
signal and corresponding G4–ChIP<sup>8,10</sup> in indicated cell lines. G4H2.0 predictions are indicated below. **e**, Peak size distribution in G4–ChIP and G4access. Vertical dashed lines indicate the maximum (bulk) sizes of fragments in the sequenced samples. **f**, Observed/expected G4Hunter predictions 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*<sup>6</sup>. The difference in distribution is highly significant ( $P < 2.2 \times 10^{-16}$ , 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 (Methods).

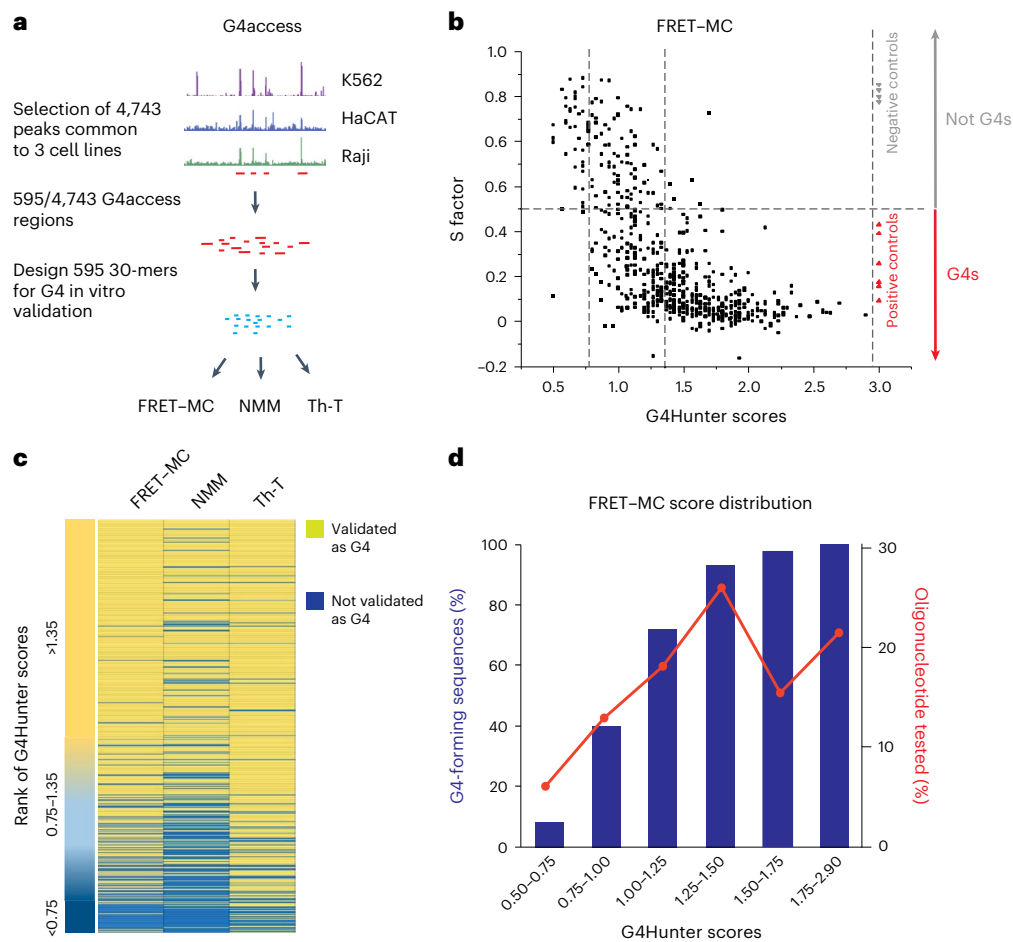
(Extended Data Fig. 3d). As expected, their GC and CpG content at promoters is also higher than average, consistent with a strong association with CGIs (Extended Data Fig. 3e). We also observed G4access-enriched promoters at genes expressed in multiple cell types, with the possible exception of DNA repair 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 G4s in vitro**

G4Hunter-pG4s were previously validated in vitro on a set of mitochondrial sequences<sup>6</sup>. We, therefore, asked if G4access-identified G4FS could, at a large scale, be validated using three independent in vitro assays.





**Fig. 2 | Experimental validation of G4 structure in G4access selected sequences.**

**a**, Strategy for in vitro validation of G4access sequences common to the three cell lines as indicated. **b**, FRET-MC assessment of the G4 structures on the 596 representative 30-nt oligonucleotides. All oligonucleotides with an S factor < 0.5 are considered G4s. Positive and negative controls are shown on

the side (Supplementary Table 2). **c**, Validated (yellow) and nonvalidated (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).

Firstly, we identified 4,743 common G4access peaks in all three cell lines and then selected 596 representative 30-nucleotide (nt) regions with G4Hunter scores > 0.5 (Methods; Fig. 2a). We next assessed their abilities to form G4 structures by performing Fluorescence Resonance Energy Transfer–Melting Competition (FRET–MC)<sup>18</sup>, 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 Th-T 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.

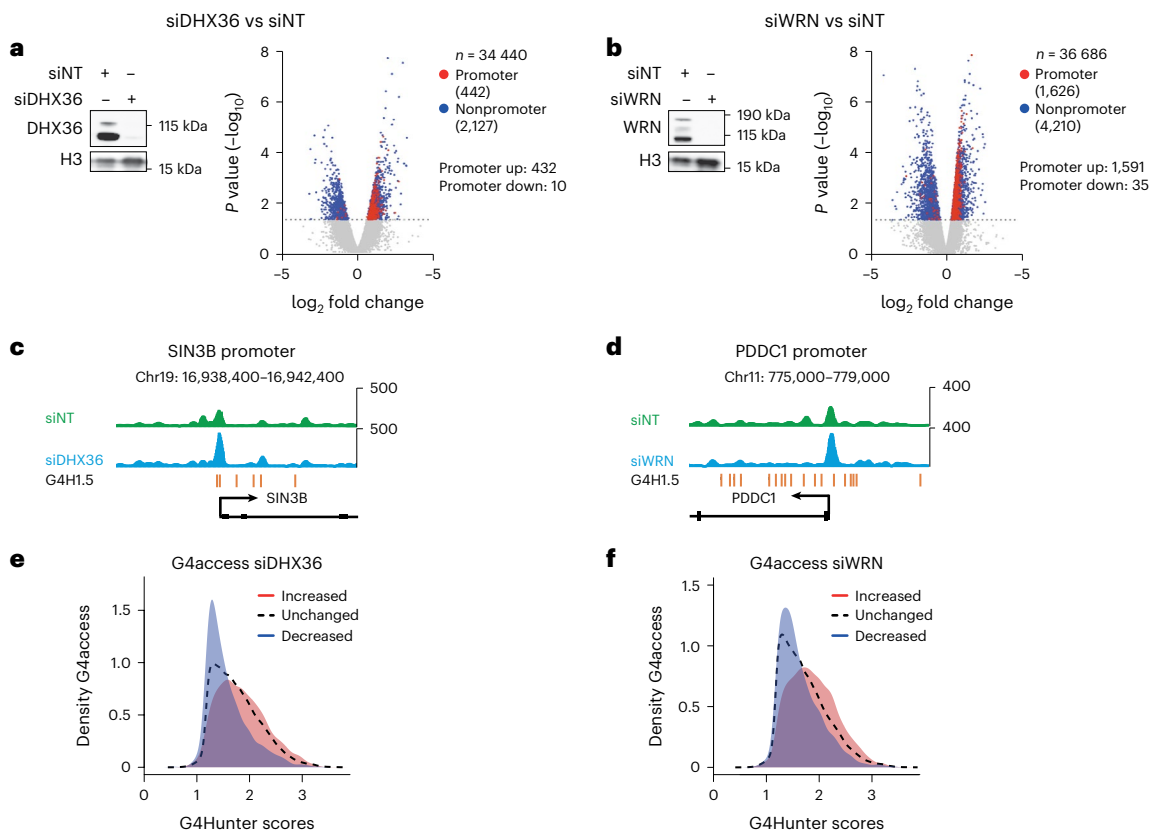
In sum, our in vitro analyses confirm that G4access strongly enriches G4FS, in line with our genomic observations.

#### 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<sup>19</sup>. To get further insights into the G4 ligand's action, we treated the cells with pyridostatin (PDS), a well-known G4 ligand<sup>19</sup>, 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 because 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 analyzed previously published genome-wide in vitro data using G4seq experiments with or without PDS<sup>20</sup>. 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.



**Fig. 3 | G4access measures G4 dynamics in response to siRNA targeting G4 helicases.** **a**, siRNA targeting DHX36 efficiently reduces the helicase expression level and promotes the 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 knockdown cells. The resulting volcano plot, performed on sequences with  $G4H > 1.2$ , is displayed, (red) differential promoter signal, (blue) differential non-TSS signal, (gray) unaffected signals (DESeq,  $P < 0.05$ ). 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, (gray) unaffected signals (DESeq,  $P < 0.05$ ). Majority (97.8%) of differential promoters show an increased G4access signal (1,591/1,626;  $n = 2$  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 a 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.

Next, we investigated the impact of known G4 helicases disruption by siRNA knockdown experiments of the WRN and DHX36 helicases, previously described as unwinding G4s<sup>21,22</sup>. 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 upregulated 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 upregulated 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 nonpromoter regions of the genome. To improve the accuracy of our analysis, we selected G4access peaks associated with medium stringency G4 annotations (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 nonpromoter 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 four classes of nucleosome positioning and depletion patterns. We observed that nucleosome positioning is generally conserved in  $\frac{3}{4}$  of the regions (classes 2 and 3), while a minority of regions in which G4s deplete or position more accurately nucleosomes (class 1 and class 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 pG4s cannot open chromatin and thus do not affect the 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<sup>23</sup> (Extended Data Fig. 7a).

Because G4access also isolates non-G4FS, we wondered whether non-G4s showed similar properties. In contrast to the pG4-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 nucleosome-depleted region (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 torsional stress that may stabilize G4s *in vitro*. Our recent observations suggested the opposite, that is, that G4s would promote chromatin opening and Pol II recruitment<sup>3,8,24</sup>. To further address this question, we treated the cells with transcription inhibitors specific for initiation or elongation (triptolide and KM05283, respectively) and analyzed if this resulted in the loss of the G4access signal. As previously reported, we found that triptolide treatment stripped the Pol II signal across all genic features while KM05283 only removed Pol II from gene bodies<sup>25,26</sup> (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 with 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<sup>27</sup>, although in this case, chromatin immunoprecipitation followed by sequencing (ChIP-seq) did not identify changes in the G4 signal in the presence of inhibitors. This difference might originate from the cell types, the techniques used or the bioinformatic setup 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 assay for transposase-accessible chromatin with sequencing (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 trichostatin A (TSA), an histone deacetylase (HDAC) inhibitor. While TSA globally relaxes chromatin through histone hyperacetylation (Extended Data Fig. 9a), it can reduce ATAC-seq signals at open chromatin regions<sup>28,29</sup>. We confirmed this effect in MNase-seq following 24 h TSA treatment where nucleosome densities increased at NDR locations. At both promoter and non-promoter G4access locations, we also observed a consistent reduction

of the signals where pG4s are present, indicating that nucleosomes tend to reposition 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<sup>8,27</sup>. However, when growth conditions such as hypoxia result in chromatin compaction at pG4 sites, G4 formation is also reduced<sup>26</sup>.

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, pG4s 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<sup>30</sup>. Because of their parental DNA methylation imprints, DMRs are stably repressed on their methylated allele and are transcriptionally active on the nonmethylated allele<sup>30</sup>. 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 mESCs obtained through reciprocal crosses between *Mus musculus domesticus* C57BL/6J and *Mus musculus molossinus* JF1 (ref. 31) 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-characterized mouse-imprinted DMRs (Supplementary Table 1; Methods). We observed differential allelic G4access signals at seven of them, of 11 DMRs showing signal ( $P < 0.05$ ; 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, *Peg3*, 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 favor of nucleosome association.

### DNA methylation antagonizes G4 formation and open chromatin

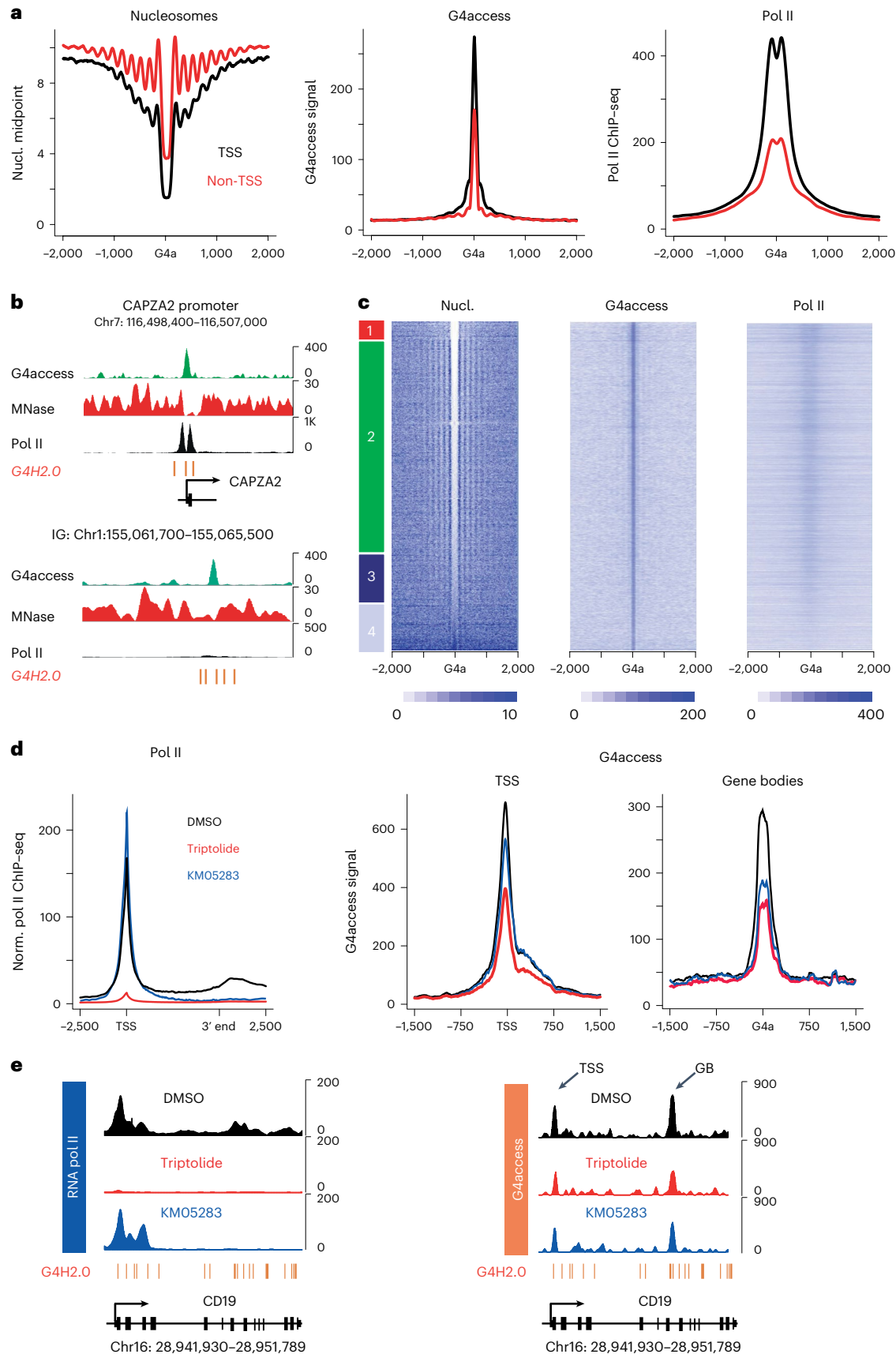
To gain further insight into this question, we analyzed methylation at G4Hunter-predicted regions with or without G4access signal (Fig. 6a,b). These analyses were performed at medium G4 predictions (G4HL.2) that contained at least two CpG nucleotides. We found that experimentally

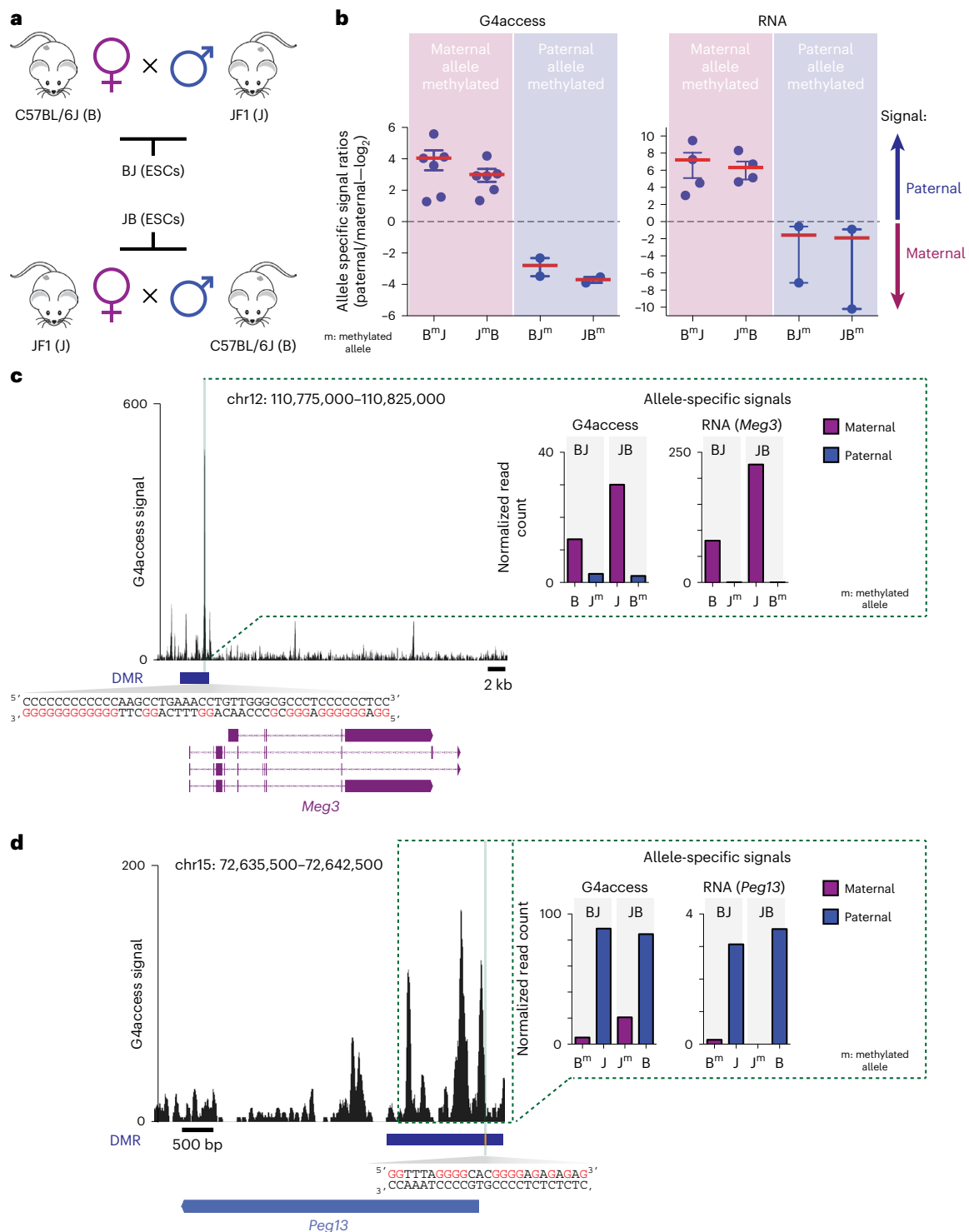
**Fig. 4 | G4s are associated with open chromatin, nucleosome positioning and Pol II and are partially independent of transcription.** **a**, Nucleosome positioning and Pol II recruitment centered on G4access sites overlapping G4 predictions at promoter (TSS,  $n = 4,619$ ) and nonpromoter sites (non-TSS,  $n = 11,807$ ; see also Methods). **b**, Examples of G4access peaks, nucleosome depletion and positioning at promoter (upper panel, chr7:116.498.400–116.507.000) and nonpromoter (lower panel, chr1:155.061.700–155.065.500) sites in Raji cells. G4H regions with a score  $>2.0$  are shown below. **c**, G4access peaks associate with strong nucleosome depletion, Pol II recruitment and define highly positioned nucleosomes at most genomic G4s. Heatmaps are shown ranked by decreasing MNase signals around G4access peaks overlapping weak/

moderate G4 annotations (Methods). Four groups were defined based on the features of the individual MNase profiles as follows: (1) 844, strong depletion and low positioning; (2) 6,215, depletion and strong positioning; (3) 2,073, moderate depletion and positioning and (4) 2,675, 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,  $n = 1,808$ ) or G4access at promoters ( $n = 1,808$ ) and gene bodies ( $n = 349$ ) following triptolide and KM05283 treatment. **e**, Examples of transcription inhibition effects on Pol II and G4access signal over the *CDI9* gene (chr16:28.941.930–28.951.789) in Raji cells. G4H regions with a score  $>2.0$  are shown below.

identified G4s are associated with a loss of CpG methylation, consistent with previous observations<sup>10</sup>. Yet, we also found that the G4s corresponding to weak/medium predictions tend to be more methylated

when compared to neighboring DNA. These results suggest that G4 formation favors open chromatin formation, while pG4 DNA methylation antagonizes this process. To further address this issue at the



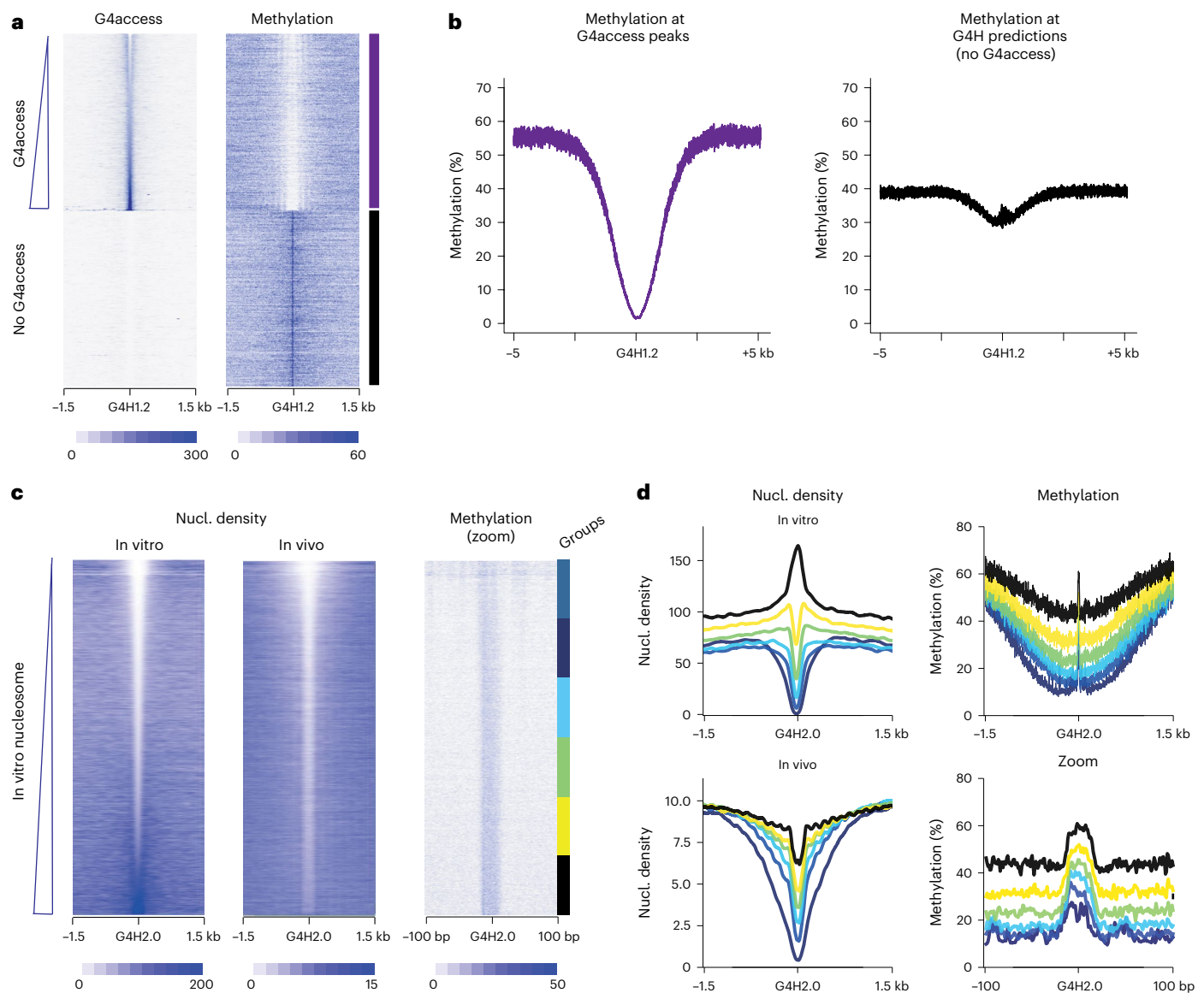


**Fig. 5 | Differential G4access signal at active and inactive imprinted genomic loci in mouse hybrid ES cells. a**, Reciprocal crosses of hybrid mouse strains (JF1 and C57/Bl6) used in this study for maternal and paternal imprinting (2i-medium derived cell lines). **b**, Allele-specific G4access and 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<sup>m</sup> or J<sup>m</sup>, depending on the strain it originates from (see Supplementary Table 1). Individual datasets, mean with s.e.m. 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.

most basic level, in the context of naked chromatin, we re-analyzed data from *in vitro* nucleosome assembly using human granulocyte genomic DNA<sup>32</sup>. We ranked the *in vitro* nucleosome densities at pG4s

by increasing signal (Fig. 6c) and analyzed corresponding signals for *in vivo* nucleosomes and DNA methylation. This analysis indicates a clear relationship between methylation levels at pG4s and nucleosome



**Fig. 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 two CpG are separated into two groups: top, G4access peaks overlapping pG4s ranked by increasing G4access signal  $n = 16,820$  and bottom, randomly selected pG4s (G4H > 1.2) at sites that do not show any G4access peak  $n = 16,820$ . 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  $\pm 5$  kb. **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,  $\pm 50$  bp) that contain at least two CpGs are shown ( $n = 52,854$ ). They are ranked by increasing in vitro MNase-seq signals. Six groups of equal sizes ( $n = 8,809$ ) were defined (right, colors are displayed). **d**, Corresponding metaprofiles of the heatmaps and the six groups defined in **c** are shown. A zoom-in at 100 bp of the DNA methylation signals is also displayed.

formation in vitro and in vivo (Fig. 6d). These results propose that high methylation of G4FS results in a sequence context favorable to intrinsic nucleosome formation and G4 structure destabilization.

#### G4access enriches 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 *Drosophila melanogaster* and *Saccharomyces 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 toward 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 an S288C yeast strain, following an adjusted workflow (Methods). We compared the procedure's performance to isolate G4FS in the three 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<sup>14,33</sup>.

## Discussion

In this study, we introduce G4access, an efficient procedure for scoring G4FS associated with open chromatin in cells. Because G4 formation requires the 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<sup>+</sup> concentration during the 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<sup>3,34</sup>. We further show that G4s are maintained without active transcription, albeit reduced. This indicates that G4s may be formed before transcription and that transcription further stabilizes their structure, reconciling previous conflicting observations<sup>8,27</sup>. This result is also consistent with the idea that negative supercoiling upstream of the transcription front increases G4 stability<sup>24</sup>.

Our study also finds that G4s might contribute to imprinted gene expression. Among 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 have 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 disfavors 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<sup>35</sup>, 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*<sup>36–38</sup>. This could originate from the difference in topology between oligonucleotides and dsDNA embedded in nucleosomes *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 not only enable future investigations to better understand the role of G4s in transcription control but also other genomic processes such as replication<sup>39</sup>, DNA repair and the role of helicase in genome stability<sup>40</sup>, as well as deepening the mechanistic understanding of various G4-targeting drugs, including some that are believed to display an anticancer potential<sup>41</sup>.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-023-01437-4>.

## References

- Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* **10**, 161–172 (2009).
- 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 (2012).
- Esnault, C. et al. G-quadruplexes are promoter elements controlling nucleosome exclusion and RNA polymerase II pausing. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.02.24.529838> (2023).
- Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770–780 (2012).
- 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 (2017).
- Bedrat, A., Lacroix, L. & Mergny, J. L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746–1759 (2016).
- Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908–2916 (2005).
- Hansel-Hertsch, R. et al. G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267–1272 (2016).
- Li, C. et al. Ligand-induced native G-quadruplex stabilization impairs transcription initiation. *Genome Res.* **31**, 1546–1560 (2021).
- Mao, S. Q. et al. DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol.* **25**, 951–957 (2018).
- 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 (2020).
- Tran, P. L. T. et al. Folding and persistence times of intramolecular G-quadruplexes transiently embedded in a DNA duplex. *Nucleic Acids Res.* **49**, 5189–5201 (2021).
- van Holde, K. & Zlatanova, J. Unusual DNA structures, chromatin and transcription. *BioEssays* **16**, 59–68 (1994).
- Hershman, S. G. et al. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **36**, 144–156 (2008).
- Dingwall, C., Lomonosoff, G. P. & Laskey, R. A. High sequence specificity of micrococcal nuclease. *Nucleic Acids Res.* **9**, 2659–2673 (1981).
- Horz, W. & Altenburger, W. Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic Acids Res.* **9**, 2643–2658 (1981).
- 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.* **25**, 725–735 (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* **112**, e23415 (2020).
19. Rodriguez, R. et al. Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301–310 (2012).
20. Chambers, V. S. et al. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33**, 877–881 (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 (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 (2005).
23. Natoli, G. & Andrau, J. C. Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* **46**, 1–19 (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 Res.* **46**, 7418–7424 (2018).
25. 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 (2014).
26. 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 (2005).
27. Shen, J. et al. Promoter G-quadruplex folding precedes transcription and is controlled by chromatin. *Genome Biol.* **22**, 143 (2021).
28. Cusack, M. et al. Distinct contributions of DNA methylation and histone acetylation to the genomic occupancy of transcription factors. *Genome Res.* **30**, 1393–1406 (2020).
29. Vaid, R., Wen, J. & Mannervik, M. Release of promoter-proximal paused Pol II in response to histone deacetylase inhibition. *Nucleic Acids Res.* **48**, 4877–4890 (2020).
30. 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 (2013).
31. Sanli, I. et al. Meg3 non-coding RNA expression controls imprinting by preventing transcriptional upregulation in cis. *Cell Rep.* **23**, 337–348 (2018).
32. Valouev, A. et al. Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516–520 (2011).
33. 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 (2010).
34. 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 (2008).
35. Lay, F. D. et al. The role of DNA methylation in directing the functional organization of the cancer epigenome. *Genome Res.* **25**, 467–477 (2015).
36. 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 (2021).
37. 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 (2014).
38. Wang, Z. F. et al. Cytosine epigenetic modification modulates the formation of an unprecedented G4 structure in the WNT1 promoter. *Nucleic Acids Res.* **48**, 1120–1130 (2020).
39. Valton, A. L. & Prioleau, M. N. G-Quadruplexes in DNA replication: a problem or a necessity? *Trends Genet.* **32**, 697–706 (2016).
40. Mendoza, O., Bourdoncle, A., Boule, J. B., Brosh, R. M. Jr. & Mergny, J. L. G-quadruplexes and helicases. *Nucleic Acids Res.* **44**, 1989–2006 (2016).
41. Asamitsu, S., Obata, S., Yu, Z., Bando, T. & Sugiyama, H. Recent progress of targeted G-quadruplex-preferred ligands toward cancer therapy. *Molecules* **24**, 429 (2019).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023

## 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 Roswell Park Memorial Institute medium (RPMI-1640, Thermo Fisher Scientific, 11875085) supplemented with 10% heat-inactivated fetal bovine serum (Sigma, 12103C), 100 units per ml penicillin, and 100  $\mu\text{g ml}^{-1}$  streptomycin (Thermo Fisher Scientific, 15140122) at 37 °C in a humidified 5% CO<sub>2</sub> incubator. Human Raji cells, grown at an approximative density of 10<sup>6</sup> cells per ml, were treated with 10  $\mu\text{M}$  PDS (Sigma, SML0678) or H<sub>2</sub>O (control) for 30 min, with 2  $\mu\text{M}$  TSA (Sigma, T8552) or dimethyl sulfoxide (DMSO, control) for 24 h 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 h with DMSO 0.1% (control), 1  $\mu\text{M}$  triptolide (inhibiting initiation stage; Sigma Merck) or 100  $\mu\text{M}$  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 Institut de Recherche en Cancérologie de Montpellier (IRCM) screening platform in oncology) was cultured in DMEM-high glucose medium (Thermo Fisher Scientific, 11965084) supplemented with 10% heat-inactivated fetal bovine serum (Sigma, 12103C), 100 units per ml penicillin and 100  $\mu\text{g ml}^{-1}$  streptomycin (Thermo Fisher Scientific, 15140122) at 37 °C in a humidified 5% CO<sub>2</sub> 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 per l) and glutamine (2.9 mg l<sup>-1</sup>) at 37 °C and 5% CO<sub>2</sub>. Knockdown experiments were performed using Invitrogen siRNA Silencer Select (siDHX36 and siWRN, 4392420) control siRNA (Dharmacon, ON-TARGETplus Nontargeting Pool) and lipofectamine RNAiMAX (Thermo Fisher Scientific) following manufacturer instructions. Knockdown efficiencies were analyzed by western blot 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<sup>31</sup>. 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 collection 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<sub>2</sub> 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 per ml penicillin and 100  $\mu\text{g ml}^{-1}$  streptomycin (Thermo Fisher Scientific, 15140122) at 27 °C and collected at 2 × 10<sup>6</sup> cells per ml. Yeast S288C (BY4741, gift from M. Radman-Livaja) haploid cells were grown up to and collected at the mid-log phase (OD<sub>600</sub> = 0.5) in 100 ml of Yeast extract-Peptone-Dextrose (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 for 10 min at 4 °C. For S2 semi-adherent cells, cells were resuspended by pipetting up and down multiple times. All cell pellets were rinsed twice in 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<sub>2</sub>PO<sub>4</sub>, 5 mM MgCl<sub>2</sub>, 0.5 mM CaCl<sub>2</sub> and 35 mM N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES, pH 7.4)) and homogenized using a dounce (40 strokes). HeLa cells tend to clump under such conditions. Therefore, nuclei were first purified under a sucrose cushion as described<sup>42</sup>. 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 MgCl<sub>2</sub>, 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  $\mu\text{g ml}^{-1}$  aprotinin) before the addition of ice-cold buffer 2 (15 mM Tris-HCl (pH 7.5), 300 mM sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl<sub>2</sub>, 0.1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 3.6  $\mu\text{g ml}^{-1}$  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 MgCl<sub>2</sub>, 0.1 mM EGTA, 0.5 mM DTT, 0.1 mM PMSF, 3.6  $\mu\text{g ml}^{-1}$  aprotinin). Cells were then centrifuged at 10,000g for 20 min at 4 °C. The supernatants were discarded, and the nuclei-containing pellets were resuspended in MNase digestion buffer for later steps (see two sentences below). For all mammalian cell lines, 5 × 10<sup>6</sup> cells per titration point were then resuspended in 50  $\mu\text{l}$  of prewarmed permeabilization buffer supplemented with 0.2% (vol/vol) NP40 and incubated for 5 min at 37 °C before digestion. The same procedure was performed for 50 × 10<sup>6</sup> cells for S2 *Drosophila* cells. MNase digestions were then performed by adding a volume of 500  $\mu\text{l}$  of prewarmed MNase reaction buffer (150 mM sucrose, 50 mM Tris-HCl (pH 8), 50 mM NaCl and 2 mM CaCl<sub>2</sub>) supplemented with either 3, 6, 12, 25 or 50 U of MNase (Merck, 10107921001). Digestions were incubated at 37 °C for 10 min and stopped on ice and by adding 11  $\mu\text{l}$  of 500 mM EDTA to each reaction followed by the addition of 550  $\mu\text{l}$  of SDS lysis buffer (1% (vol/vol) SDS, 10 mM EDTA and 50 mM Tris-HCl (pH 8)). Samples were then incubated for 10 min on ice. Before DNA purification, 1 ml of water was added to dilute the SDS, and the samples were incubated with 5  $\mu\text{l}$  of RNase A (Thermo Fisher Scientific, EN0531) at 37 °C for 2 h and with 8  $\mu\text{l}$  of proteinase K (Euro-medex, 09-0911) at 56 °C for 2 h to complete the lysis. To then quality control the MNase digestions, 125  $\mu\text{l}$  of each sample was 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% ( $\pm 5\%$ ) of mononucleosomes (Fig. 1c). Notably, this assessment should be performed on purified DNA that does not contain the subnucleosomal fraction, using bioanalyzer equipment. The remaining samples were 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 the best G4 sequence recovery. We also recommend that the bioanalyzer and agarose gel Quality Controls (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 nonpermeabilized 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 PBS. Pellets from 100 ml culture were suspended in 600  $\mu\text{l}$  of cell-breaking buffer (20% glycerol and 100 mM Tris (pH 7.5)), 600  $\mu\text{l}$  zirconia beads (0.5 mm) and 10  $\mu\text{l}$  of 100× protease inhibitors (Roche, 06538282001). Beads beating was performed in 1.5 ml screw-cap tubes in a Bullet Blender (Next Advance) for 4 × 3 min 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 5 min at 20g, and the pellets that contained the nuclei were resuspended in 300  $\mu\text{l}$  of prewarmed Nuclear Preparation (NP) buffer (50 mM NaCl, 10 mM Tris (pH 7.4), 5 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 0.2% NP40 (vol/vol), 0.5 mM spermidine (Sigma, S0266-1G) and 0.007%

$\beta$ -mercaptoethanol (Sigma, M3148-100ML)). Digestion took place by adding a volume of 300  $\mu$ l 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  $\mu$ l of stop buffer (5% SDS, 50 mM EDTA). Before DNA purification, digestions were incubated with 5  $\mu$ l of RNase A (Thermo Fisher Scientific, EN0531) at 37 °C for 2 h and with 10  $\mu$ l of proteinase K (Euromedex, 09-0911) at 56 °C for 2 h. Purification was performed with two consecutive phenol and one chloroform steps followed by ethanol and linear acrylamide precipitation. Notably, purified DNAs were once again incubated with 5  $\mu$ l RNase A (Thermo Fisher Scientific, 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<sup>+</sup> 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  $\mu$ g of each digestion product was migrated in a 4–20% polyacrylamide Novex TBE gel (Thermo Fisher Scientific, EC6225BOX) at 100 V for 60 min. The gels were stained with Syber Gold (Thermo Fisher Scientific, 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  $\mu$ m needle. These tubes were inserted into 1.5-ml collection tubes and centrifuged for 10 min at 15,300 rcf to collect the gel through the hole, generating gel beads. To elute the DNA from the beads, 700  $\mu$ l of water was added and the tubes were incubated overnight at 55 °C in a thermomixer at 1500 r.p.m. DNA was purified by transferring the samples (DNA eluate and gel beads) to the top of a 0.22  $\mu$ m spin filter (Agilent, 5185-5990). Spin filters were centrifuged for 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 mononucleosomes 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 datasets 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 adapter-ligated DNA was performed by using CleanNGS beads (CNGS-0050) with a bead:DNA ratio of 2:1. The purified products were amplified with eight 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), paired 50–30 bp reads (K562, mES cells) or an Illumina HS4000 using paired 76 bp reads (HaCaT cells). Libraries from HeLa, Raji cells treated with TSA, Drosophila and yeast cells were sequenced on Novaseq 6000 Sequencer in paired-end (50–50 bp) 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 LBI (50 mM HEPES (pH 7.5), 140 mM NaCl, 1 mM EDTA (pH 8), 10% glycerol, 0.75% NP40, 0.25% Triton X-100) at 4 °C for 20 min on a rotating wheel. Nuclei were pelleted down by spinning at 1,350 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 1 ml 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 (LBI, 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 20,000 rcf and 4 °C for 10 min to remove particulate matter. After taking aside a 50  $\mu$ l 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 2 $\times$  elution buffer (100 mM Tris (pH 8.0), 20 mM EDTA, 2% SDS) and incubated at 65 °C for 12 h 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 (0.2  $\mu$ g ml<sup>–1</sup>) at 37 °C for 1 h and proteinase K (0.2  $\mu$ g l<sup>–1</sup>) for 2 h at 55 °C. DNA was isolated by phenol:chloroform:isoamyl alcohol (25:24:1, pH 8) extraction followed by Qiaquick PCR Purification (Qiagen). Purified DNA was then analyzed on a 1.5% agarose gel and on Bioanalyzer (Agilent) 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 chromatin (10  $\times$  10<sup>6</sup> Raji cells equivalent and 5  $\times$  10<sup>6</sup> Drosophila S2 cells equivalent (2:1 spike-in ratio)) were added to pre-coated beads (250  $\mu$ l), and the mix was incubated overnight at 4 °C on a rotating wheel. After incubation with chromatin, beads were washed seven times with wash buffer (50 mM HEPES (pH 7.6), 500 mM LiCl, 1 mM EDTA (pH 8), 1% NP40, 0.7% Na-deoxycholate, 1 $\times$  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  $\mu$ l 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 h to reverse-crosslink the chromatin followed by treatment with RNase A and proteinase K and purification of DNA, as described in the previous paragraph 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 (Thermo Fisher Scientific). In total, 1 ng of ChIP DNA was used to prepare sequencing libraries with NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, E7645S). After end-repair and adapter ligation, library fragments were amplified by 13 cycles of PCR. Barcoded libraries from different samples were pooled together and sequenced on Illumina NextSeq-500 Sequencer in paired-end (50–30 bp) sequencing runs.

### Nascent chromatin-associated RNAs sequencing (ChrRNA-seq)

ChrRNAs presented in Extended Data Fig. 3g were either analyzed from GSE90238 or isolated from 2  $\times$  10<sup>7</sup> Raji cells as follows. Nuclear fractionation was performed by incubating cells in 4 ml of buffer 1 (10 mM Tris–HCl (pH 7.5), 10 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 0.5% IGEPAL CA-630) on ice for 5 min. Next, we carefully underlaid 1 ml of buffer 2 (10 mM Tris–HCl (pH 7.5), 10 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 0.5% IGEPAL CA-630, 10% sucrose) before collecting the nuclear fraction at 1400 rcf for 5 min at 4 °C. Nuclei were resuspended with 125  $\mu$ l of NUN1 buffer

(20 mM Tris-HCl (pH 8), 75 mM NaCl, 0.5 mM EDTA, 50% glycerol) followed by 1.2 ml of NUN2 buffer (20 mM HEPES-KOH (pH 7.6), 300 mM NaCl, 0.2 mM EDTA, 7.5 mM MgCl<sub>2</sub>, 1% IGEPAL CA-630, 1 M urea). After 15 min of vigorous vortexing, the chromatin was centrifuged at 15,000 rcf for 10 min at 4 °C and washed with 500 µl of NUN2 buffer. After discarding the supernatant, the chromatin was resuspended in 500 µl 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 (1 ml–200 µl–10 µl) and then pushing through very small syringe needles. Then RNA was extracted from chromatin according to the Trizol manufacturer's protocol and resuspended in 50 µl of nuclease-free water (Invitrogen, AM9906) followed by TurboDNase (Invitrogen, AM2238) treatment. Trizol RNA extraction and TurboDNase were performed two more times, and RNA was resuspended in 20 µl of nuclease-free water. Purified RNAs were quantified by Qubit, and quality was assessed using the RNA Assay kit (Agilent RNA 6000 Pico reagents, 1567-1513) with Bioanalyzer 2100 (Agilent Technologies). ChrRNAs were then subjected to library preparation using the True-seq stranded total RNA library prep gold kit (Illumina, 220599) using 1 µg of ChrRNA, with a total of 15 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 nonsize 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 TGACCGCAAAGCTGTACAC. Quantitative PCR reaction was performed using the Platinum Taq DNA polymerase (Thermo Fisher Scientific, 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 s/65 °C for 30 s/72 °C for 15 s, followed by melting curve analysis. Results presented in 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 G4s and one hairpin to test their stabilities in MNase digestion conditions. FRET buffers were lithium cacodylate supplemented with 140 mM KCl and 10 mM KCl with 45 mM NaCl for physiological and MNase 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 °C. 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 4,743 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 synthesized and purchased at Eurogentec (Seraing) with RP cartridge gold purification, and further used for FRET-MC, Th-T 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 (Bio-Rad). Oligonucleotides were annealed at 7.5 µM strand concentration (95 °C, 5 min) 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 (T<sub>m</sub>) was defined when the normalized signal was 0.5. ΔT<sub>m</sub> corresponds to the difference of T<sub>m</sub> between the oligonucleotide with and without PhenDC3. Each sequence was tested in an intraday 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 Th-T 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 fluorescent 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 intraday duplicate. The threshold of positive or negative sequences was determined using different controls (G4s with different topologies, duplex and single 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 dataset processing, peak calling and differential analysis.

For both native and re-analyzed published datasets considered in this study, raw sequencing reads were aligned using Bowtie2 (ref. 43; 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<sup>44</sup> or using MACS2 (ref. 45; version 2.1.2), which also allows peak calling for G4-CHIP and G4access (peaks were considered below a *P* value of 10<sup>-10</sup> from the narrow peaks table). PASHA was used for ChIP-seq and MNase-seq datasets, using Drosophila spike-in for ChIP normalization<sup>46</sup> (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<sup>45</sup>. Bedgraph files generated by MACS2 were then converted to wig files (bin10) and scaled using the sequencing depth with PASHA. Wiggle files representing the average enrichment score every 10 bp 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 (Figs. 1a and 4a and Extended Data Figs. 6a, 7a and 9c). Finally, for nucleosome densities representation and analyses (MNase-seq), we smoothed the signal by replacing each 10 bp bin by the average of the five surrounding bins on each side and using an in-house script.

To assess differences in G4access peak intensities, DESeq<sup>47</sup> (version 1.26.0) was used with the MACS2 peak definition as genomic reference. 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 the mouse genome (mm9) or human genome (hg19) using TopHat2 for RNA-seq and ChrRNA-seq.

Alignment files were then treated using Cufflinks (v2.2.1) to generate RPKM used in Extended Data Fig. 3g. SNP analyses for mouse RNA-seq data from Fig. 5a are described in the section 'Allelic data analysis of mESC'.

**Motif search analysis and NDRs determination.** To analyze motifs associated with open chromatin/NDRs at promoters, we focused on CGI-containing promoters because those display an established intrinsic property to exclude nucleosomes<sup>2</sup>. 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<sup>48</sup> (<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  $-500$  and TSS (Fig. 1a) and on G4access or G4-ChIP (Extended Data Fig. 4) datasets using Multiple Em for Motif Elicitation tool MEME-ChIP<sup>49</sup> (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 Jaspas 2020 core nonredundant 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 was used (<http://www.yeastss.org/download/>). The first three motifs are displayed ranked by site numbers (Fig. 1a and Extended Data Figs. 4 and 10c).

**Allelic data analysis of mESC.** Raw sequencing reads were aligned strictly to the mm9 or JF1 (ref. 50) 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 G4access 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 was performed using DESeq<sup>47</sup> (version 1.26.0; Fig. 5b–d and 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<sup>51–53</sup> 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<sup>6,54</sup> (see also two sentences below), and their distributions were plotted using R (Figs. 1g and 3d,h and Extended Data Figs. 2d,e, 6c,d, 7c and 10b). To compare G4access and G4-ChIP, all peaks were resized at 90 bp ( $\pm 45$  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 (ten 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<sup>8</sup>—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 and Extended Data Figs. 3a and 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<sup>54</sup> with specific functions (Supplementary Software). 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 the same size distribution were used as control.

To assess GC richness effects in G4Hunter score distributions, all peaks were resized at 90 bp 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<sup>6</sup> (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 90 bp 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  $-200$  bp to gene TSSs were used (Extended Data Fig. 3e).

**Gene expression and gene ontology analyses.** Genes exhibiting G4access peaks in their promoters (within  $-200$  bp and their TSS) were determined using Bedtools<sup>55</sup>. Fragments per kilobase per million (FPKM) of all genes were determined using Cufflinks<sup>56</sup> (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<sup>57</sup> (<https://david.ncifcrf.gov/>) on the top 3,000 promoters ranked by G4access signals (Extended Data Fig. 3f).

**Methylation analysis.** Whole-genome bisulfite sequencing datasets were analyzed using Bismarck (version 0.22.3)<sup>58</sup> for Fig. 6a,b (K562 cells) and the computed beta files provided in ref. 59 (GSE186458) for Fig. 6c,d (human granulocytes). A selection on G4H1.2 exhibiting at least 2 CpG within the annotation and overlapping with G4access was performed (Fig. 6). A subset of 100,000 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 (Figs. 1a and 4a,d and Extended Data Figs. 1a, 7a,b, 8 and 9c), R scripts were developed and used for retrieving bin scores in defined regions from 10 bp bin-sized wiggle files<sup>44</sup>. Heatmaps were generated, viewed and color-scaled according to sample read depth using Java TreeView<sup>60</sup> (version 1.2.0-os; Figs. 4c and 6a,c). Regions were defined as centered on the G4access peak summits (from the narrow peak table of MACS2 at  $P < 10^{-10}$ ). For Fig. 4a,c, and because we applied a filter of G4access peaks with weak/moderate G4 predictions, G4access peaks (from the narrow peak table of MACS2,  $P < 10^{-8}$ ) overlapping G4 predictions G4H1.2 or QP long loops were considered ( $n = 10,018$ ). The long loop predictions were generated using the Quad-Parser consensus<sup>14</sup> 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 5 kb before the TSSs and after 5 kb of annotated termination

sites were determined. Based on the gene list selections, bin scores from wiggle files were used to rescale values between TSSs and transcription termination sites (gene body) of all genes using linear interpolation. In total, 1,000 points were interpolated for the gene body of each selected gene in all average profiles presented.

Metaprofiles of G4 access at gene bodies were performed as above at sites that do not overlap with H3K27ac ChIP-seq signals to avoid enhancers and unannotated promoters (Fig. 4d, right). Finally, DeepTools<sup>61</sup> (version 3.3.0) was used to generate metaprofiles in Fig. 6.

**Frequency analysis of pG4 in the human, mouse and yeast genomes.** pG4s were defined by G4Hunter using a window of 25 nt and thresholds ranging from G4H0.25 to G4H2.0 (refs. 6,54). Number of hits per kb of the sequenced genome was then determined. For TSSs, pG4 densities found in the TSS area (−200 bp to TSS) were calculated and expressed as pG4s/kb (Extended Data Fig. 10a).

**Analysis of genomic locations of pG4s or G4 access peaks.** pG4s 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 G4 access peaks detected by MACS2 (described in section ‘Average binding profiles and heatmaps’) were defined as follows (Extended Data Fig. 3b): promoter (−200 bp to TSS), 5′ gene body (TSS to +400 bp), gene body (from +400 bp after the TSS to −200 bp 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 G4 access signals; Extended Data Fig. 1e).

### Statistics and reproducibility

**Sample size.** No statistical method was used to predetermine the sample size. In the case of Fig. 2 analysis, 596 G4 access peaks were randomly selected within the 4,743 G4 access peaks that were found in common in the three cell lines (Hela, Raji and 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,c, and because we applied a filter of G4 access peaks with weak/moderate G4 prediction scores, G4 access peaks (from the narrow peak table of MACS2,  $P < 10^{-8}$ ) overlapping G4 predictions G4H1.2 and/or QP long loops were considered. The long loop 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 was repeated ten times for enrichment analyses. In Figs. 1g and 6a and Extended Data Figs. 2d,e, 6c,d, 7c and 10a,b genomic sequences were randomly selected and compared to experimental datasets of the same sizes.

**Blinding.** A total of 596 G4 access peaks were blindly tested using in vitro assays (Fig. 2) using coded oligonucleotides.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The genomic datasets published in this study are available under GSE187007. All data used from the previously published study are referenced in Supplementary Table 4. Source data are provided with this paper.

### Code availability

Code generated and used for this study has been deposited in the linked Zenodo repository: <https://zenodo.org/record/7912528> (ref. 62). The G4Hunter algorithm version code and functions used in this article are included as supplementary files (G4Hunter.r and G4HunterAccess\_function.r).

### References

- 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 (2021).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- Fenouil, R. et al. Pasha: a versatile R package for piling chromatin HTS data. *Bioinformatics* **32**, 2528–2530 (2016).
- Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
- Orlando, D. A. et al. Quantitative ChIP-seq normalization reveals global modulation of the epigenome. *Cell Rep.* **9**, 1163–1170 (2014).
- Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
- 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 (2014).
- Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
- 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 Res.* **44**, 1118–1132 (2016).
- Morison, I. M., Ramsay, J. P. & Spencer, H. G. A census of mammalian imprinting. *Trends Genet.* **21**, 457–465 (2005).
- Schulz, R. et al. WAMIDEX: a web atlas of murine genomic imprinting and differential expression. *Epigenetics* **3**, 89–96 (2008).
- Xie, W. et al. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816–831 (2012).
- Lacroix, L. G4HunterApps. *Bioinformatics* **35**, 2311–2312 (2019).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- Trapnell, C. et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
- 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 (2009).
- Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
- Loyfer, N. et al. A DNA methylation atlas of normal human cell types. *Nature* **613**, 355–364 (2023).
- Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).

61. 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).
62. Makrini, A., Esnault, C., Andrau, J. C. & Magat, T. Scripts and codes for G4access analysis. Zenodo. <https://zenodo.org/record/7912528> (2023).

## Acknowledgements

This work was supported by JCA lab (grant ANR-20-CE12-0023), FRM (grant AJE20130728183), INCA PLbIO (grant 2020-117) and CNRS 8Oprime 2021 (grant DeciphG4). This project has received financial support from the CNRS through the MITI interdisciplinary programs. C.E. was supported in part by an ARC grant (retour postdoc). We thank B. Lorioid and the Transcriptomics and Genomics Marseille-Luminy (TGML) platform for sequencing the G4access samples. We are grateful to D. Monchaud for providing us with G4-interfering molecules in exploratory experiments. TGML is a member of the France Génomique Consortium (ANR-10-INBS-0009). E.G.O., T.M. and S.B. 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, and decision to publish or preparation of the manuscript. We also thank P. Navarro and E. Kremer for critical reading of the manuscript and the Raman-Livaja lab for help in the yeast extract preparation.

## Author contributions

C.E. and E.G.O. performed important initial setup experiments for G4access that were further optimized by C.E. and T.M. A.Z.E.A. performed most of the bioinformatic analyses with the help of C.E. and

T.M., and L.L. developed and provided the modified G4Hunter scripts used in our analyses. A.C. contributed most of the G4 biophysical assays described in the manuscript under the supervision of J.L.M. and with the help of Y.L. and D.V. in the initial stages of the project. S.B. and T.M. contributed to the helicases, HDAC and transcription inhibition, while E.G.O. and C.E. performed the PDS treatment. T.M. and L.G. performed and analyzed the G4access in yeast and *Drosophila* with the help of AZEA for some bioinformatic analyses. D.L. and R.F. provided support for the mouse hybrid ES model by providing the ES cells, advice on cell culture and by establishing the list of DMR regions listed in Supplementary Table 1. S.S. helped and advised us for sequencing some of the G4access samples at the TGML. J.C.A. conceived the frame of the study, supervised the work and wrote the manuscript. All authors read, corrected and proofread the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

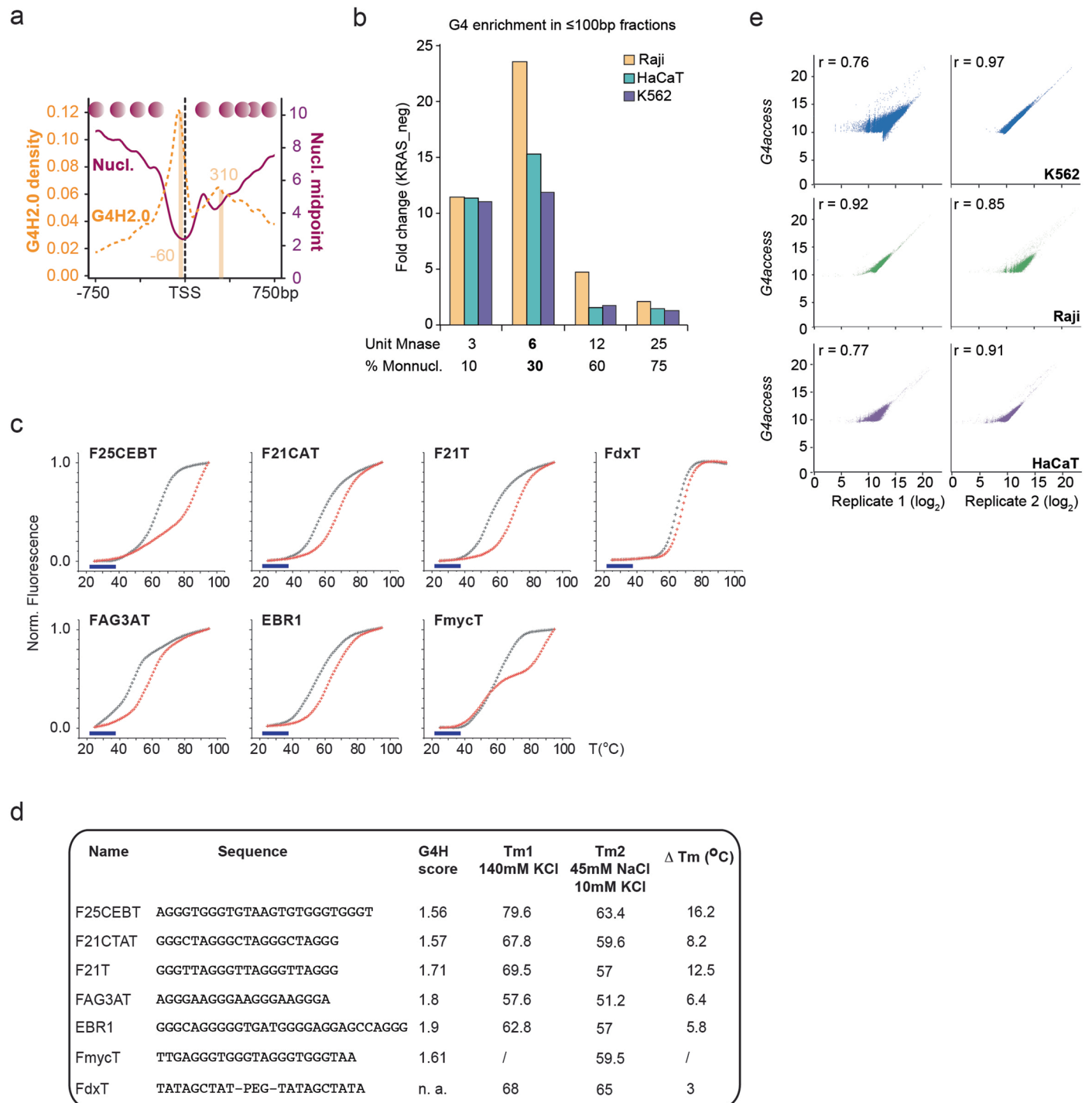
**Extended data** is available for this paper at <https://doi.org/10.1038/s41588-023-01437-4>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41588-023-01437-4>.

**Correspondence and requests for materials** should be addressed to Jean-Christophe Andrau.

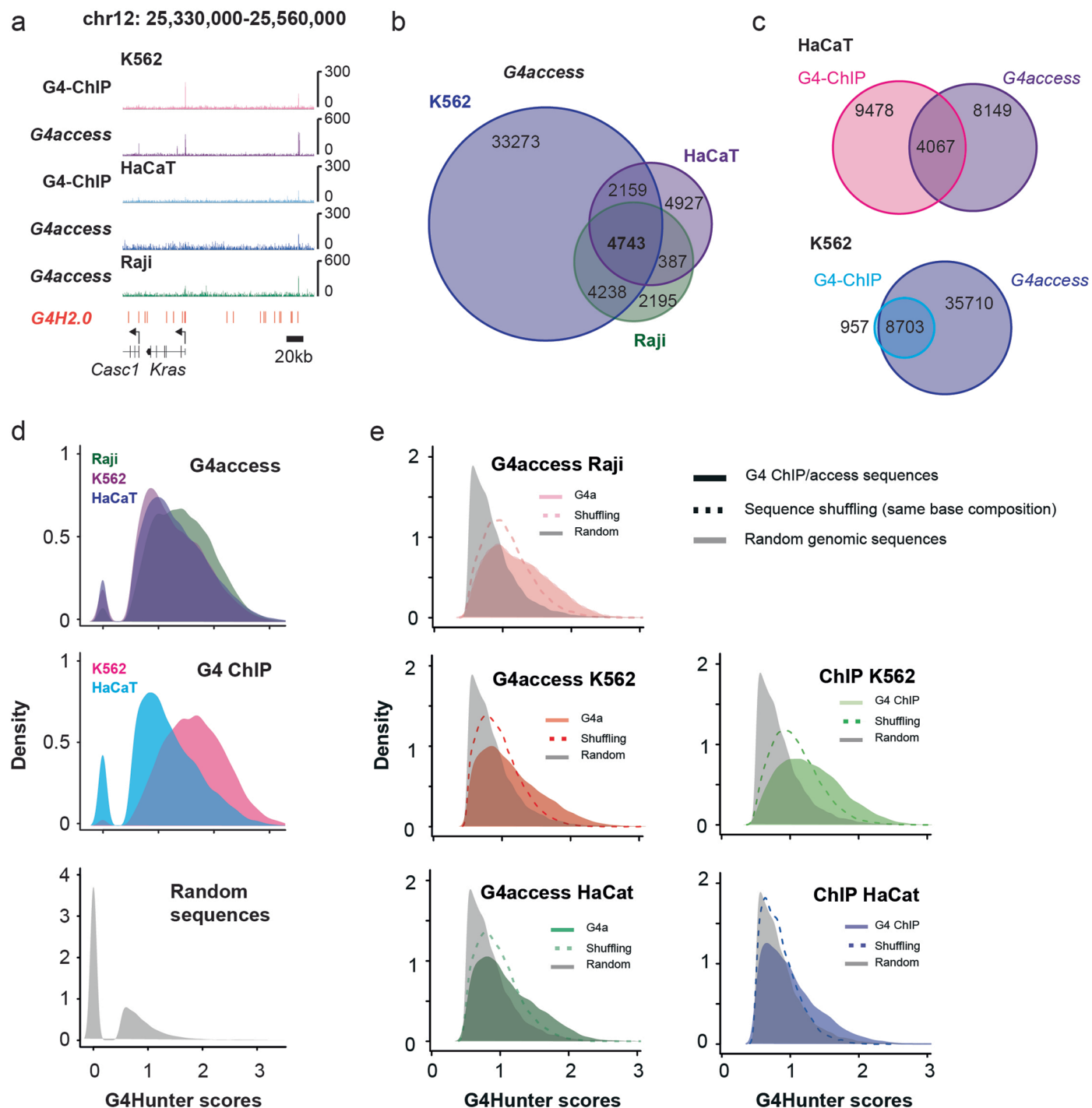
**Peer review information** *Nature Genetics* thanks F. Brad Johnson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Extended Data Fig. 1 | G4access signal optimization and characterization.**

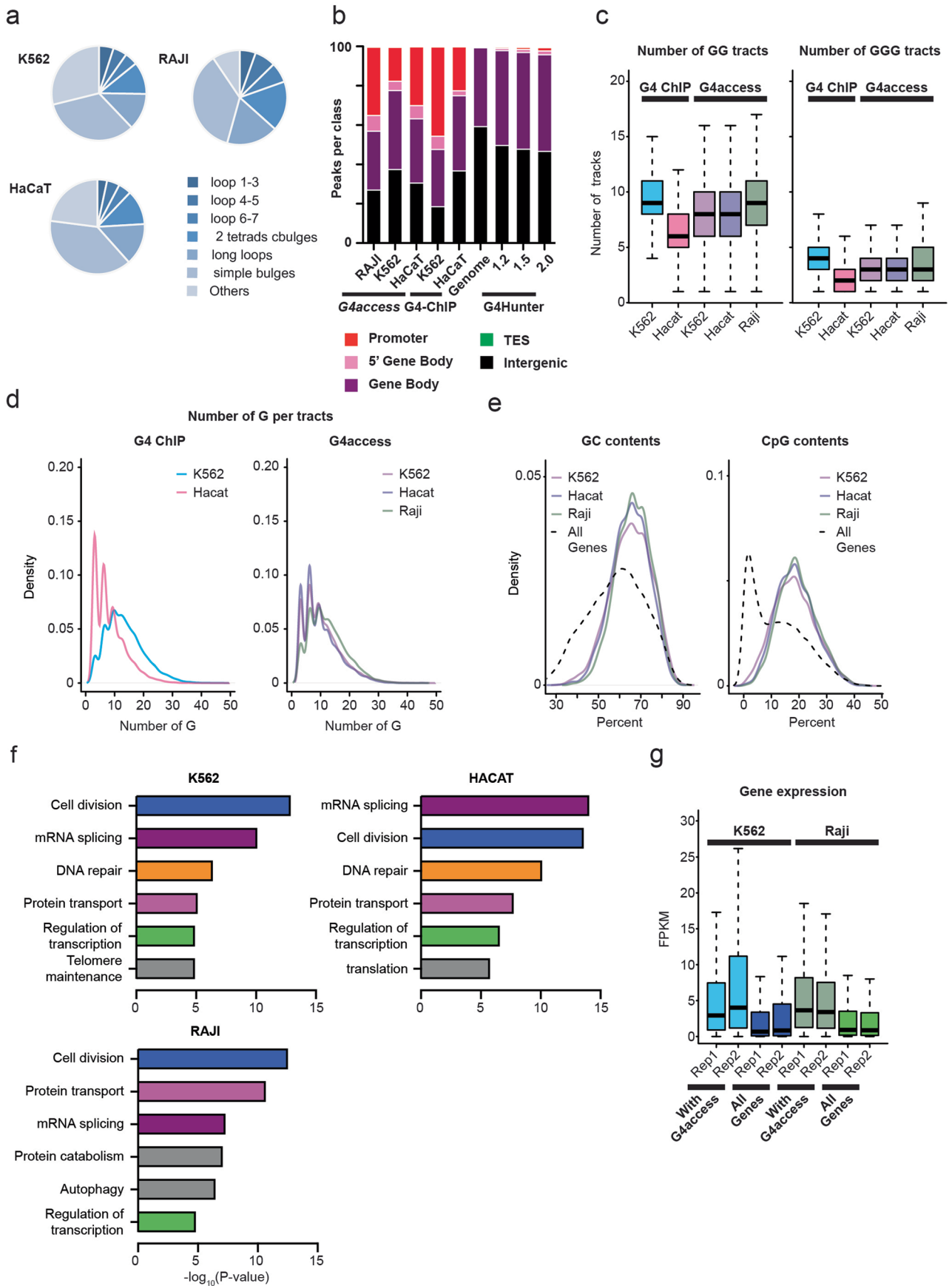
**a**- G4FS at all promoters are associated to open regions upstream of TSSs. The graph shows nucleosome and G4H2.0 densities in Raji cells, as well as the location of upstream and downstream peak's locations of G4 and nucleosome deep of all CGI containing promoters. **b**- qPCR quantification of a model G4 (human MAZ locus) in G4access preparations. 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  $\Delta Tm$  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 37  $^{\circ}C$  (blue bar). **e**- Correlation plots of G4access merged signals compared to individual biological replicates.

**Extended Data Fig. 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 tests of the overlaps  $<1 \times 10^{-4}$ ) **c-** Venn diagram of overlapping G4access and G4-ChIP peaks in the HaCaT and K562 cell lines. (Fisher tests of the overlaps  $<1 \times 10^{-4}$ ) **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

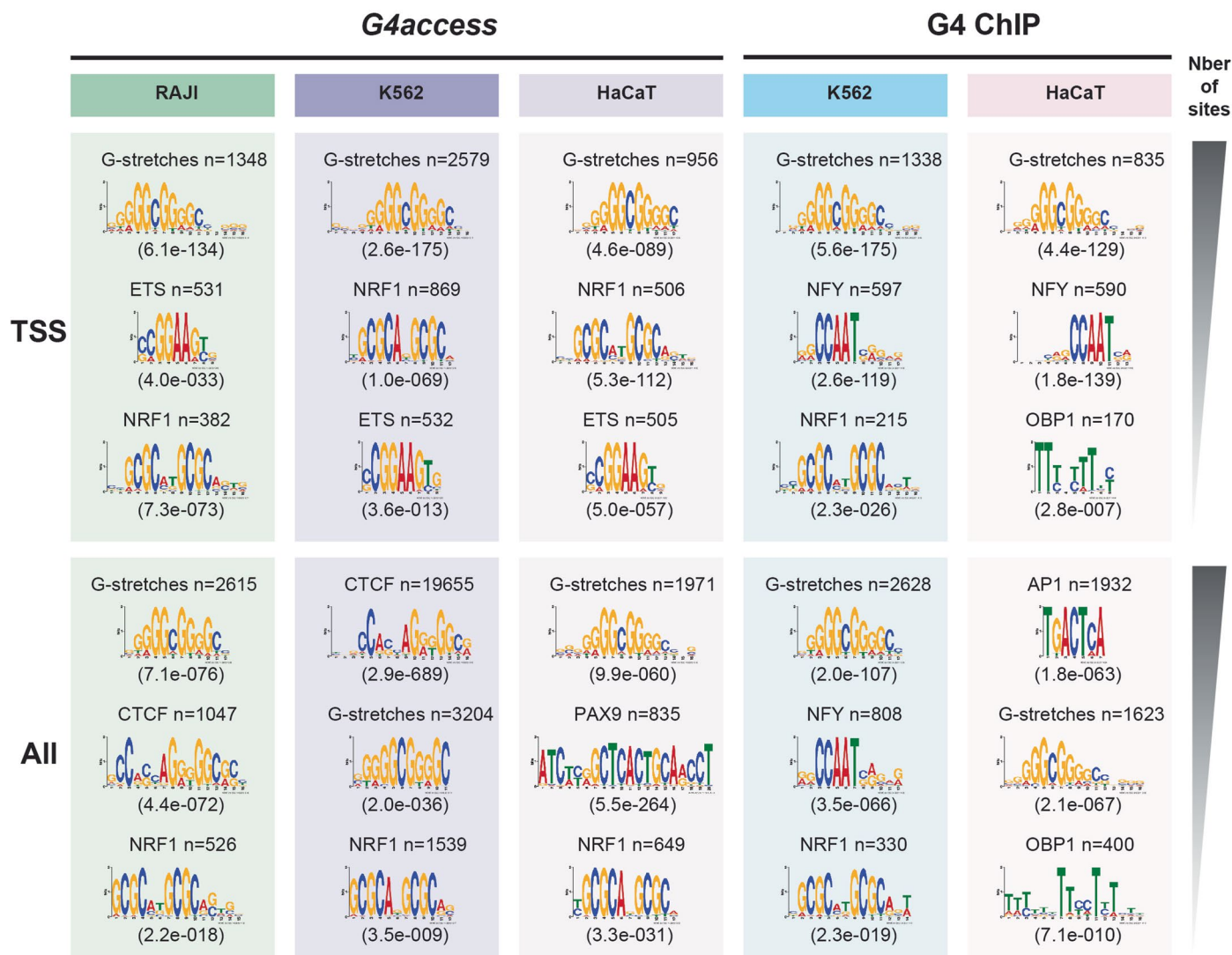
at 90 bp in G4-ChIP, G4access peaks and genomic DNA (40.000 annotations; see Methods). All distributions are highly significant compared to random selections (not shown) using a two-sided Wilcoxon test ( $p$ -value  $< 2 \times 10^{-16}$ ). **e-** G4Hunter prediction scores compared to shuffled sequences 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  $< 2 \times 10^{-16}$ ).



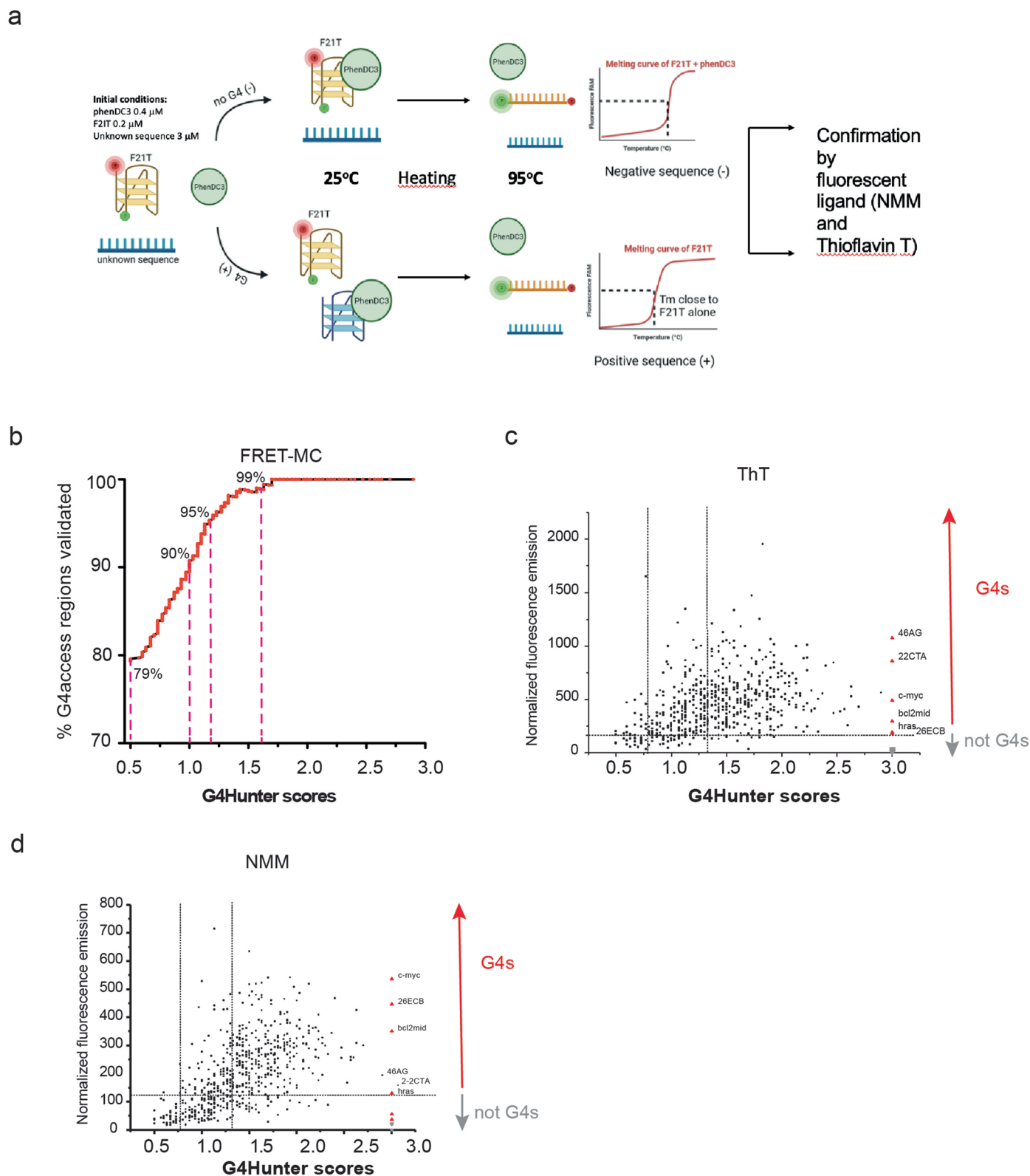
Extended Data Fig. 3 | See next page for caption.

**Extended Data Fig. 3 | G4access genomic localization, sequence characterization and association to gene expression programs.** **a-** G4 subtypes identified in the 3 cell lines (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, 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.

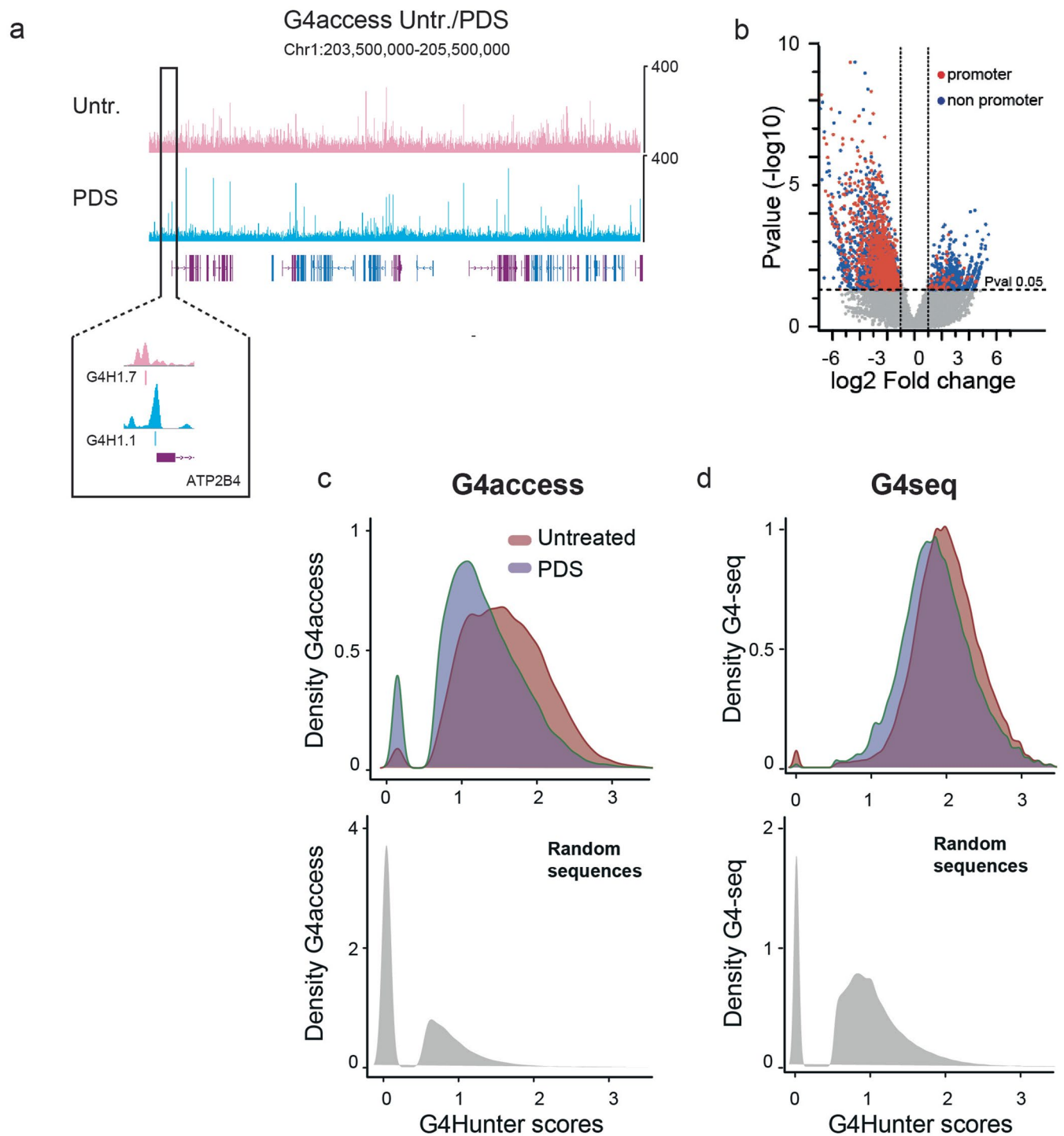


**Extended Data Fig. 4 | Motifs associated to G4access and G4 CHIP peaks in the 3 model cell lines (2 in the case of CHIP) at TSS and all sites as indicated. The sequence logos and statistics associated to this analysis were generated using the MEME algorithm. Presented motifs are ranked by occurrence (top 3). MEME-CHIP e-value are displayed.**



**Extended Data Fig. 5 | In vitro G4 characterization and validation.** **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).

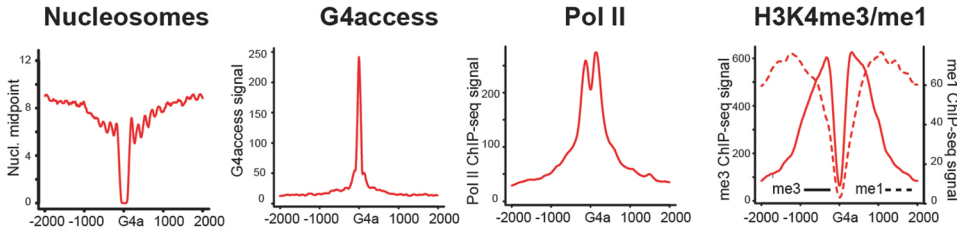


**Extended Data Fig. 6 | G4access measures G-quadruplex dynamics in response to cell treatments with a G4 ligand.** **a**- Genome browser view illustrating Pyridostatin (PDS, 10  $\mu$ M for 30 min) effect on G4access peaks dynamics in Raji cells (Chr1: 203,500,000-205,500,000). In the zoom area is shown the promoter *ATP2B4*, in which the main G4access signal redistributes from strong to weak G4Fs. **b**- DESeq analysis of G4access signal following 30 min

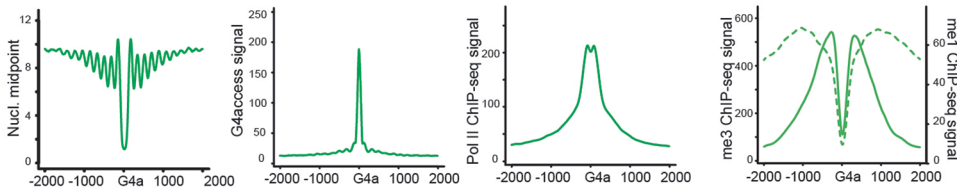
of treatment by PDS. The promoter-proximal (TSS) and non-promoter G4s are indicated in red and blue respectively (DESeq, p-value < 0.05). **c**- G4access score density is shifted toward weaker G4s following PDS treatment. **d**- G4seq score density is shifted toward weaker G4s following PDS treatment, although to a lesser extent than for G4access.

a

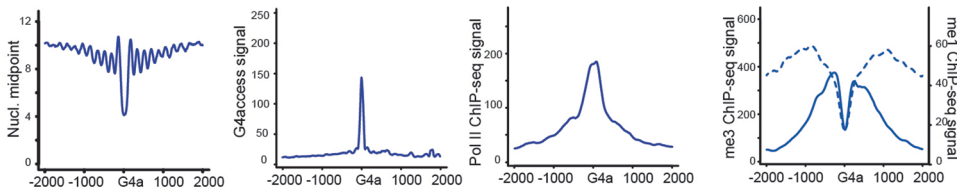
**Group 1 (n=844): Strong nucleosome depletion, low positioning**



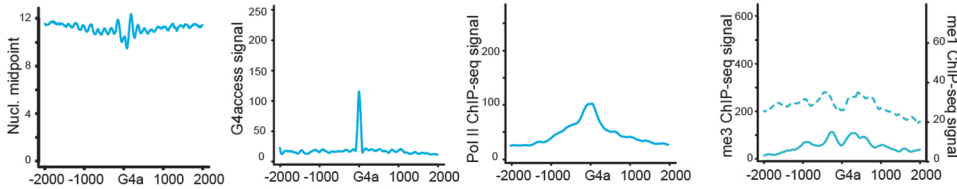
**Group 2 (n=6215): Strong nucleosome depletion, strong positioning**



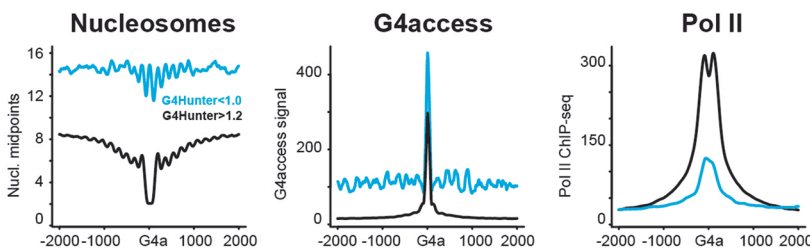
**Group 3 (n=2073): Moderate nucleosome depletion, strong positioning**



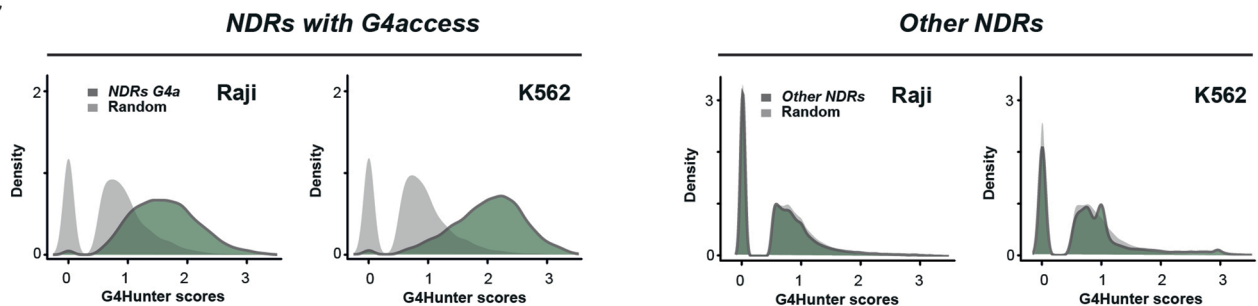
**Group 4 (n=2675): No nucleosome depletion, weaker positioning**



b



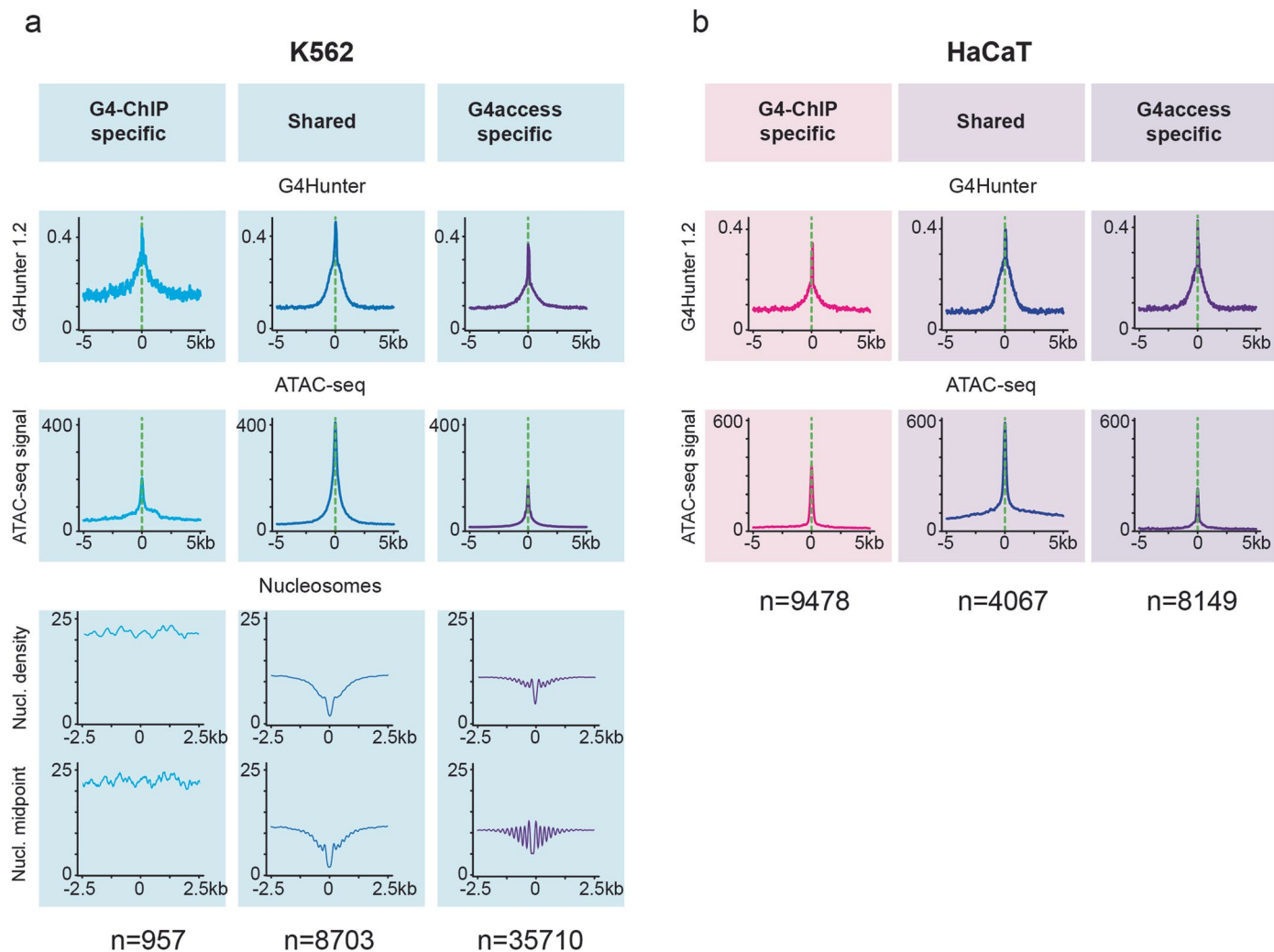
c



Extended Data Fig. 7 | See next page for caption.

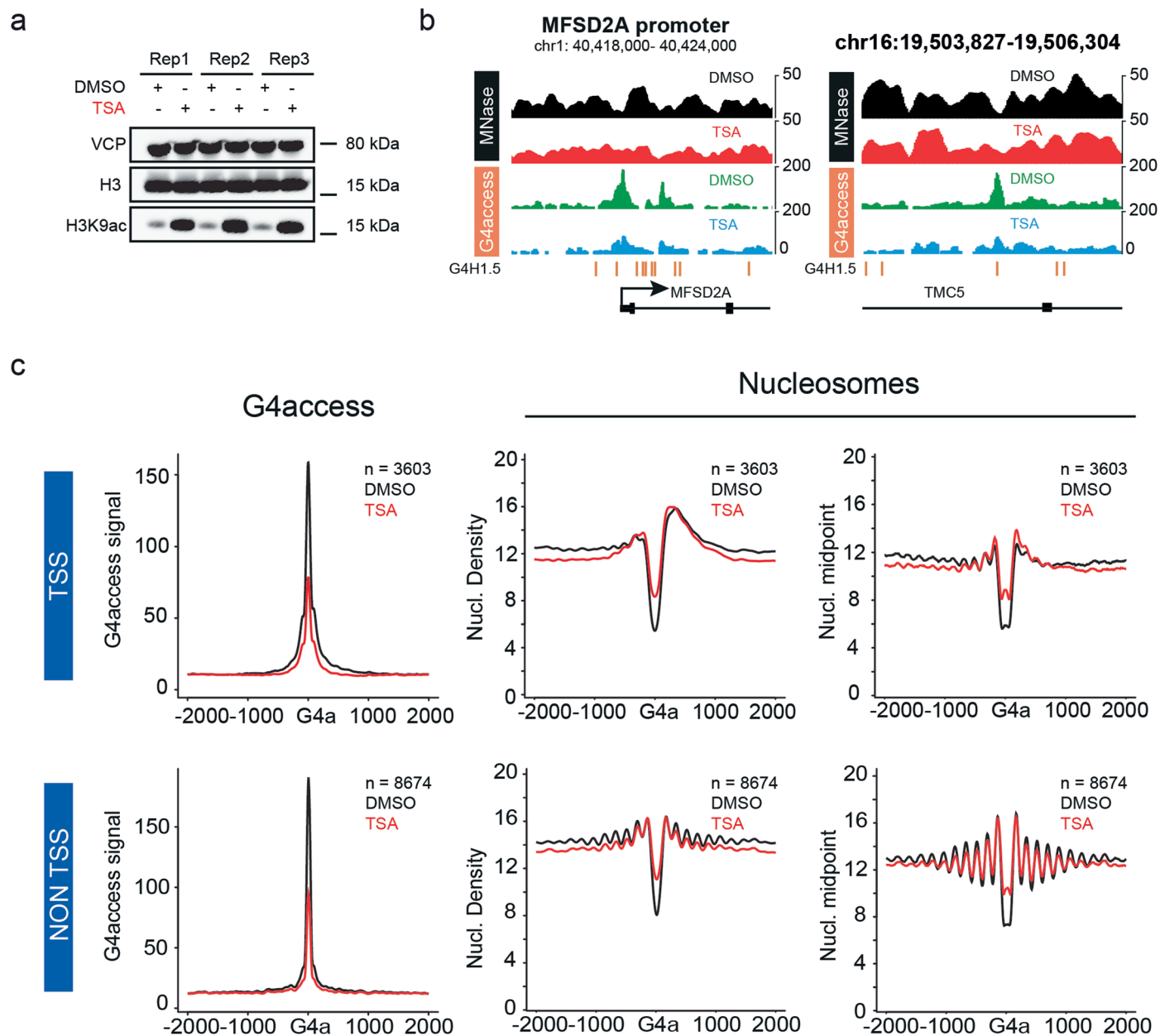
**Extended Data Fig. 7 | Nucleosome and Pol II features at G4access peaks, with or without strong G4 predictions.** **a-** average profiles of G4access regions depending of their nucleosome depletion level (relates to Fig. 4). Metaprofiles of MNase-seq (Nucleosome midpoints), G4access and Pol II ChIP-seq centered on G4access 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 gray. While distributions of scores at G4access associated NDRs are highly significant compared to random selections (using a two-sided Wilcoxon test,  $p$ -value <  $2.2 \times 10^{-16}$ ), distributions of G4Hunter scores at other NDRs are not significantly different to random selections.



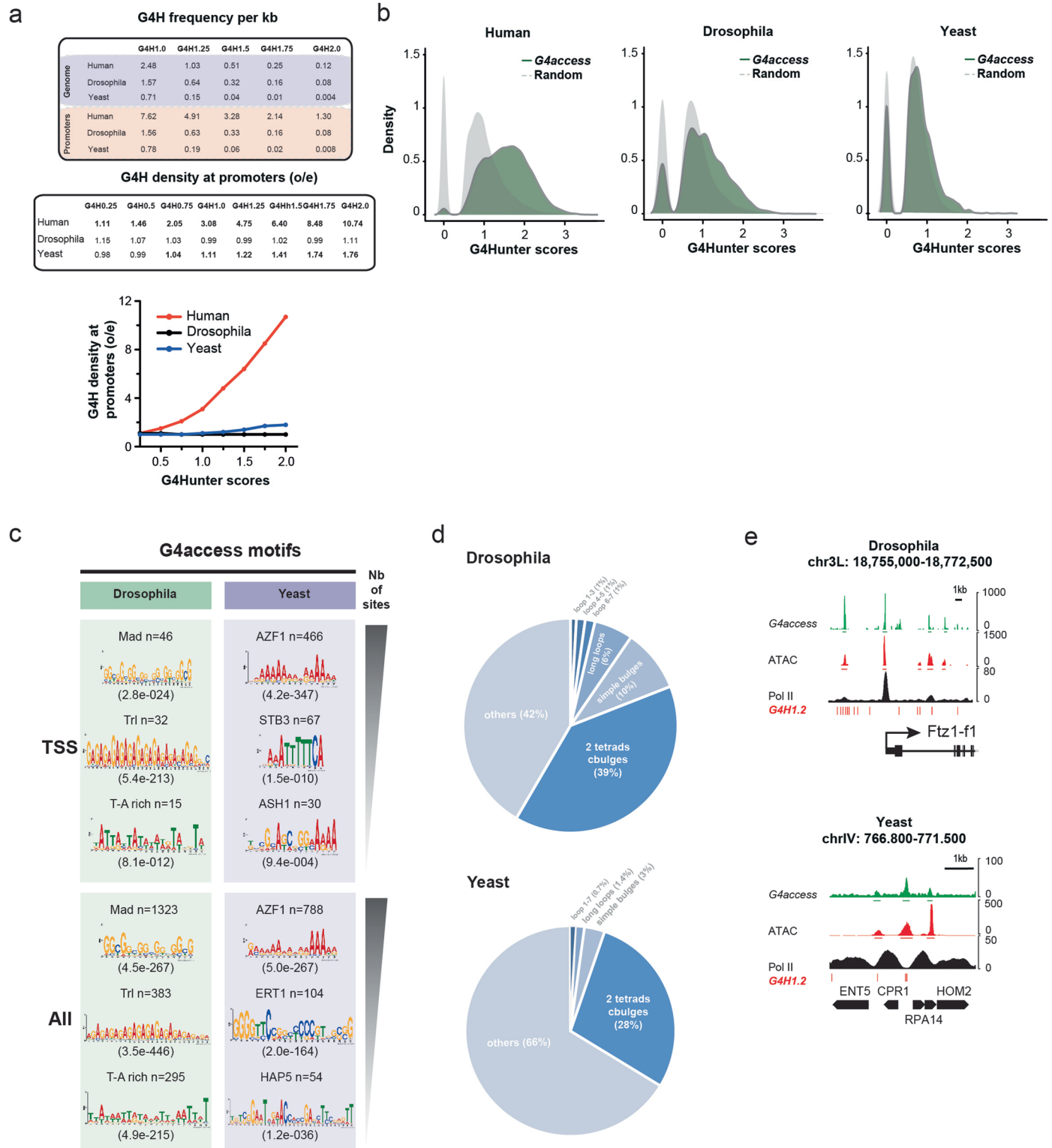
**Extended Data Fig. 8 | Nucleosome depletion at experimentally defined G-quadruplexes.** **a-** G4Hunter (G4H1.2) and chromatin landscape (ATAC-seq and MNase-seq density or positioning) profiling in K562 cells at sites with common or specific G4access and G4-ChIP peaks as indicated. **b-** G4Hunter (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.



**Extended Data Fig. 9 | G4access dynamics in response to nucleosome perturbation by the HDAC inhibitor TSA.** **a**- TSA treatment for 24 hours leads to H3K9acetylation increase. Western-blot of VCP and total H3 (loading controls) and of H3K9ac in 3 independent replicates are shown. **b**- Representative examples of G4access decrease associated to NDR closure at the *MFSD2A*

promoter (chr1: 40.418.000- 40.424.000) and the chr16:19.503.827-19.506.304 genomic region. **c**- TSA treatment for 24 hours leads to a global decrease of chromatin accessibility at NDRs associated to G4access decrease signal. Metaprofiles of G4access (left) and MNase-seq density and positioning (right) are shown at all TSSs (up) and non TSS (bottom) sites.



**Extended Data Fig. 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.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

ENCODE  
GEO  
SRA-Toolkit (version 2.9.0) was used for data collection.

Data analysis

bedtools (v2.21.0)  
Bismark (0.22.3)  
bowtie2 (2.1.0)  
HTSeq (0.6.1p1)  
cufflinks (v2.2.1)  
deeptools (3.3.0)  
FastQC (0.10.1)  
MACS2 (2.1.2)  
MEME-ChIP (5.0.2)  
PASHA (0.99.21)  
R (3.3.1)  
TopHat2 (v2.0.10)  
TreeView (1.2.0-osx)  
DESeq (version 1.26.0)  
G4hunter (DOI 10.5281/zenodo.7816663)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All genomic raw datasets used in this study are available.

The extended table 3 lists all genomic datasets used in this study.

Genomic datasets produced in this study are deposited under the accession number GSE187007

Figures with associated raw data:

Fig. 1d-h, 3, 4, 5, 6

ED Fig. 1E, 2, 3, 4, 6, 7, 8, 9, 10,

Datasets used in this study:

- 1 Raji Homo Sapiens Mnase-seq None Nucleosome Ours\_GSE187007 This study
- 2 Raji Homo Sapiens ChIP-seq None PolII Ours\_GSE187007 This study
- 3 Raji Homo Sapiens ChIP-seq None H3H4me1 GSE52914 published
- 4 Raji Homo Sapiens ChIP-seq None H3H4me3 GSE52914 published
- 5 Raji Homo Sapiens ChIP-seq None H3K27ac GSE52914 published
- 6 Raji Homo Sapiens G4access None G-quadruplexes Ours\_GSE187007 This study
- 7 Raji Homo Sapiens G4access PYRIDOSTATIN G-quadruplexes Ours\_GSE187007 This study
- 8 Raji Homo Sapiens G4access DMSO G-quadruplexes Ours\_GSE187007 This study
- 9 Raji Homo Sapiens G4access TRIPTOLIDE G-quadruplexes Ours\_GSE187007 This study
- 10 Raji Homo Sapiens G4access KM05283 G-quadruplexes Ours\_GSE187007 This study
- 11 Raji Homo Sapiens ChIP-seq DMSO PolII Ours\_GSE187007 This study
- 12 Raji Homo Sapiens ChIP-seq TRIPTOLIDE PolII Ours\_GSE187007 This study
- 13 Raji Homo Sapiens ChIP-seq KM05283 PolII Ours\_GSE187007 This study
- 14 Raji Homo Sapiens RNA-seq None ChrRNA Ours\_GSE187007 This study
- 15 K562 Homo Sapiens G4 ChIP None G-quadruplexes GSE107690 published
- 16 K562 Homo Sapiens G4 ChIP None H3K27ac ENCODE/Uw published
- 17 K562 Homo Sapiens G4 ChIP None PolII ENCODE/HAIB published
- 18 K562 Homo Sapiens ATAC-seq None Open chromatin GSE99173 published
- 19 K562 Homo Sapiens Mnase-seq None Nucleosome ENCODE published
- 20 K562 Homo Sapiens G4access None G-quadruplexes Ours\_GSE187007 This study
- 21 K562 Homo Sapiens RNA-seq None ChrRNA GSE90238 published
- 22 K562 Homo Sapiens WGBS None DNA methylation GSE86747 published
- 23 Hacat Homo Sapiens G4access None G-quadruplexes Ours\_GSE187007 This study
- 24 Hacat Homo Sapiens ChIP-seq None G4-ChIP GSE76688 published
- 25 Hacat Homo Sapiens ATAC-seq None Open chromatin GSE76688 published
- 26 Granulocytes Homo Sapiens Mnase-seq None Nucleosome GSE25133 published
- 27 Granulocytes (in vitro reconstituted) Homo Sapiens Mnase-seq None In\_Vitro\_nucl GSE25133 published
- 28 Granulocytes Homo Sapiens WGBS None DNA methylation GSE186458 published
- 29 Hela Homo Sapiens G4access DMSO G-quadruplexes Ours\_GSE187007 This study
- 30 Hela Homo Sapiens G4access TSA G-quadruplexes Ours\_GSE187007 This study
- 31 Hela Homo Sapiens G4access Control siRNA G-quadruplexes Ours\_GSE187007 This study
- 32 Hela Homo Sapiens G4access DHX36 siRNA G-quadruplexes Ours\_GSE187007 This study
- 33 Hela Homo Sapiens G4access WRN siRNA G-quadruplexes Ours\_GSE187007 This study
- 34 ES - cells (C57BL/6JxJF1) Mus musculus G4access None G-quadruplexes Ours\_GSE187007 This study
- 35 ES - cells (JF1xC57BL/6J) Mus musculus G4access None G-quadruplexes Ours\_GSE187007 This study
- 36 ES - cells (C57BL/6JxJF1) Mus musculus RNA-seq None mRNA GSE58523 published
- 37 ES - cells (JF1xC57BL/6J) Mus musculus RNA-seq None mRNA GSE58523 published
- 38 S2 Drosophila melanogaster G4access None G-quadruplexes Ours\_GSE187007 This study
- 39 S2 Drosophila melanogaster ATAC-seq None Open chromatin GSE103177 published
- 40 S2 Drosophila melanogaster ChIP-seq None PolII GSE41440 published
- 41 S288C (BY4741) Saccharomyces cerevisiae G4access None G-quadruplexes Ours This study
- 42 S288C (DBY12000) Saccharomyces cerevisiae ATAC-seq None Open chromatin GSE118581 published
- 43 S288C (BY4742) Saccharomyces cerevisiae ChIP-seq None PolII GSE59370 published
- 44 Human genome Homo Sapiens G4 prediction None G4Hunter2.0 track Ours\_GSE187007 This study based on Bedrat et al., 2016
- 45 Human genome Homo Sapiens G4 prediction None G4Hunter1.5 track Ours\_GSE187007 This study based on Bedrat et al., 2016
- 46 Mouse genome Mus musculus G4 prediction None G4Hunter2.0 track Ours\_GSE187007 This study based on Bedrat et al., 2016
- 47 Mouse genome Mus musculus G4 prediction None G4Hunter1.5 track Ours\_GSE187007 This study based on Bedrat et al., 2016

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Referring to Fig2: 595 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 is 3.68% for Confidence Level of 95% following the central limit theorem.
Data exclusions	For Fig.4a and 4c, and because we applied a filter of G4 access peaks with weak/moderate G4 predictions, G4access peaks (from the narrow peak table of MACS2, p-value < 10 <sup>-8</sup> ) overlapping G4 predictions G4H1.2 and/or QP longloops were considered. The longloop predictions were generated using the Quad-Parser consensus. Genes that are not expressed (RPKM=0) were not included in the gene expression analysis presented in ED Fig.3g. In MNase-seq heatmaps, saturated or absence of signals expanding all along the displayed genomic areas were removed.
Replication	Experiments were repeated in replicates as indicated in the manuscript, (Between 2 and 4 times). All replications were successful. Please note that EDFig1b has been performed as technical replicates only.
Randomization	Randomization of genome sequences were repeated 10 times for enrichment analyses. In Fig. 1g, 6a, ED Fig. 2d-e,6c-d,7c and 10a-b genomic sequences were randomly selected and compared to experimental datasets of same sizes.
Blinding	595 G4access peaks were blindly tested using in vitro assays (Fig. 2) using coded oligonucleotides.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	WRN (W0393-200UL - Dilution 1 / 2000) DHX36 (Ab70269 - Dilution 1 / 5000) VCP (Ab11433 - Dilution 1 / 2000) H3K9ac (C15410004 - lot a.4243-0013p - Dilution 1 / 1000) Total H3 (Ab1791 - lot gr291516 -1 - Dilution 1 / 10 000) Total PolII (F-12 - sc-55492 - lot H2019 - 0.8ug/IP - Santa Cruz Biotechnology) Total PolII (N20 - sc899 - lot H3115 - 0.8ug/IP - Santa Cruz Biotechnology)
Validation	WRN (W0393-200UL - This antibody is widely used by the scientific community (eg. PMID: 27672210)) DHX36 (Ab70269 - This antibody is widely used by the scientific community (eg. PMID: 32461552)) VCP (Ab11433 - This antibody is widely used by the scientific community (eg. PMID: 36732333)) H3K9ac (C15410004 - This antibody is widely used by the scientific community (eg. PMID: 32826895)) Total H3 (Ab1791 - This antibody is widely used by the scientific community (eg. PMID: 33135214)) Total PolII (F-12 - sc-55492 - This antibody is widely used by the scientific community (eg. PMID:35081363)) Total PolII (N20 - sc899 - This antibody is widely used by the scientific community (eg. Nature. 2016 Feb 4; 530(7588): 113–116.; Nature. 2015 Feb 12;518(7538):249-53.; Nature Genetics volume 43, pages630–638(2011)...) and were validated by western-blot and ChIP-seq in our laboratory (eLife. 2014; 3: e02105.; Mol Cell. 2018 Jan 4;69(1):48-61.e6.; Nat Struct Mol Biol. 2011 Jul 17;18(8):956-63.).

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

- B lymphocyte, The Raji line of lymphoblast-like cells was established by R.J.V. Pulvertaft in 1963 from a Burkitt's lymphoma of the left maxilla of an 11-year-old Black male. (PMID: 14086209). ATCC-CCL-86  
 - The continuous cell line K-562 was established by Lozzio and Lozzio from the pleural effusion of a 53-year-old female with chronic myelogenous leukemia in terminal blast crises. (PMID: 95026). ATCC-CCL-243  
 - HaCaT cells, human keratinocytes were kindly provided by the screening platform in oncology, Montpellier Cancer Research Institute (IRCM).  
 - HeLa cells were isolated in 1951 from a cervical carcinoma derived from a 31-year-old patient. HeLa H9 cells were a gift of S. Emiliani, Institut Cochin Paris. (see Tantale et al. (2021). Nature communications 12, 4503.  
 - Drosophila melanogaster S2 cells, Schneider's Drosophila Line 2, this line was established in 1969 by I. Schneider from several hundred 20 to 24 hour embryos. (PMID: 4625067). ATCC-CRL-1963  
 - 2i-medium-derived ESCs hybrid between M. m. domesticus strain C57BL/6J and M. m. molossinus strain JF1 were derived recently (Sanli, I. et al. Cell Rep 23, 337-348, doi:10.1016/j.celrep.2018.03.044 (2018). 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.  
 - BY4741 is part of a set of deletion strains derived from S288C in which commonly used selectable marker genes were deleted by design in order to minimize or eliminate homology to the corresponding marker genes in commonly used vectors without significantly affecting adjacent gene expression. The yeast strains were all directly descended from FY2, which is itself a direct descendant of S288C. Variation between BY4741 and S288C is miniscule. BY4741 was used as a parent strain for the international systematic Saccharomyces cerevisiae gene disruption project.

Authentication

None of the cell line used were authenticated

Mycoplasma contamination

We confirm that all cell lines were tested negative for mycoplasma contamination

Commonly misidentified lines  
(See [ICLAC](#) register)

No commonly misidentified cell lines (ILAC, Register of Misidentified Cells version 12)

## ChIP-seq

### Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).  
 Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

*May remain private before publication.*

GSE187007  
 To review GEO accession GSE187007:  
 Go to <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE187007>

Additional processed data files provided:

dm6\_G4H1.2.bed  
 dm6\_G4H1.5.bed  
 dm6\_G4H2.0.bed  
 hg19\_G4H1.2.bed  
 hg19\_G4H1.52.bed  
 hg19\_G4H2.0.bed  
 mm9\_G4H1.2.bed  
 mm9\_G4H1.5.bed  
 mm9\_G4H2.0.bed  
 sc3\_G4H1.2.bed  
 sc3\_G4H1.5.bed  
 sc3\_G4H2.0.bed  
 ES\_B6xJF1\_G4access\_bin10.wig  
 ES\_JF1xB6\_G4access\_bin10.wig  
 HeLa\_H9\_siDHX36\_rep1\_rep2.wig  
 HeLa\_H9\_siINT\_rep1\_rep2.wig  
 HeLa\_H9\_siWRN\_rep1\_rep2.wig  
 Raji\_DMSO\_control\_TSA\_G4access\_merge\_bin10.wig  
 Raji\_DMSO\_control\_TSA\_mnase\_merge\_bin10\_smoothed5.wig  
 Raji\_DMSO\_control\_TSA\_mnase\_Midpoint\_merge\_bin10\_smoothed5.wig  
 Raji\_TSA\_G4access\_merge\_bin10.wig  
 Raji\_TSA\_mnase\_merge\_bin10\_smoothed5.wig  
 Raji\_TSA\_mnase\_midpoint\_merge\_bin10\_smoothed5.wig  
 dm6\_G4access\_bin10.wig  
 HaCaT\_G4access\_bin10.wig  
 K562\_G4access\_bin10.wig  
 Raji\_G4access\_bin10.wig  
 Raji\_G4access\_DMSO\_bin10.wig  
 Raji\_G4access\_KM\_bin10.wig  
 Raji\_G4access\_PDS\_24h\_bin10.wig  
 Raji\_G4access\_PDS\_30min\_bin10.wig

Raji\_G4access\_PDS\_60min\_bin10.wig  
 Raji\_G4access\_Triptolide\_bin10.wig  
 Raji\_Mnase-seq\_untreated1\_bin10\_Scaled\_Smooth5.wig  
 Raji\_Mnase-seq\_untreated2\_bin10\_Scaled\_Smooth5.wig  
 Raji\_PDS\_chRNA\_negative\_NormalizedERCC.wig  
 Raji\_PDS\_chRNA\_positive\_NormalizedERCC.wig  
 Raji\_Pol\_II\_DMSO\_bin10\_NormalisedDroso\_Smoothed5.wig  
 Raji\_Pol\_II\_KM\_bin10\_NormalisedDroso\_Smoothed5.wig  
 Raji\_Pol\_II\_Triptolide\_bin10\_NormalisedDroso\_Smoothed5.wig  
 Raji\_PolII\_untreated1\_bin50\_Scaled.wig  
 Raji\_PolII\_untreated2\_bin50\_Scaled.wig  
 sc3\_G4access\_bin10.wig  
 K562\_input\_bin10.wig

## Files in database submission

Samples:

dm6 G4access Rep1  
 dm6 G4access Rep2  
 ES B6xJF1 Rep1 1  
 ES B6xJF1 Rep1 2  
 ES B6xJF1 Rep2 1  
 ES B6xJF1 Rep2 2  
 ES JF1XB6 Rep1 1  
 ES JF1XB6 Rep1 2  
 ES JF1XB6 Rep2 1  
 ES JF1XB6 Rep2 2  
 HaCaT G4access Rep1  
 HaCaT G4access Rep2  
 K562 G4access Rep1 1  
 K562 G4access Rep1 2  
 K562 G4access Rep1 3  
 K562 G4access Rep2  
 K562 Input G4access Rep1  
 Raji G4access DMSO Rep1  
 Raji G4access DMSO Rep2  
 Raji G4access KM Rep1  
 Raji G4access KM Rep2  
 Raji G4access PDS 30min Rep1  
 Raji G4access PDS 30min Rep2  
 Raji G4access Rep1  
 Raji G4access Rep2  
 Raji G4access Triptolide Rep1  
 Raji G4access Triptolide Rep2  
 Raji PolII DMSO Rep1  
 Raji PolII DMSO Rep2  
 Raji PolII KM Rep1  
 Raji PolII KM Rep2  
 Raji PolII Triptolide Rep1  
 Raji PolII Triptolide Rep2  
 sc3 G4access Rep1  
 sc3 G4access Rep2  
 Hela\_H9\_siDHX36\_rep1  
 Hela\_H9\_siDHX36\_rep2  
 Hela\_H9\_siNT\_rep1  
 Hela\_H9\_siNT\_rep2  
 Hela\_H9\_siWRN\_rep1  
 Hela\_H9\_siWRN\_rep2  
 Raji\_DMSO\_control\_TSA\_G4access\_rep1  
 Raji\_DMSO\_control\_TSA\_G4access\_rep2  
 Raji\_DMSO\_control\_TSA\_G4access\_rep3  
 Raji\_DMSO\_control\_TSA\_mnase\_rep1  
 Raji\_DMSO\_control\_TSA\_mnase\_rep2  
 Raji\_DMSO\_control\_TSA\_mnase\_rep3  
 Raji\_TSA\_G4access\_rep1  
 Raji\_TSA\_G4access\_rep2  
 Raji\_TSA\_G4access\_rep3  
 Raji\_TSA\_mnase\_rep1  
 Raji\_TSA\_mnase\_rep2  
 Raji\_TSA\_mnase\_rep3  
 Raji\_PolII\_Rep1  
 Raji\_PolII\_Rep2  
 Raji\_Mnase-seq\_Rep1  
 Raji\_Mnase-seq\_Rep2  
 Raji\_chrRNA-seq\_rep1  
 Raji\_chrRNA-seq\_rep2

[https://genome.ucsc.edu/s/Cyril\\_Esnault/hg19\\_tracks\\_garcia\\_et\\_al\\_manuscript\\_DMSO\\_KM\\_TRP](https://genome.ucsc.edu/s/Cyril_Esnault/hg19_tracks_garcia_et_al_manuscript_DMSO_KM_TRP)  
[https://genome.ucsc.edu/s/Cyril\\_Esnault/hg19\\_tracks\\_garcia\\_et\\_al\\_manuscript\\_G4access](https://genome.ucsc.edu/s/Cyril_Esnault/hg19_tracks_garcia_et_al_manuscript_G4access)  
[https://genome.ucsc.edu/s/Cyril\\_Esnault/hg19\\_tracks\\_garcia\\_et\\_al\\_manuscript\\_G4access\\_PDS](https://genome.ucsc.edu/s/Cyril_Esnault/hg19_tracks_garcia_et_al_manuscript_G4access_PDS)  
[https://genome.ucsc.edu/s/Cyril\\_Esnault/mm9\\_tracks\\_garcia\\_et\\_al\\_manuscript\\_G4access\\_EScells](https://genome.ucsc.edu/s/Cyril_Esnault/mm9_tracks_garcia_et_al_manuscript_G4access_EScells)  
[https://genome.ucsc.edu/s/Cyril\\_Esnault/sacCer3\\_tracks\\_garcia\\_et\\_al\\_manuscript\\_Yeast](https://genome.ucsc.edu/s/Cyril_Esnault/sacCer3_tracks_garcia_et_al_manuscript_Yeast)  
[https://genome.ucsc.edu/s/Cyril\\_Esnault/dm6\\_tracks\\_garcia\\_et\\_al\\_manuscript\\_Drosophila](https://genome.ucsc.edu/s/Cyril_Esnault/dm6_tracks_garcia_et_al_manuscript_Drosophila)

## Methodology

### Replicates

All analyses on original data were performed as two independent replicates.

### Sequencing depth

Samples	Sequenced reads	% aligned
dm6 G4access Rep1	6632108	73.16
dm6 G4access Rep2	6921075	70.95
ES B6xJF1 Rep1 1	18874845	64.17
ES B6xJF1 Rep1 2	61770373	87.26
ES B6xJF1 Rep2 1	28546506	91.73
ES B6xJF1 Rep2 2	37592995	94.25
ES JF1XB6 Rep1 1	29293079	93.38
ES JF1XB6 Rep1 2	57252725	94.65
ES JF1XB6 Rep2 1	34415134	93.26
ES JF1XB6 Rep2 2	69553782	94.48
HaCaT G4access Rep1	33988016	91.97
HaCaT G4access Rep2	40842705	81.90
K562 G4access Rep1 1	5209244	82.38
K562 G4access Rep1 2	2255501	81.64
K562 G4access Rep1 3	5468993	74.27
K562 G4access Rep2	84111027	98.32
K562 Input G4access Rep1	47501193	98.20
Raji G4access DMSO Rep1	12404725	61.70
Raji G4access DMSO Rep2	11380878	64.26
Raji G4access KM Rep1	8180651	54.12
Raji G4access KM Rep2	5888879	53.52
Raji G4access PDS 30min Rep1	36081308	83.08
Raji G4access PDS 30min Rep2	21972460	85.30
Raji G4access Rep1	11069970	83.37
Raji G4access Rep2	32245753	91.46
Raji G4access Triptolide Rep1	6280935	53.23
Raji G4access Triptolide Rep2	9626350	63.21
Raji PolII DMSO Rep1	52623435	93.12
Raji PolII DMSO Rep2	62034709	93.68
Raji PolII KM Rep1	58052471	94.21
Raji PolII KM Rep2	59236618	91.13
Raji PolII Triptolide Rep1	52658499	92.46
Raji PolII Triptolide Rep2	66729771	90.76
sc3 G4access Rep1	7165216	89.16
sc3 G4access Rep2	10281530	89.25
Raji_DMSO_control_TSA_G4access_rep1	45218419	87.0
Raji_DMSO_control_TSA_G4access_rep2	53284287	86.5
Raji_DMSO_control_TSA_G4access_rep3	54394222	85.2
Raji_DMSO_control_TSA_mnase_rep1	134476425	96.9
Raji_DMSO_control_TSA_mnase_rep2	143115951	97.0
Raji_DMSO_control_TSA_mnase_rep3	122740642	97.0
Raji_TSA_G4access_rep1	32490552	87.2
Raji_TSA_G4access_rep2	42400525	88.0
Raji_TSA_G4access_rep3	49477053	88.4
Raji_TSA_mnase_rep1	128263282	96.8
Raji_TSA_mnase_rep2	132580546	96.2
Raji_TSA_mnase_rep3	133833620	96.7
Hela H9 siNT Rep1	89181169	71.91
Hela H9 siNT Rep2	90171415	84.24
Hela H9 siDHX36 Rep1	94588942	78.36
Hela H9 siDHX36 Rep2	69206660	73.30
Hela H9 siWRN Rep1	55171189	69.42
Hela H9 siWRN Rep2	53649252	68.35
Raji_PolII_Rep1	76259482	92.97
Raji_PolII_Rep2	55373464	92.55
Raji_Mnase-seq_Rep1	62261437	98.08
Raji_Mnase-seq_Rep2	60791243	97.97
Raji_chrRNA-seq_rep1	56655887	93.60
Raji_chrRNA-seq_rep2	64989206	95.90

### Antibodies

Total PolII (F-12 - sc-55492 - lot H2019 - 0.8ug/IP - Santa Cruz Biotechnology)

Antibodies	Total PolII (N20 - sc899 - Santa Cruz Biotechnology)
Peak calling parameters	<p>For both native and re-analyzed published data sets considered in this study, raw sequencing reads were aligned using Bowtie2 to the human (hg19), mouse (mm9), drosophila (dm6) and yeast (SC3) genomes. Aligned reads were elongated in silico using the DNA fragment size inferred from paired-reads using an in-house developed PASHA R pipeline or using MACS2 which also allows peak calling for G4-ChIP and G4access (peaks were considered below a p-value of 10<sup>-10</sup> from the narrow peaks table). PASHA was used for ChIP-seq and MNase-seq datasets, using drosophila spike-in for ChIP normalization (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. 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 ED Fig. 1a, 7a, 7b, 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.</p> <p>To assess differences in G4access peak intensities, DESeq 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.</p>
Data quality	Data quality was assessed with FastQC (0.10.1)
Software	All sequencing reads were aligned to the hg19,mm9, dm6 and SC3 genome assemblies using Bowtie2 and analyzed with MACS2.